



Utilizing a fusion of remote sensing data and machine learning models to forecast flood risks to agriculture in Hanoi City, Vietnam

Anh Ngoc Thi Do^{1,2}

Received: 6 November 2023 / Accepted: 25 May 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Floods represent a substantial and consequential form of natural disasters within Hanoi City. To minimize the detrimental effects on agriculture, an all-encompassing decision support instrument is necessary for flood management and alert systems. The primary aim of the current study is to delineate flood susceptible regions by employing SPOT satellite imagery and a hybrid Principal Component Analysis-Support Vector Machine (PCA-SVM) model, thereby gauging the influence of floods on land utilization for agricultural purposes in Hanoi City. The prediction results demonstrate a high model performance with $R^2_{\text{test}}=0.904$, and $\text{AUC}=0.921$. Areas classified as exhibiting high to very high flood susceptible encompass 55.882% of the overall expanse, while those classified as having low and very low flood risk account for 10.357% and 6.278% respectively. The amalgamation of satellite imagery and the PCA-SVM model in the formulation of flood susceptible zoning maps confers valuable insights to bolster flood prevention endeavors. The current research findings will make a substantial contribution to the strategic planning and preservation of food security for the nation.

Keywords Flood susceptibility · Agriculture · SPOT satellite · CRNN model

✉ Anh Ngoc Thi Do
anhngocthido269@gmail.com

¹ Thai Nguyen university– Lao Cai campus, Lao Cai, Vietnam

² Faculty of Natural resources and Environmental management, Thai Nguyen university– Lao, Cai campus, Lao Cai, Vietnam

1 Introduction

Floods are a frequently occurring and highly destructive occurrence in the natural realm, resulting in significant harm to agricultural crops and posing a threat to food security (Hirabayashi et al. 2008; Bhattacharya et al. 2019). In recent years, floods have escalated in severity, exerting substantial influence on the agricultural sector, disrupting infrastructure and causing economic and social setbacks (Lee and Mohamad 2014; Su et al. 2021). Consequently, there has been an increasing emphasis on the mapping of flood-susceptible areas for the purposes of early warning systems and impact evaluation (Do et al. 2022a, b, c). Traditional approaches to flood mapping rely on ground surveys and aerial observations, but the irregular and extensive nature of floods renders these methods time-consuming, expensive, and impedes the prompt assessment of flood-related effects on the economy and livelihoods (Peng and Peng 2018; Chen et al. 2019; Do and Tran 2023a, b, c).

Hanoi City has recently encountered frequently instances of flooding during prolonged rainfall (Anh 2021). Due to rapid urbanization, numerous main roads have been expanded without coordinated drainage planning, leading to localized flooding in various areas during heavy rainfall. Among these areas, three-quarters of the total area comprise plains, with agriculture still occupying a significant portion of the economic structure. Consequently, the impact of floods has resulted in localized inundation (Anh 2023). Nowadays, one- and two-dimensional models are progressively being employed to enhance the precision of simulation results (Lin et al. 2006; Liu et al. 2015). However, the primary drawback of this technique is the substantial amount of input data required for the model, necessitating extensive time and effort for field surveys, data collection, and model calibration (Klemas 2015; Lin et al. 2016). Furthermore, in susceptible areas due to torrential rain, the current hydraulic models have not been able to provide a comprehensive resolution (Brakenridge et al. 1994). The development of remote sensing technology and GIS has supplied powerful tools for data acquisition, spatial analysis, and graphical representation in the monitoring and identification of flooded areas (Zaharia et al. 2017; Al-Abadi 2018). Remote sensing data possesses the capacity to gather information over vast areas and for extended periods with a high repetition frequency (Do et al. 2022b). The integration of remote sensing and GIS within machine learning models facilitates rapid calculation and assessment of areas at susceptible of flooding and inundation (Do et al. 2022a, b, c).

In recent years, the utilization of machine learning techniques and data mining has proven to be valuable in the realm of flood prediction (Mosavi et al. 2018; Do and Tran 2023a). One particularly popular method employed in this context is the Support Vector Machine (SVM) non-linear model (Khan et al. 2019). The distinguishing feature of SVM lies in its ability to employ kernel functions to transform the original feature space, thereby facilitating the handling of non-linear features (Do and Tran 2023c). Consequently, SVM is able to classify flood data based on non-linear features such as the interplay between environmental factors, topography, and weather. Additionally, the SVM model is effective in managing large datasets, thus reducing the time and resources required for computation (Costache 2019). However, it should be noted that SVM is susceptible to data noise and exhibits relatively high computa-

tional complexity, especially when complex kernel functions are employed or when dealing with substantial amounts of data (Do and Tran 2023c). Therefore, in the current study, the Principal Component Analysis (PCA) algorithm has been utilized to identify the principal components of the data, the components with the highest variance, with the aim of enhancing the performance of the SVM prediction model (Xu and Wang 2005).

In order to assess the effects of floods on agriculture, it is necessary to possess a spatial distribution map of agricultural land (Do et al. 2022a, b, c). Over the past few years, there has been rapid development in various machine learning algorithms for mapping land cover/land use (LULC), including agricultural land (Pham et al. 2023a, b, 2024; Do et al. 2023). The efficacy of machine learning methods in handling limited sample data has been demonstrated through their generalization and noise resistance capabilities. As a result, machine learning algorithms have become invaluable tools in processing remote sensing data and offering solutions within the realm of agriculture (Anh 2023). Currently, algorithms based on the Convolutional Recurrent Neural Network (CRNN) have emerged as a prominent subject of interest in this field, utilizing convolutional layers and pooling techniques (Rajendran et al. 2020; Moharram and Sundaram 2023). The CRNN model the ability to automatically extract features and information from the original images based on the combination of two types of network, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), thereby enhancing the accuracy of classification results (Cao et al. 2019).

In light of this reality, the primary objectives of the current investigation are as follows: (i) to examine the practicality of the CRNN model in extracting agricultural land using SPOT 7 satellite imagery; (ii) to map flood-susceptible areas using the PCA-SVM model; and (iii) to evaluate the impacts of floods on agricultural land use in Hanoi City.

2 Materials and methods

2.1 Study area

The Hanoi City area is distinguished by numerous advantages for the advancement of high-quality agriculture. It is the largest city in Vietnam, encompassing an area of roughly 3,360km², and it ranks second in terms of population and population density among Vietnam's 63 provinces and cities. Situated in the northwest of the central Red River Delta, within the latitude range of 20°34' to 21°18' north and the longitude range of 105°17' to 106°02' east, Hanoi is situated in the triangle of the Red River Delta, an area known for its fertile and abundant land (Fig. 1). Hanoi has an extensive hydrological system consisting of various small and large rivers, including the Red River, Duong River, Da River, Nhue River, Cau River, Day River, and Ca Lo River. The city possesses all the necessary prerequisites for the development of a contemporary agricultural sector and serves as a market for high-quality rice, catering to a substantial and steadfast demand.

