

Predicting land use effects on flood susceptibility using machine learning and remote sensing in coastal Vietnam

Van Tich Vu^a, Huu Duy Nguyen^{a,*}, Phuong Lan Vu^a, Minh Cuong Ha^b, Van Dong Bui^a, Thi Oanh Nguyen^c, Van Hiep Hoang^c and Thanh Kim Hue Nguyen^d

^aUniversity of Science, Vietnam National University, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

^bVNU – University of Engineering and Technology, Vietnam National University, 144 Xuan Thuy str., Hanoi, Vietnam

^cVNU – School of Interdisciplinary Sciences, Vietnam National University, 144 Xuan Thuy str., Hanoi, Vietnam

^dCenter for Water Resources Warning and Forecast – National center for water resources Planning and Investigation

*Corresponding author. E-mail: nguyenhuuduy@hus.edu.vn

ABSTRACT

Flood damage is becoming increasingly severe in the context of climate change and changes in land use. Assessing the effects of these changes on floods is important, to help decision-makers and local authorities understand the causes of worsening floods and propose appropriate measures. The objective of this study was to evaluate the effects of climate and land use change on flood susceptibility in Thua Thien Hue province, Vietnam, using machine learning techniques (support vector machine (SVM) and random forest (RF)) and remote sensing. The machine learning models used a flood inventory including 1,864 flood locations and 11 conditional factors in 2017 and 2021, as the input data. The predictive capacity of the proposed models was assessed using the area under the curve (AUC), the root mean square error (RMSE), and the mean absolute error (MAE). Both proposed models were successful, with AUC values exceeding 0.95 in predicting the effects of climate and land use change on flood susceptibility. The RF model, with AUC = 0.98, outperformed the SVM model (AUC = 0.97). The areas most susceptible to flooding increased between 2017 and 2021 due to increased built-up area. The results of the study confirm machine learning's capacity to assess differences in flood susceptibility.

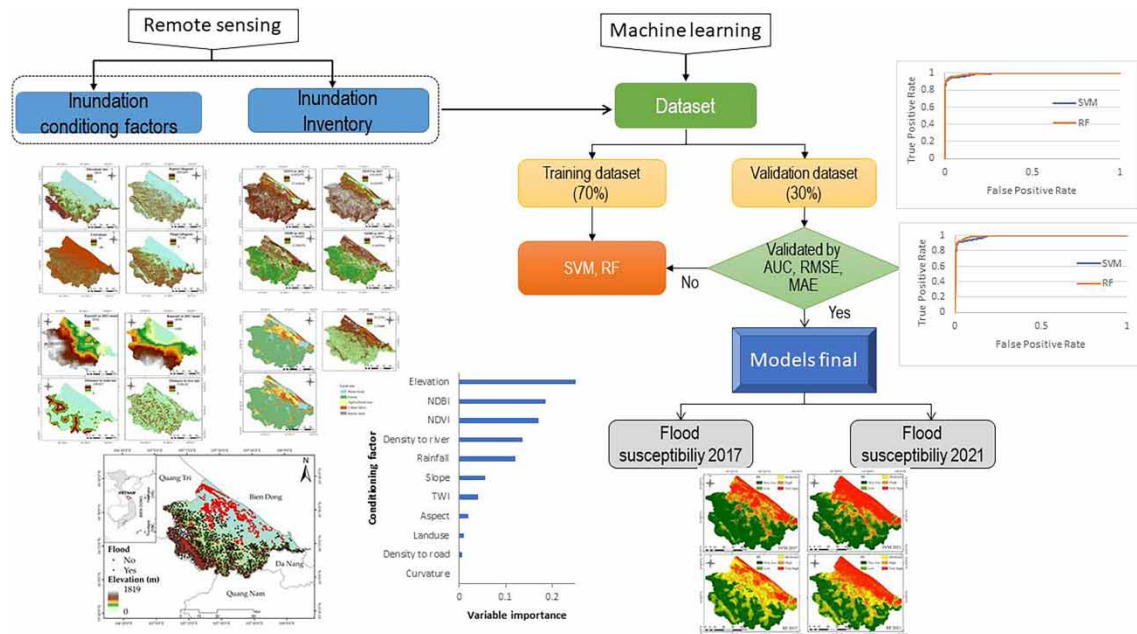
Key words: climate change, flood susceptibility, land use, machine learning