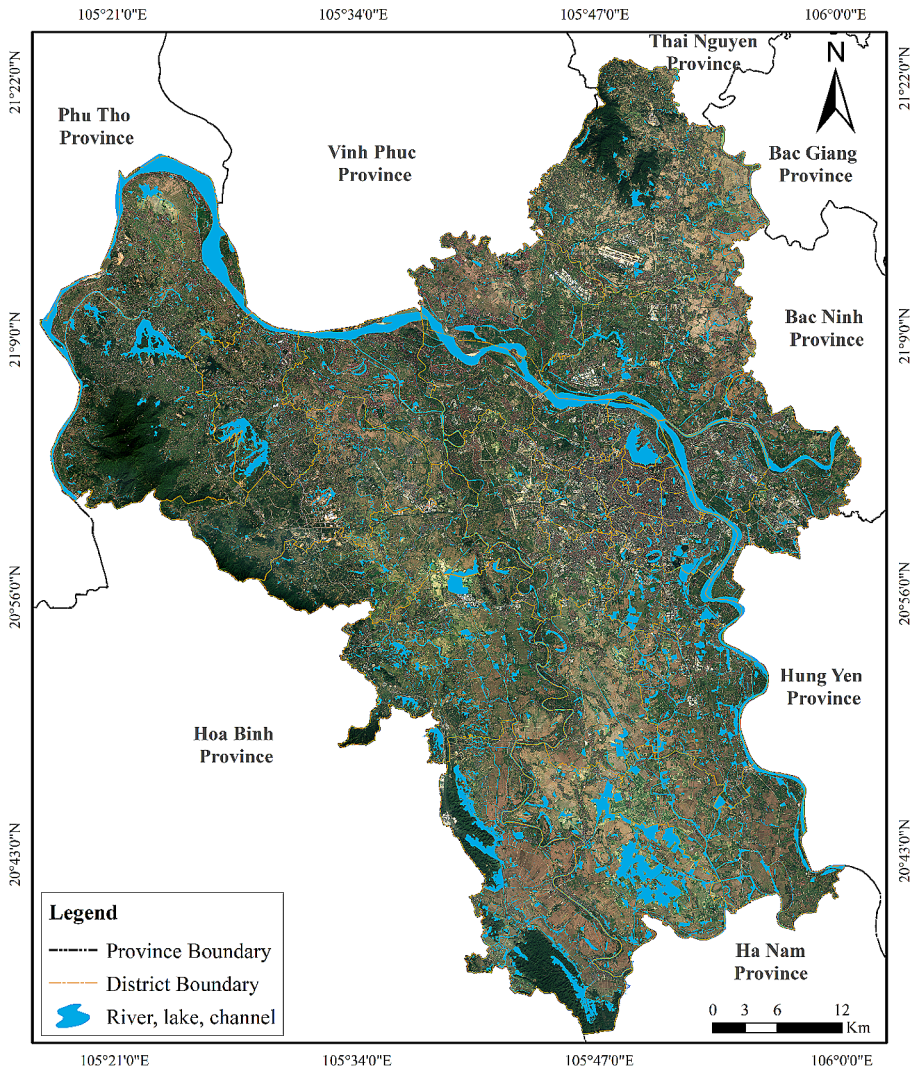


Fig. 1 A map showing Hanoi city

2.2 Data collection and SPOT image preprocessing

This investigation collected satellite imagery data from SPOT 7 in January 2023, featuring a resolution of 1.5 m, while ensuring that the acquired images contained a cloud cover of less than 10% in the study area. The SPOT 7 images underwent atmospheric and spectral correction to derive radiometric values. Several atmospheric correction models, such as COST, DOS, MODTRAN, ATCOR, or FLAASH, can be employed for atmospheric correction (Pham et al. 2023a, b, 2024; Do, 2024). To enhance accuracy, the ATCOR (Atmospheric and Topographic Correction) atmospheric correction model was integrated into the PCI Geomatica 2018 software to

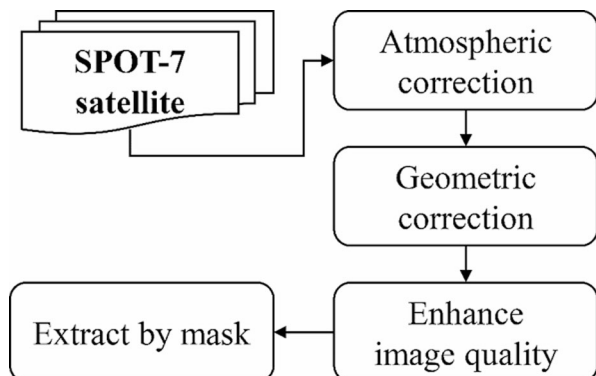
execute atmospheric correction. The quality of satellite imagery is heavily dependent on the image processing. Typically, the acquired satellite image channels have pixel values that are distributed within a narrow range compared to the display capabilities of the image. Each individual channel tends to be relatively dark or bright when displayed. Therefore, to enhance the contrast of the image, we need to perform an image stretching operation. This is done to transform the actual gray-level range of the original image into a gray-level range that the display device is capable of showing. After the processing, the satellite image will have good image quality, appropriate contrast, accurate color representation, even color distribution, and will be suitable for LULC cover classification purposes. The image preprocessing procedure entailed four steps, as depicted in Fig. 2, encompassing geometric correction and enhancement facilitated by ENVI 5.3 software, utilizing the UTM projection grid, VN-2000 coordinate system, and zone 48 with a resolution of 2.5 m.

The current study executed the acquisition of sample data for image classification within the research area. In order to reference during the classification and prediction process, a land use map of Hanoi city for the year 2022 has been collected. The current study has collected a total of 363 samples encompassing all land cover classes and surveyed flood-sensitive areas throughout the entire research region. These samples include on-site data collection (95 samples), utilization of high-resolution imagery from Google Earth, historical flood locations, and direct sampling on SPOT 7 satellite imagery. Five various types of land cover have been classified, including: other land, construction, forest and urban green space, water surface, and agriculture. The collected dataset was divided into two, with 70% of the samples utilized for training the classification model (training data) and the remaining 30% utilized for validating the classification results (testing data).

2.3 Selection of LULC classification method

Currently, there exist multiple machine learning algorithms designed for land use land cover (LULC) classification using satellite imagery (Nahuelhual et al. 2012; Hua 2017). Among these algorithms is the Convolutional Recurrent Neural Network (CRNN), which combines the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Connectionist Temporal Classification (CTC) (Fig. 3).

Fig. 2 Simple flow of SPOT image preprocessing



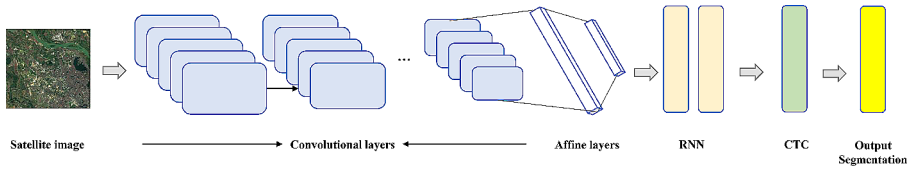


Fig. 3 Land use land cover classification steps using CRNN model

CRNN is commonly employed for image classification tasks (Cao et al. 2019). This neural network architecture seamlessly integrates feature extraction, sequence modeling, and decoding within a unified framework. CNN is utilized to extract structural features from the images (Kattenborn et al. 2021), RNN is employed to model the sequential information (Mou et al. 2017), and CTC is utilized to enhance the performance of the CRNN model (Hsu and Li 2021). The training process of the CRNN model for image segmentation generally involves the utilization of training data consisting of image-label pairs. The model is adjusted through the optimization of a loss function, such as cross-entropy loss, to attain the most accurate segmentation results (Kattenborn et al. 2021). Numerous studies have demonstrated the commendable performance of CRNN in LULC classification (Wu and Prasad 2017; Zhao and Zetsu 2018). To evaluate the classification performance, the current study employed the overall accuracy (OA) and cross-validated accuracy (CV) metrics (Anh 2021; Do et al. 2022b).

$$OA = (TP + TN) / (TN + TP + FN + FP) \quad (1)$$

$$CV = \left(\frac{2TP^2}{(TP + FN) \cdot (TP + FP)} \right) / \left(\frac{2TN^2}{(TP + FN) \cdot (TP + FP)} \right) \quad (2)$$

where TP represents accurately classified agricultural objects, TN represents accurately classified non-agricultural objects, FP represents incorrectly classified agricultural objects, and FN represents incorrectly classified non-agricultural objects.

2.4 PCA-SVM model in susceptible prediction of floods

Principal Component Analysis (PCA) is a technique utilized to diminish the dimensionality of data within the feature space by identifying the principal components of the data (Do et al. 2022b). Within this investigation, the PCA algorithm was employed to eliminate less significant components from a total of 16 input variables (Table 1) in order to decrease the dimensionality of the data and optimize the model for better performance. Following the reduction of data dimensionality through PCA, the principal components were selected as input features for the Support Vector Machine (SVM) model, which is a supervised machine learning model utilized for regression problems aiming to discover an optimal hyperplane for classifying data into flood-susceptible and flood-insensitive (Do et al. 2022b; Do and Tran 2023c). When employing a linear kernel, the SVM decision function takes the form (Gao et al. 2003):

Table 1 Database for flood susceptibility mapping

Variables	Description
Altitude	The height of a point on the Earth's surface in relation to sea level
Aspect	Derived from a digital elevation model
Slope	DEM
Curvature	
Land cover/land use	Land categories include agricultural land, forest land, building land, water surface...
(Normalized difference vegetation index) NDVI	$(\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$
Temperature	Data was gathered from hydrological monitoring stations
Rainfall	
Agriculture density	The degree of concentration of agricultural activities in a specific area
Forest density	The degree of concentration of forests in a specific area
Construction density	The degree of concentration of construction works in a specific area
Water density	The degree of concentration of water bodies (such as lakes, rivers, and streams) in a specific area
Distance to agriculture	The distance from a specific location to the nearest agricultural areas
Distance to forest	The distance from a specific location to the nearest forested areas
Distance to construction	The distance from a specific location to the nearest construction sites
Distance to water surface	The distance from a specific location to the nearest water surface

$$f(x) = \text{sign}(w^T x + b) \quad (3)$$

where w is the weight vector, x is the feature vector of the data sample, b is the bias term, and $\text{sign}()$ is the sign function. In the present study, the parameter C was used to regulate the model's regularization. A larger C value directs the SVM model to prioritize error minimization and tolerate fewer violations, while a smaller value of C prioritizes minimizing the magnitude of w and allows more violations of the margin. In the SVM model, the optimal value of the parameter C is determined through techniques such as grid search or error optimization. This process involves training and evaluating the model using both the training and testing datasets.

To evaluate the accuracy of the PCA-SVM model in predicting flood sensitivity in the study area, the study utilized the coefficient of determination (R^2), and the root mean square error (RMSE) (A. N. T. Do et al. 2022a, b, c; T. A. T. Do et al., 2022; Do et al., 2024; Do, 2024).

$$R^2 = \frac{\sum_{i=1}^k [(Y_i - Y)(X_i - X)]}{\sqrt{\sum_{i=1}^k (Y_i - Y)^2} \cdot \sqrt{\sum_{i=1}^k (X_i - X)^2}} \quad (4)$$

Table 2 Area and percentage of area of each type of LCL

Land cover/land use	Area (ha)	Percentage (%)
Other land	1,428.591	0.425
Construction	108,418.382	32.269
Forest, urban green space	44,699.180	13.304
Water surface	39,457.853	11.744
Agriculture	141,982.871	42.258

Table 3 Performance of the CRNN classification model

Land cover/land use	OA (%)	CV
Other land	85.934	0.820
Construction	87.395	0.842
Forest, urban green space	91.385	0.899
Water surface	95.027	0.931
Agriculture	80.282	0.782
LCLU	88.005	0.855

$$\text{RMSE} = \sqrt{\frac{1}{k} \sum_{i=1}^k (Y_i - X_i)^2} \quad (5)$$

where Y_i , and Y represent the predicted variable and the mean value, respectively; X_i , and X represent the observed variable and the mean value, respectively, and k is the sample size.

3 Results

3.1 CRNN model in agricultural land classification

In the current study, the utilization of the CRNN model has been employed for the purpose of classifying LULC in Hanoi city. The results of this classification are displayed in Table 2, which the area and percentage for each respective LULC type. Agricultural land, with an expansive area of 141,982.871 ha (42.258%); followed by construction (32.259%); forest and urban green space (13.304%); water surface (11.744%); and finally, other land only accounting for a mere 0.425% of the total area. To gauge of precision of the SPOT 7 image classification results, the study utilized the overall accuracy (OA) and the coefficient of variation (CV) as evaluation metrics. The evaluation results are as follows: the overall accuracy achieved 88.005% and the CV reached 0.855 (Table 3). Among these results, the water surface category displayed the highest classification accuracy (OA=95.027%, and CV=0.931), followed by forest, urban green space, other land, and finally agriculture (OA=80.282%, and CV=0.782). Overall, the CRNN model demonstrated a high level of accuracy and suitability for classifying land cover within the Hanoi city area.

Figure 4 displays the spatial distribution map of agricultural land, which was classified using the CRNN model and SPOT satellite imagery. It can be observed that agricultural land encompasses a significant area and is primarily distributed in suburban areas. Areas devoid of agricultural land are predominantly situated in inner-city