HIGHLIGHT

- Machine learning algorithms were applied for flood susceptibility modeling. The RF model had the highest AUC value (0.98) for areas highly susceptible to flooding, which increased between 2017 and 2021.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

GRAPHICAL ABSTRACT



INTRODUCTION

Floods affect infrastructure and the environment, not just the social economy (Khosravi *et al.* 2019; Islam *et al.* 2021). Between 1995 and 2015, approximately 150,000 flood events happened worldwide and caused around 157,000 deaths (Islam *et al.* 2021). Several studies point out that flooding affects approximately 250 million people and cause USD 40 billion in losses each year worldwide (Nguyen 2022a; Nguyen *et al.* 2022b).

Fluvial, flash, coastal, and urban floods can be classified on the nature of their occurrence. Floods are one of the most devastating natural events, causing damage to human life and infrastructure due to their onset attributes and rapid flow speeds (Islam *et al.* 2021). In developing countries, flood damage often has more impact than in developed countries due to climate change (Janizadeh *et al.* 2021). Climate change can lead to weather conditions outside predicted capacity: changes in precipitation and/or heat wave frequency and increases in the number and intensity of extreme weather events such as typhoons or floods (Janizadeh *et al.* 2021). Flood frequency change may be caused by changes in climatic conditions – e.g., precipitation and/or temperature – and may be apparent in low-lying, riverside, coastal, and urban areas. Flooding occurs not only under extreme climatic conditions but also through the effects of changing ecosystems. In the global scenario, human activities such as deforestation, dam construction, population, and/or urban growth can all change flood intensity (Junger *et al.* 2022; Pal *et al.* 2022; Roy *et al.* 2022). Since the ‘Doi Moi (Innovation)’ renewal policy in 1986, Vietnam has embarked on sustained urban growth. This occurs more rapidly in coastal regions, where it is often affected by typhoons or flooding. Therefore, improving flood risk assessment using state-of-the-art methods can reduce flood risk (Nguyen *et al.* 2022a).

Four factors, namely topographic, climatic, hydrologic, and human activity cause floods. Topographic factors include elevation, slope, curvature, and aspect, which are relatively stable (Ghosh *et al.* 2022; Prasad *et al.* 2022). Climatic and hydrologic factors, precipitation, and drainage networks are dynamic factors because they change every year. Chen *et al.* (2022) report that climate change, particularly the modification of precipitation, significantly influences the probability of flooding occurring and the speed of flow. At the same time, human activity like land use and land cover change (LULC), road distance, and distance to settlement are considered strong dynamic factors, especially in LULC. Several studies note that human activity can lead to changes in the flow and speed of floods, with the risk of flooding becoming more severe. Deforestation can influence the mechanical disturbance capacity of soils. Changes in the type of land cover have an important effect on the soil’s infiltration capacity (Nguyen *et al.* 2022b; Sugianto *et al.* 2022). All of this can influence the likelihood of flood occurrence, so assessing the effects of climate change and LULC on flood susceptibility is a critical task in understanding the

causes of change in flood susceptibility, which can support decision-makers in proposing appropriate measures to reduce flood damage.

Previous studies using different methods to assess the effects of climate change and LULC on flood susceptibility can be grouped into four main approaches: physical models, remote sensing, and statistical and machine learning models. Physical models include Mike Flood (Mangukiya & Yadav 2022), HEC-RAS (Afzal *et al.* 2022; Shah *et al.* 2022). Soil and Water Assessment Tool (SWAT) (Liu *et al.* 2022) is considered promising as it can simulate floods with details such as flow velocity, and flood depth and duration. However, physical models require deep understanding of hydrologic processes to adjust parameters, which poses challenges for researchers in model development. These models also require detailed *in situ* data, limiting their application to data-rich regions. Over past decades, remote sensing data, including synthetic aperture radar (SAR) data and optical sensor images combined with GIS, have been used to assess floods (Dewan *et al.* 2006; Alarifi *et al.* 2022). These methods have been used to build flood occurrence lists. Many different studies use remote sensing databases to assess floods hazards, combined with statistical techniques such as frequency ratio (FR), logistic regression (LR), weights of evidence (WoE model), fuzzy logic (FL) (Nandi *et al.* 2016; Rahmati *et al.* 2016; Sepehri *et al.* 2020; Ullah & Zhang 2020). These techniques can process spatial data to predict the likelihood of flood occurrence with high accuracy.