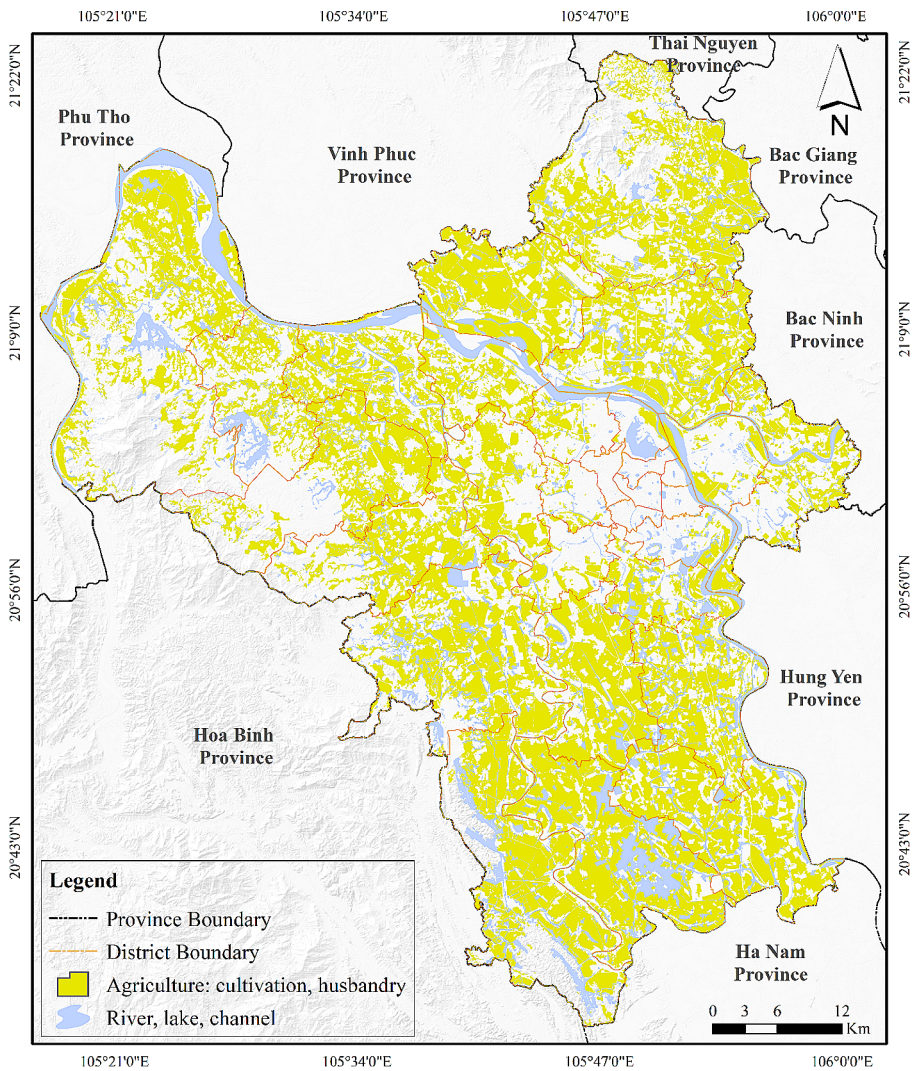


Fig. 4 Spatial distribution of agricultural land in Hanoi city in 2023

areas such as Ba Dinh and Dong Da districts, as well as in highland areas such as Ba Vi district (Fig. 4). In general, agriculture plays a prominent role in the economy and food production within the research area. However, it exhibits an uneven distribution, primarily concentrated in suburban areas, delta regions, and areas in close proximity to rivers and streams. With water surfaces covering 11.744% of the total area, flood events can cause significant losses to crops in these areas.

3.2 The importance of variables

Sixteen variables were selected as input variables to prediction flood susceptibility in the research area, as depicted in Fig. 5. Nevertheless, incorporating an excessive number of input variables would give rise to overfitting concerns for the SVM model. Therefore, principal component analysis (PCA) was employed to diminish the dimensionality of the data and improve the predictive performance of the model. The results of the PCA analysis indicate a distinct separation of data points into separate clusters. The principal component PC1, which represents rainfall, accounts for a significant portion of the data's variance, with, with $R^2=0.342$ (Fig. 6) indicating its utmost importance in relation to flood variables. Rainfall holds a critical role in flood modeling in Hanoi, an area characterized by a humid tropical climate with a

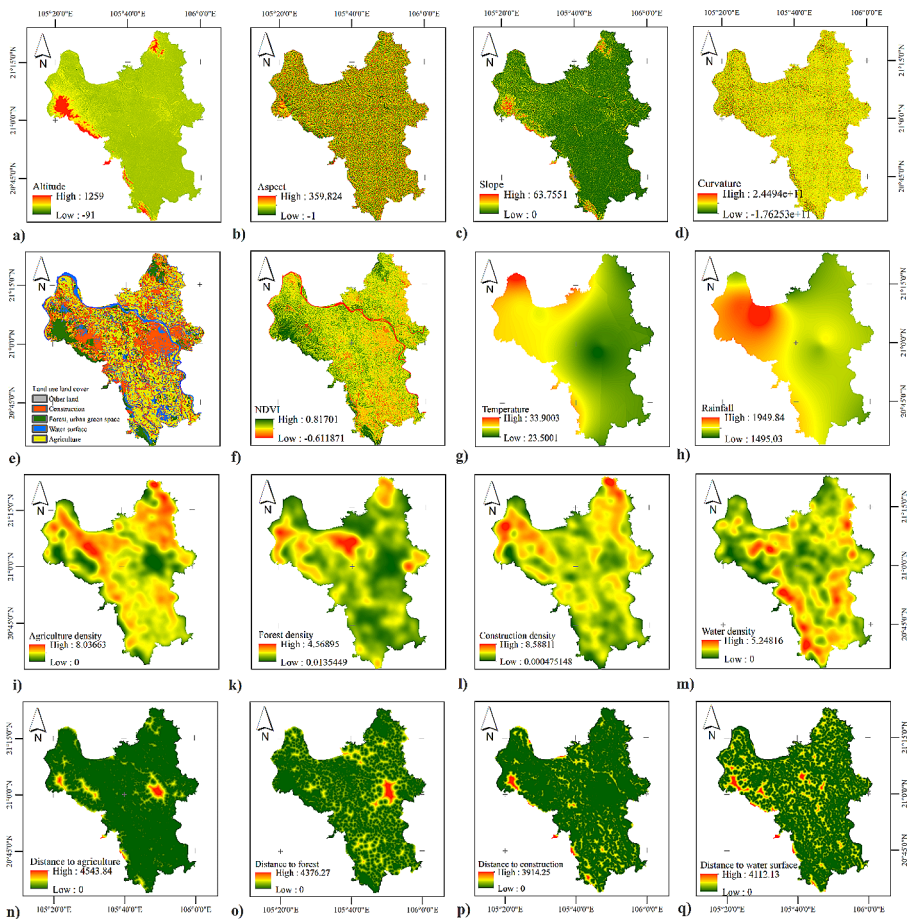


Fig. 5 Predictors of flood susceptibility in the study area. **a)** altitude; **b)** aspect; **c)** slope; **d)** curvature; **e)** LCLU; **f)** NDVI; **g)** temperature; **h)** rainfall; **i)** agriculture density; **k)** forest density; **l)** construction density; **m)** water density; **n)** distance to agriculture; **o)** distance to forest; **p)** distance to construction; and **q)** distance to water surface

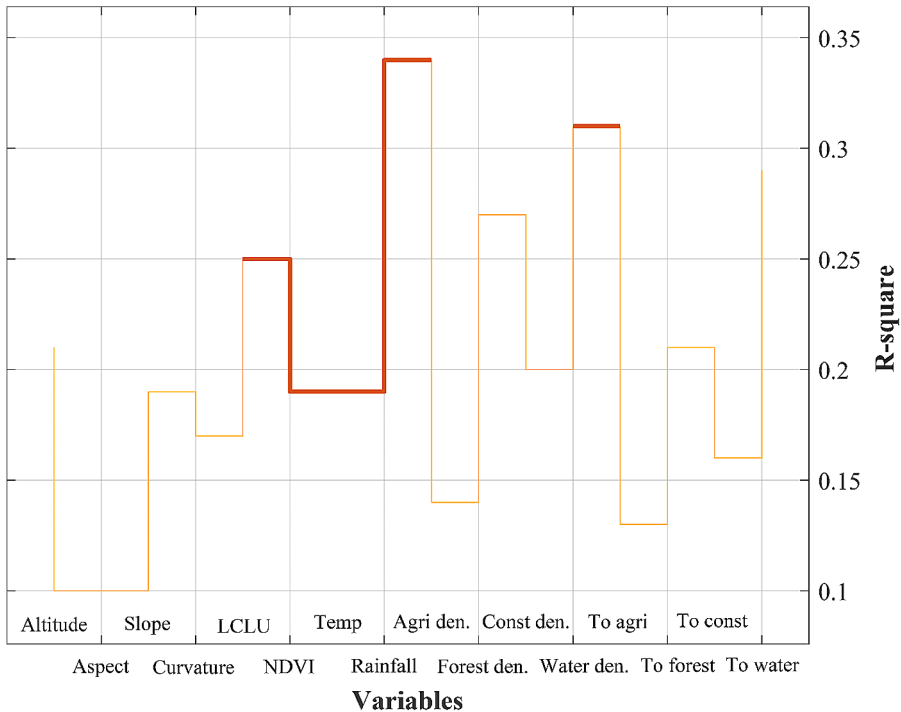


Fig. 6 The coefficient of determination between flood susceptibility and sixteen metrics applied from SPOT satellite

total area of 3,360 km², Large and abrupt increases in rainfall can result in flooding and give rise to various flood-related issues. Consequently, constructing a flood modeling approach sensitive to Hanoi requires comprehensive data on the intensity and distribution of rainfall within the region. This result underscores the significant role of rainfall in instigating floods and reaffirms its importance in the flood susceptibility prediction model within the research area. The variables exhibiting correlations greater than 0.25 include water density ($R^2=0.316$), distance to water surface ($R^2=0.290$), forest density ($R^2=0.274$), and LULC ($R^2=0.252$). These results indicate that LULC plays a pivotal role in identifying susceptible areas. The level of susceptible and susceptibility heavily relies on the interplay between land cover and rainfall. Forests and land use can influence watershed runoff, soil permeability, and water absorption capacity. Regions with a substantial forest cover can mitigate flood susceptible in the vicinity. Altitude, aspect, slope, curvature, NDVI, temperature, agriculture density, construction density, distance to agriculture, distance to forest, distance to construction, and distance to water surface possess R-square values ranging from 0.107 to 0.216 (Fig. 6). Although these variables exhibit a certain degree of correlation with flood susceptibility, their significance is not as pronounced as the aforementioned variables.

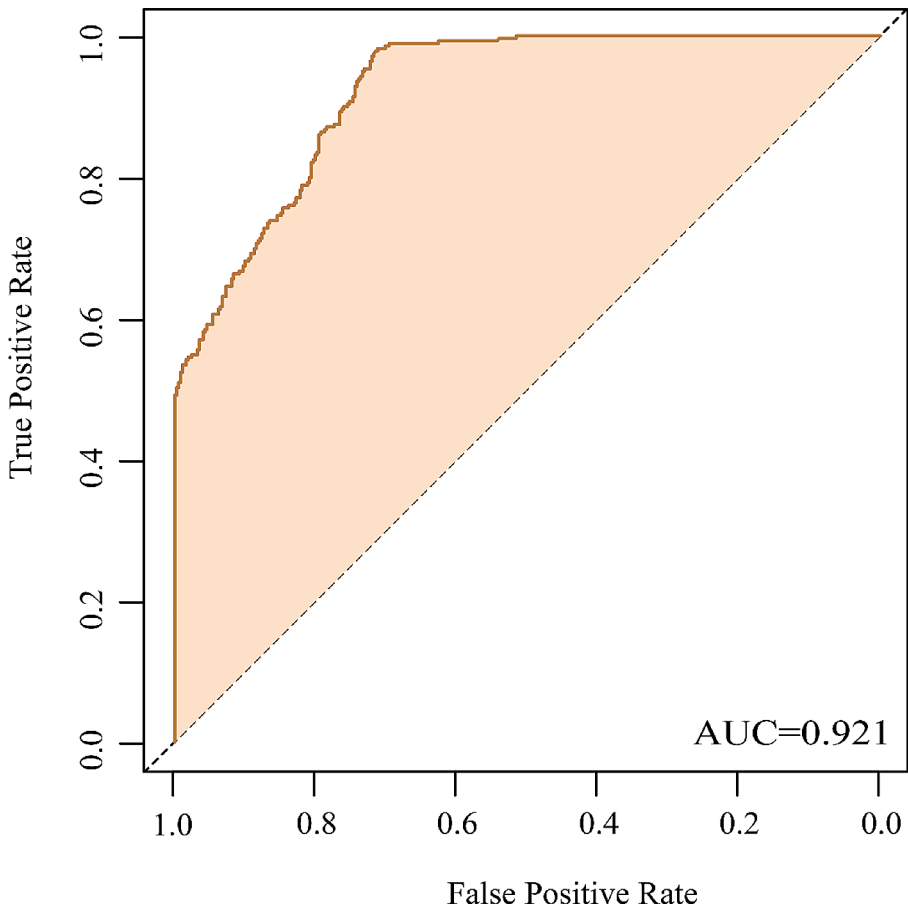


Fig. 7 The ROC plot shows the performance of the flood prediction model using the PCA-SVM model

Table 4 Performance of PCA-SVM on flood susceptibility in Hanoi city

Coefficient	SVM	PCA-SVM
R^2_{train}	0.830	0.938
R^2_{test}	0.794	0.904
RMSE	8.801	5.752
AUC	0.752	0.921

3.3 Mapping flood-susceptible areas

To evaluate the performance of the flood sensitivity prediction model, this study utilized the R^2 , RMSE, and ROC curve metrics, as indicated in Fig. 7; Table 4. The R^2_{train} value of 0.938 signifies that the model accounts for 93.8% of the variability in the training data, suggesting a strong fit to the training data (Table 4). Additionally, the PCA-SVM model demonstrated commendable performance on the test data, achieving an R^2_{test} value of 0.904. The results revealed an AUC of 0.921 (Fig. 7),

signifying the model’s ability to accurately classify flood susceptibility at a rate of 92.1%. Therefore, the PCA-SVM model proves valuable for flood susceptibility prediction in the study area.

Drawing on the prediction results obtained from the PCA-SVM model, this current study was able to identify and depict the spatial distribution map of flood-susceptible areas in Hanoi city (Fig. 8). Generally, susceptible areas are predominantly found along major rivers, especially the Red River. Additionally, flood-susceptible areas are typically located in low-lying regions with inadequate drainage, high river density, and substantial rainfall. Table 5 presents the distribution of areas based on flood susceptible zones, with the very low and low-risk levels accounting for 6.278%