Several researchers have applied machine learning to develop flood susceptibility models. These include artificial neural networks (ANNs), random forest (RF), support vector machine (SVM), LR, and decision trees (DT) (Tehrany *et al.* 2014; Al-Juaidi *et al.* 2018; Chen *et al.* 2020; Khoirunisa *et al.* 2021; Wang *et al.* 2021). However, flood susceptibility model development using machine learning has several challenges, like selecting appropriately from the hundreds of different models and each model providing different results. Therefore, testing the models to assess flood susceptibility with different regions and types of natural hazards is recommended.

This study's objective was to test and compare machine learning models in assessing flood sensitivity in the context of climate change and land use in Thua Thien Hue province, Vietnam. This is the first time flooding has been assessed in relation to climate change and LULC in the province, which is one of those most affected by climate change and LULC in Vietnam. The key contributions of this study are: (i) selection of appropriate models to determine flood susceptibility zones; (ii) construction of a flood susceptibility map at different periods and (iii) evaluation of the effects of climate change and LULC on flood susceptibility.

DATA USE AND METHODS

Study area

Thua Thien Hue province is in Vietnam's central coastal region and covers approximately 5,000 km² (Figure 1). The province's topography includes mountains, midlands, and plains. Mountainous regions comprise over 50% of the total, with altitudes ranging from 500 to 1,819 m. The midlands, ranging from 20 to 200 m, with some peaks reaching 400 m, occupy 33% of the area. The plains cover 10% of the total, with altitudes between 0 and 20 m.

The study area has a dense hydrologic system, comprising the rivers Perfume, Bo, and Ong Lau. The province also has two lagoons, Tam Giang – Cau Hai and Lang Co. The study area's climate has two main seasons: rainy and dry. The rainy season, from September to December, accounts for 70% of annual precipitation. Rainfall during this season is often brief but intense, leading to floods that can cause significant socio-economic and environmental damage.

Thua Thien Hue province is frequently affected by storms. According to the Ministry of Natural Resources and Environment, the study area is impacted by approximately 30% of the storms that occur in the country.

Flood inventory

Flood inventory maps are vital in developing flood susceptibility models. They present records of past flood events, and can be used to predict the probability of flood occurrence in future by evaluating the relationships between past events and conditioning factors (Ghosh *et al.* 2022; Yaseen *et al.* 2022). In this study a flood inventory map was constructed using flood marks measured in 2020 and 2021. Sentinel 1A images from October 16, 2020 were also used to determine flood points, to ensure the flood susceptibility model's reliability. In total, 986 flood points were collected to build the model. Because the proposed model is classified as binary, non-flooding points must also be collected, for example, in the high elevation and slope regions, never affected by floods.

In the end, 1,864 flood and non-flood points were divided into two groups: 70% as training data to train the models and 30% as validation data to evaluate the models.