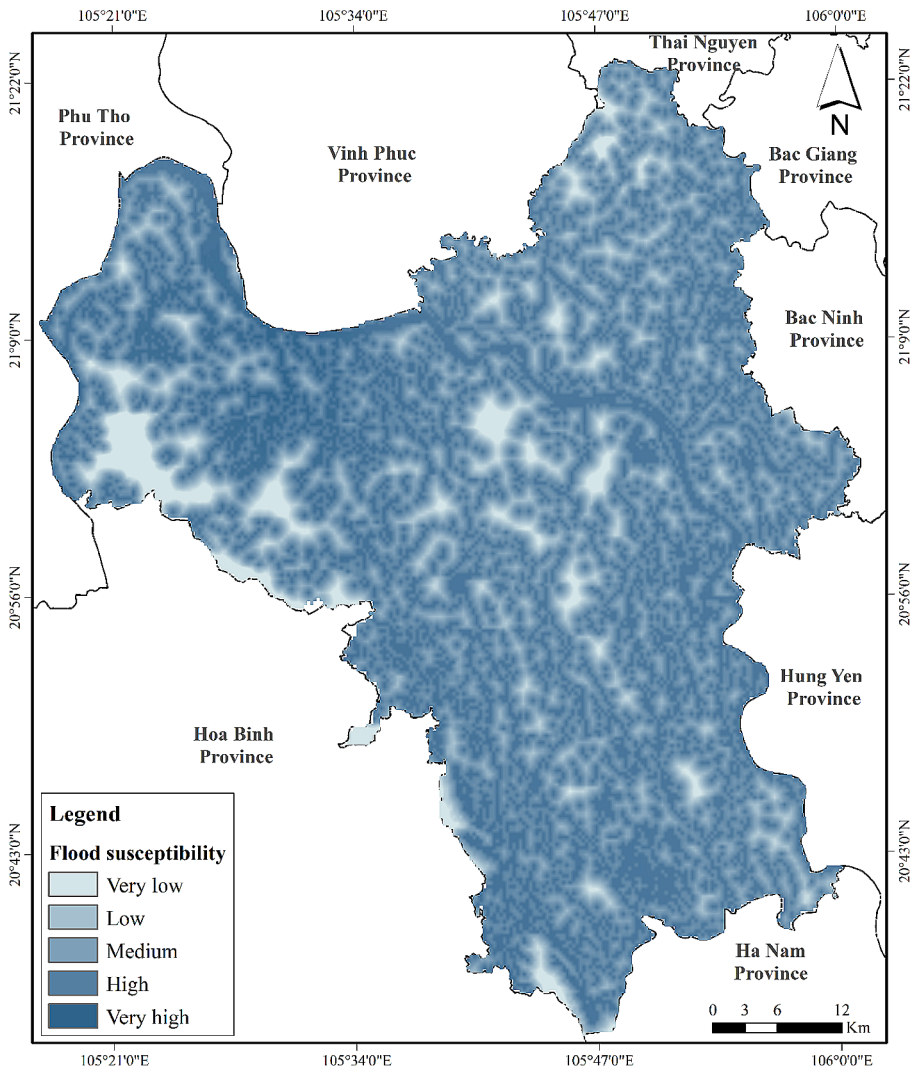


Fig. 8 Flood susceptibility map in the study area

Table 5 Area and percentage of each flood risk level in Hanoi city

Flood susceptibility	Area (ha)	Percentage (%)
Very low	21,093.256	6.278
Low	34,798.161	10.357
Medium	92,339.273	27.483
High	149,483.921	44.491
Very high	38,272.265	11.391

and 10.357% respectively, primarily concentrated in hilly areas with low river density (Fig. 5). The average flood susceptible level encompasses a substantial portion of approximately 27% of the total area (Table 5). In contrast, the highest proportion (over 55%) of high and above flood susceptible levels is concentrated in the delta region with high river density and heavy rainfall. Through the overlay of the flood classification layer beneath the agricultural land use layer (Fig. 9), this current study has successfully identified flooded agricultural land in areas with high to very high flood susceptible. Overall, more than 70% of agricultural land is situated in high-risk and above areas, indicating that flooding can result in inundation, crop loss, and a decline in agricultural quality and productivity.

4 Discussion

Floods are severe natural phenomena that annually cause substantial damage in Hanoi city annually (Do et al. 2022a, b, c; Anh 2021). Therefore, the analysis of flood susceptible contributes to the examination of areas at an unacceptable high susceptibility to flooding and the identification of locations for risk mitigation actions (Liu and Wu 2011; Costache 2019). Historically in Vietnam, the delimitation of flood zone heavily relied on hydro-meteorological monitoring data obtained from stations (Do et al. 2022a, b, c). However, these stations are widely spaced, with an average coverage of approximately 1,650 km² per station (Anh 2023). Therefore, the utilization of data from Earth observation satellites, which provide detailed information and have shorter repetitive cycles, has been regarded as a superior approach to complement the traditional monitoring methods facilitated by existing stations (Lee and Mohamad 2014; Lin et al. 2016).

To evaluate the impact and extent of floods on agriculture, the classification of LULC becomes essential (Ahmadlou et al. 2019). In the realm of multi-level tasks, deep learning algorithms have progressively outperformed traditional algorithms in terms of fast processing speed, and classification accuracy (Amitrano et al. 2018; Su et al. 2021; Do and Tran 2023b). Among them, the CRNN model has demonstrated remarkable effectiveness in LULC classification (Wu and Prasad 2017). The findings of the study unveil that the CRNN model attains favorable classification accuracy, with OA=88.005%, and CV=0.855 (Table 3). The research results indicate that the CRNN model achieves commendable classification accuracy, with an OA of 88.005% and CV of 0.855 (Table 3). Based on the computed results, agriculture encompasses nearly half of the total area (Table 2) and is mainly concentrated in low-lying and riverine areas. Nevertheless, the agricultural land area in Hanoi city is diminishing, as documented by studies conducted by Anh (2021, 2023). Despite being the largest

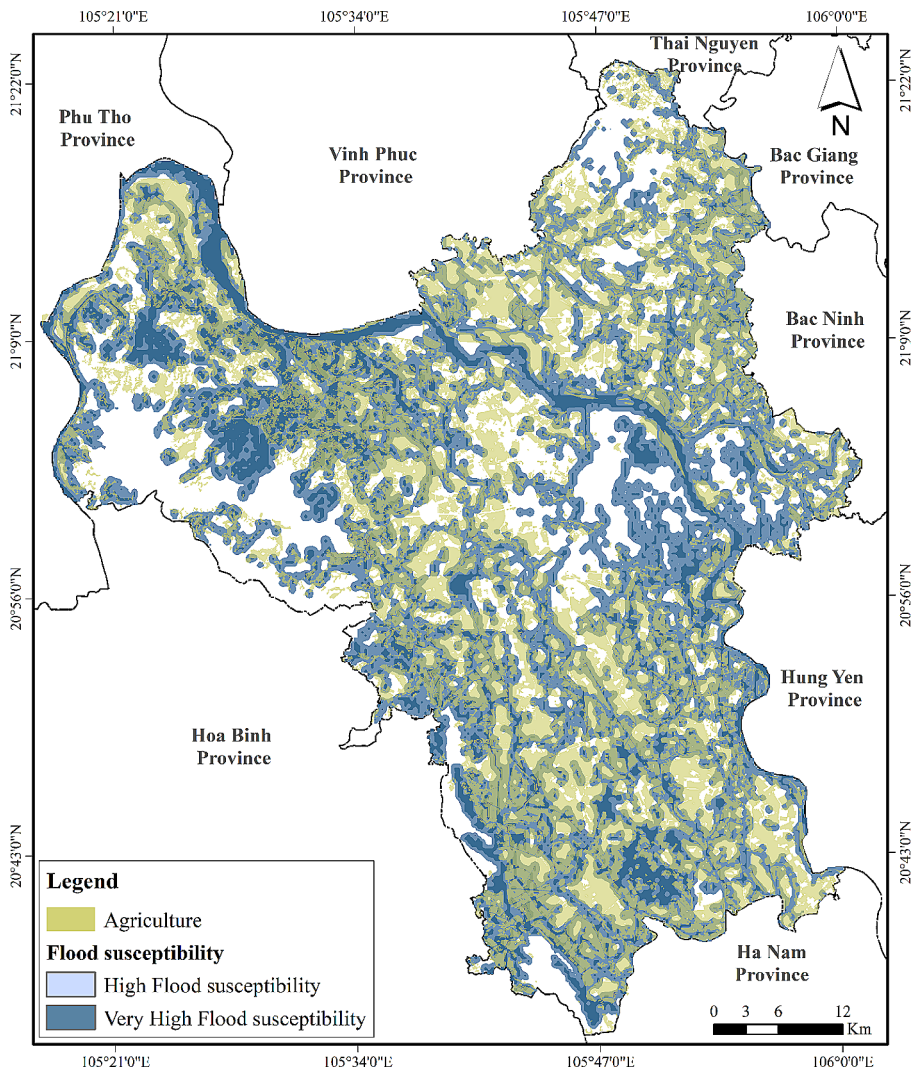


Fig. 9 Distribution of high and very high flood risk levels compared to agricultural land

city in Vietnam, Hanoi still retains a significant portion of agricultural land, which not only plays a pivotal role in food provision but also in conservation and sustainable development. Therefore, when flooding transpires, it can result in inundation, reduced quality, and productivity, as well as substantial economic losses.