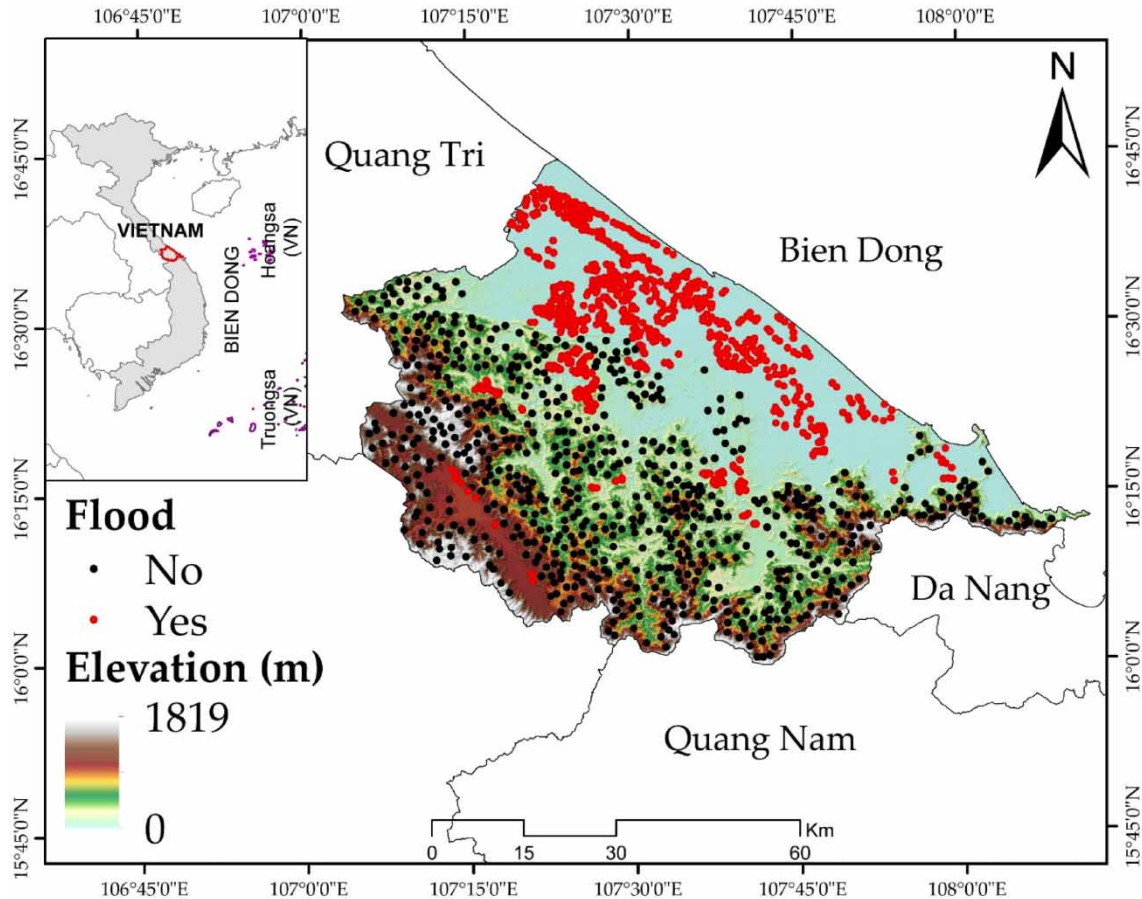


Figure 1 | Study area in Thua Thien Hue province.

Conditioning factors

Selection of the conditioning factors that best explain the flooding is vital to obtain accurate spatial estimates of flood susceptibility. The 11 conditioning factors from 2017 to 2021 were used in this study – elevation, slope, aspect, curvature, normalized difference built-up index (NDBI), normalized difference vegetation index (NDVI), land use, distance to river, distance to road, topographic wetness index (TWI), and rainfall – as input data for the proposed models (Figure 2, Table 1).

Elevation, aspect, curvature, slope, and TWI were calculated using a DEM generated from the 1:50,000 scale topographic map (available from Vietnam Department of Survey, Mapping and Geographic information (<https://www.bandovn.vn/en/topographic-maps-7>)). NDVI and NDBI in 2017 and 2021 were calculated from Landsat 08 images. Land use in 2017 and 2021 were downloaded from <https://www.arcgis.com>. Distance to river and distance to road were extracted from the 1:50,000 scale topographic map using Line Density in Arcgis 10.6. Rainfall in 2017 and 2021 were downloaded from <https://chrsdata.eng.uci.edu/>. (All factors were transformed into raster format at 10 m resolution.)

METHODS

Assessment of land cover and climate change effects on flood susceptibility has the following four stages: (i) data collection and pre-processing; (ii) SVM and RF model building; (iii) assessment of the proposed models' fit capacity; and (iv) analysis of the outcomes (Figure 3).

- (i) Data in this study include historical flood points and conditioning factors. Past flood points were collected from the field mission in 2020 and 2021, and extracted from the Sentinel 1A images in October 2020.
- (ii) The SVM and RF models were coded using the Python platform with the TensorFlow library. The hardest part is determining the model hyperparameters in SVM (C and Gamma) and RF (number of trees).