To minimize the negative impacts of floods on the agricultural sector, numerous investigations have indicated the necessity for effective measures in flood prevention and response. This process begins with the assessment and classification of flood susceptible s (Liu et al. 2015; Loc et al. 2022; Do and Tran 2023a). Various studies have successfully conducted assessments of flood susceptible Hailin et al. (2009) employed multi-year average rainfall, storm rainfall days, terrain factors, and flood

frequency in order to map flood hazards. Similarly, Hagos et al. (2022) used GIS to identify susceptible areas by considering factors such as annual rainfall, slope, drainage systems, and soil type. However, the challenge lies in the multitude of factors that influence floods, encompassing both natural and socio-economic factors. Therefore, prior to incorporating all dependent variables into the prediction model, careful consideration of the various factors that impact flood zoning is necessary. The current study utilized the PCA method to examine the relationships between factors and reduce the dimensionality of the data (Fig. 4).

Figure 6 illustrates that rainfall is the most important factor leading to an increase in the quantity of surface water, which in turn affects flow transmission. Furthermore, the majority of Ha Noi city comprises areas with low slopes, flat terrain, and predominantly agricultural land cover in the downstream areas and main river branches within the city. The study conducted by Do et al. (2022a, b, c) also demonstrated that the aspect variable has minimal influence on flood occurrence. In the current study, less significant variables will be eliminated from the input model to ensure optimal model performance. Based on the prediction results from the PCA-SVM model, flood susceptible zoning maps at different levels are presented in Fig. 8. The areas with high to very high flood susceptible are primarily located in agricultural cultivation areas (Fig. 9). Conversely, low-risk susceptible areas are typically found in hilly regions with abundant forest cover and low river density. Similar conclusions were also drawn in studies by Zaharia et al. (2017); Do and Tran (2023a).

According to the study by Anh (2021), revealed that Hanoi city is renowned for its exceptional rice and vegetable production in the Red River Delta. The region yields an impressive amount of agriculture products, including 952.7 thousand tons of rice, 72.5 thousand tons of corn, and 723.2 thousand tons of vegetables, alongside other perennial crops. Consequently, any disruptions caused by flooding could have a significantly impact on the local food supply. To address this concern, a spatial distribution map has been developed to identify areas with varying flood sensitivity levels. This map serves as an early warning system for potential hazards and aids in assessing the likelihood of flood-related inundation affecting agricultural activities in Hanoi city. The results of this study provide valuable reference materials and effective support for decision-makers when planning land use for agriculture purposes.

5 Conclusions

The findings of the current study introduce a methodology for mapping flood sensitivity, which is highly relevant in distinguishing flood-affected agricultural land areas. Through the utilization of satellite imagery data and the PCA-SVM model, the study effectively depicts flood susceptible levels, ranging from high to very high, in approximately 70% of the agricultural regions. Furthermore, the study identifies susceptible areas primarily situated in riverine zones, highlighting their high susceptibility. that the efficacy of satellite imagery data in detecting and mapping flood sensitivity is evident. These findings significantly contribute to assisting decision-makers in pinpointing susceptible locations and formulating prevention and mitiga-

tion measures to minimizing the damages incurred by floods in the agriculture sector and its related industries.

Acknowledgements This research did not receive specific grants from public, commercial, or not-for-profit sector funding agencies. The author thanks anonymous reviewers for their helpful comments and suggestions, which improved this manuscript.

Declarations

Conflict of interest The author declares no conflict of interest.

References

- Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., Panahi, M.: Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int.* **34**, 1252–1272 (2019). <https://doi.org/10.1080/10106049.2018.1474276>
- Al-Abadi, A.M.: Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: A comparative study. *Arab. J. Geosci.* **11**, 218 (2018). <https://doi.org/10.1007/s12517-018-3584-5>
- Anh, D.T.N.: Application of machine learning in analyzing the impact of urbanization on the green space pattern in Hanoi City (Thesis). (2023)
- Anh, D.T.N.: Quantitative study of rice biomass in Hanoi city area using remote sensing and GIS technology (Thesis). (2021)
- Bhattacharya, B., Mazzoleni, M., Ugay, R.: Flood Inundation Mapping of the sparsely gauged large-scale Brahmaputra Basin using Remote Sensing products. *Remote Sens.* **11**, 501 (2019). <https://doi.org/10.3390/rs11050501>
- Brakenridge, G.R., Knox, J.C., Paylor, E.D., Magilligan, F.J.: Radar remote sensing aids study of the Great Flood of 1993. *Eos Trans. Am. Geophys. Union.* **75**, 521–527 (1994). <https://doi.org/10.1029/E0075i045p00521>
- Cao, C., Dragičević, S., Li, S.: Land-use change detection with convolutional neural network methods. *Environments.* **6**, 25 (2019)
- Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., Ahmad, B.B.: Flood susceptibility modeling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **575**, 864–873 (2019)
- Amitrano, D., Di Martino, G., Iodice, A., Riccio, D., Ruello, G.: *IEEE Trans. Geosci. Remote Sens.* **56**, 3290–3299 (2018). <https://doi.org/10.1109/TGRS.2018.2797536> Unsupervised Rapid Flood Mapping Using Sentinel-1 GRD SAR Images
- Do, A.N.T., 2024. Assessing the Impact of Habitat Fragmentation on the Distribution of Juvenile and Larval Sillago Species in the Ka Long Estuary Located in Northern Vietnam. *Ocean Sci. J.* **59**, 24. <https://doi.org/10.1007/s12601-024-00149-y>
- Do, A.N.T., Do, T.A.T., Van Pham, L., Tran, H.D., 2024. Assessment of the role of mangroves for *Periophthalmus modestus* applying machine learning and remote sensing: a case study in a large estuary from Vietnam. *Aquat. Ecol.* <https://doi.org/10.1007/s10452-024-10111-8>
- Do, A.N.T., Tran, H.D.: Application of deep learning in assessing the impact of flooding on the endangered freshwater fish *Neolissochilus benasi* (Cyprinidae) in a northern province of Vietnam. *Aquat. Ecol.* (2023a). <https://doi.org/10.1007/s10452-023-10056-4>
- Do, A.N.T., Tran, H.D., Ashley, M.: Employing a novel hybrid of GA-ANFIS model to predict distribution of whiting fish larvae and juveniles from tropical estuaries in the context of climate change. *Ecol. Inf.* **71**, 101780 (2022a). <https://doi.org/10.1016/j.ecoinf.2022.101780>
- Do, A.N.T., Tran, H.D., Ashley, M., Nguyen, A.T.: Monitoring landscape fragmentation and above-ground biomass estimation in Can Gio Mangrove Biosphere Reserve over the past 20 years. *Ecol. Inf.* **101743** (2022b). <https://doi.org/10.1016/j.ecoinf.2022.101743>

- Do, A.N.T., Tran, H.D.: Combining a deep learning model with an optimization algorithm to detect the dispersal of the early stages of spotted butterflyfish in northern Vietnam under global warming. *Ecol. Inf.* 102380 (2023b). <https://doi.org/10.1016/j.ecoinf.2023.102380>
- Do, A.N.T., Tran, H.D., Do, T.A.T.: Impacts of urbanization on heat in Ho Chi Minh, southern Vietnam using U-Net model and remote sensing. *Int. J. Environ. Sci. Technol.* (2023). <https://doi.org/10.1007/s13762-023-05118-x>
- Do, A.N.T., Tran, H.D.: Potential application of artificial neural networks for analyzing the occurrences of fish larvae and juveniles in an estuary in northern Vietnam. *Aquat. Ecol.* (2023c). <https://doi.org/10.1007/s10452-022-09959-5>
- Do, T.A.T., Do, A.N.T., Tran, H.D.: Quantifying the spatial pattern of urban expansion trends in the period 1987–2022c and identifying areas at risk of flooding due to the impact of urbanization in Lao Cai city. *Ecol. Inf.* 101912 (2022). <https://doi.org/10.1016/j.ecoinf.2022.101912>
- Gao, J.B., Gunn, S.R., Harris, C.J.: SVM regression through variational methods and its sequential implementation. *Neurocomputing Support Vector Machines.* **55**, 151–167 (2003). [https://doi.org/10.1016/S0925-2312\(03\)00365-5](https://doi.org/10.1016/S0925-2312(03)00365-5)
- Hagos, Y.G., Andualem, T.G., Yibeltal, M., Mengie, M.A.: Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in upper Awash River basin, Ethiopia. *Appl. Water Sci.* **12**, 148 (2022). <https://doi.org/10.1007/s13201-022-01674-8>
- Hailin, Z., Yi, J., Xuesong, Z., Gaoliao, J., Yi, Y., Baoyin, H.: GIS-based risk assessment for regional flood disaster, in: 2009 International Conference on Environmental Science and Information Application Technology. *IEEE*, pp. 564–567. (2009)
- Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., Kimoto, M.: Global projections of changing risks of floods and droughts in a changing climate. *Hydrol. Sci. J.* **53**, 754–772 (2008)
- Hsu, C.-Y., Li, W.: Learning from Counting. Leveraging Temporal Classification for Weakly Supervised Object Localization and Detection (2021)
- Hua, A.K.: Land Use Land Cover changes in detection of Water Quality: A study based on Remote Sensing and Multivariate statistics. *J. Environ. Public. Health.* **2017**(e7515130) (2017). <https://doi.org/10.1155/2017/7515130>
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S.: Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm Remote Sens.* **173**, 24–49 (2021)
- Khan, T.A., Shahid, Z., Alam, M., Su'ud, M.M., Kadir, K.: Early flood risk assessment using machine learning: A comparative study of svm, q-svm, k-nn and lda, in: 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS). *IEEE*, pp. 1–7. (2019)
- Klemas, V.: Remote sensing of floods and flood-prone areas: An overview. *J. Coast Res.* **31**, 1005–1013 (2015)
- Lee, W.-K., Mohamad, I.N.: Flood Economy Appraisal: An overview of the Malaysian scenario. In: Hassan, R., Yusoff, M., Ismail, Z., Amin, N.M., Fadzil, M.A. (eds.) *InCIEC 2013*, pp. 263–274. Springer Singapore, Singapore (2014). https://doi.org/10.1007/978-981-4585-02-6_23
- Lin, B., Wicks, J.M., Falconer, R.A., Adams, K.: Integrating 1D and 2D hydrodynamic models for flood simulation. *Proc. Inst. Civ. Eng. - Water Manag.* **159**, 19–25 (2006). <https://doi.org/10.1680/wama.2006.159.1.19>
- Lin, L., Di, L., Yu, E.G., Kang, L., Shrestha, R., Rahman, M.S., Tang, J., Deng, M., Sun, Z., Zhang, C.: A review of remote sensing in flood assessment, in: 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics). *IEEE*, pp. 1–4. (2016)
- Liu, Q., Qin, Y., Zhang, Y., Li, Z.: A coupled 1D–2D hydrodynamic model for flood simulation in flood detention basin. *Nat. Hazards.* **75**, 1303–1325 (2015). <https://doi.org/10.1007/s11069-014-1373-3>
- Liu, W.-C., Wu, C.-Y.: Flash flood routing modeling for levee-breaks and overbank flows due to typhoon events in a complicated river system. *Nat. Hazards.* **58**, 1057–1076 (2011). <https://doi.org/10.1007/s11069-010-9711-6>
- Loc, N.D., Yin, C.T., Mei, C.H., Van, H.T., Phuong, T.T., Son, N.T., Anh, D.T.N., Huong, B.T., Huy, N.Q.: Flood susceptibility assessment in central region of Vietnam based on ensemble fuzzy logic and multivariate analysis. *Urban Arhit Construcții.* **13**, 29–44 (2022)
- Moharram, M.A., Sundaram, D.M.: Land Use and Land Cover Classification with Hyperspectral Data: A comprehensive review of methods, challenges and future directions. *Neurocomputing.* (2023)
- Mosavi, A., Ozturk, P., Chau, K.: Flood prediction using machine learning models: Literature review. *Water.* **10**, 1536 (2018)
- Mou, L., Ghamisi, P., Zhu, X.X.: Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **55**, 3639–3655 (2017)

- Nahuelhual, L., Carmona, A., Lara, A., Echeverría, C., González, M.E.: Land-cover change to forest plantations: Proximate causes and implications for the landscape in south-central Chile. *Landsc. Urban Plan.* **107**, 12–20 (2012)
- Peng, J., Peng, F.-L.: A GIS-based evaluation method of underground space resources for urban spatial planning: Part 1 methodology. *Tunn. Undergr. Space Technol.* **74**, 82–95 (2018)
- Pham, T.V., Do, T.A.T., Tran, H.D., Do, A.N.T.: Assessing the impact of ecological security and forest fire susceptibility on carbon stocks in Bo Trach district, Quang Binh province. *Vietnam Ecol. Inf.* **74**, 101962 (2023a). <https://doi.org/10.1016/j.ecoinf.2022.101962>
- Pham, T.V., Thi Do, T.A., Tran, H.D., Thi Do, A.N.: Assessing groundwater potential for mitigating salinity issues in agricultural areas of southern Dong Nai Province. *Vietnam Groundw. Sustain. Dev.* **25**, 101177 (2024). <https://doi.org/10.1016/j.gsd.2024.101177>
- Pham, V.T., Do, T.A.T., Tran, H.D., Do, A.N.T.: Classifying forest cover and mapping forest fire susceptibility in Dak Nong Province, Vietnam utilizing remote sensing and machine learning. *Ecol. Inf.* **102392** (2023b). <https://doi.org/10.1016/j.ecoinf.2023.102392>
- Rajendran, G.B., Kumarasamy, U.M., Zarro, C., Divakarachari, P.B., Ullo, S.L.: Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images. *Remote Sens.* **12**, 4135 (2020)
- Costache, R.: Water Resour. Manag. **33**, 3239–3256 (2019). <https://doi.org/10.1007/s11269-019-02301-z>
Flood Susceptibility Assessment by Using Bivariate Statistics and Machine Learning Models - A Useful Tool for Flood Risk Management
- Su, X., Shao, W., Liu, J., Jiang, Y., Wang, K.: Dynamic assessment of the impact of flood disaster on economy and population under extreme rainstorm events. *Remote Sens.* **13**, 3924 (2021)
- Wu, H., Prasad, S.: Convolutional recurrent neural networks for hyperspectral data classification. *Remote Sens.* **9**, 298 (2017)
- Xu, X., Wang, X.: An Adaptive Network Intrusion Detection Method Based on PCA and Support Vector machines. In: Li, X., Wang, S., Dong, Z.Y. (eds.) *Advanced Data Mining and Applications. Lecture Notes in Computer Science*, pp. 696–703. Springer Berlin Heidelberg, Berlin, Heidelberg (2005). https://doi.org/10.1007/11527503_82
- Zaharia, L., Costache, R., Prăvălie, R., Ioana-Toroimac, G.: Mapping flood and flooding potential indices: A methodological approach to identifying areas susceptible to flood and flooding risk. Case study: The Prahova catchment (Romania). *Front. Earth Sci.* **11**, 229–247 (2017)
- Zhao, P., Zettsu, K.: Convolution recurrent neural networks for short-term prediction of atmospheric sensing data, in: 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, pp. 815–821. (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.