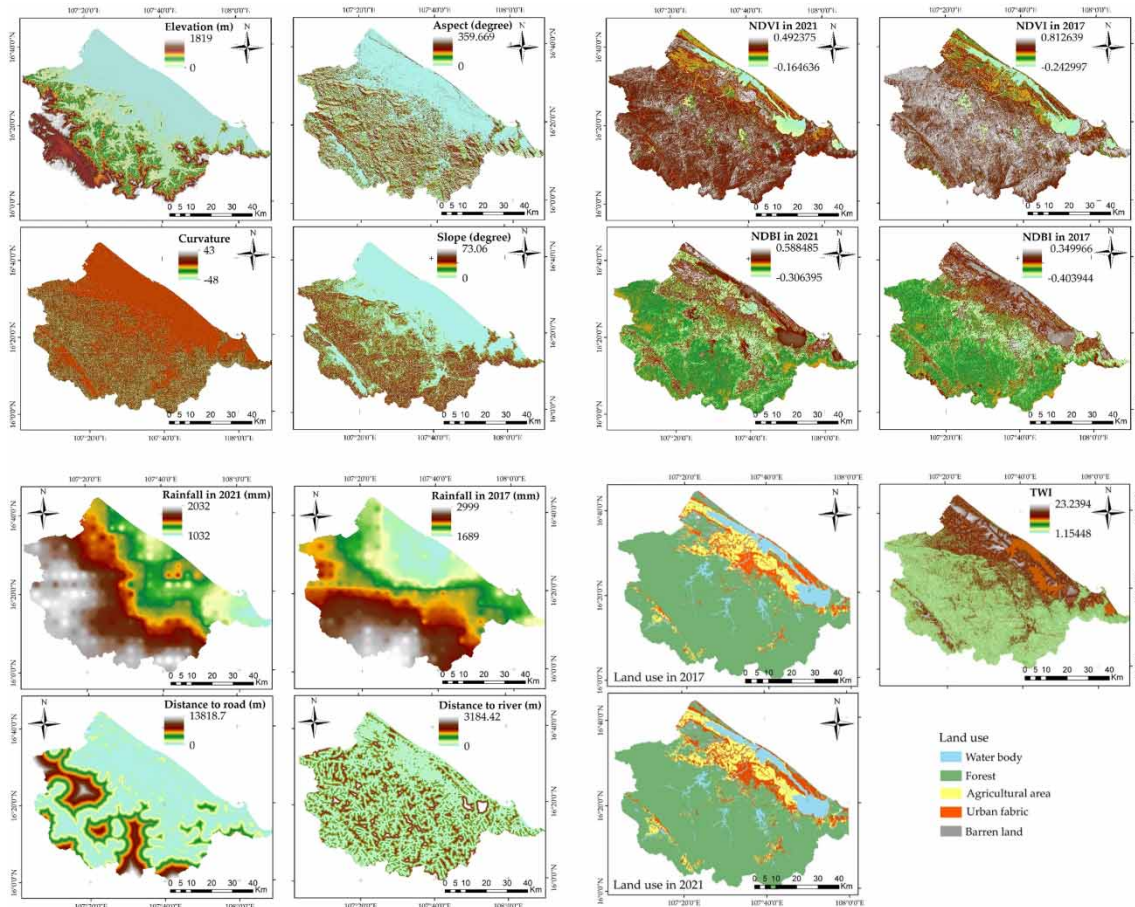


Figure 2 | Conditioning factors for the flood susceptibility models.

Hyperparameter selection for these models is based mainly on trial and error. The final values of C and Gamma in SVM were 10 and 0.2, respectively, and 200 trees were used in RF.

- (iii) Model evaluation: Several statistical indices – AUC, RMSE, and MAE – were used to evaluate model performance.
- (iv) Analysis: after validation, SVM and RF were used to build the 2017 and 2021 flood susceptibility maps. The model output values vary from 0 to 1 and have been grouped into five classes: very low, low, moderate, high and very high – using the natural break method. Notably, the range of values between types in 2017 and 2021 is similar, allowing for the assessment of changes in flood susceptibility over time.

Support vector machine

SVM was developed by Vapnik Research in 1965 but became an algorithm in 1995 (Cortes & Vapnik 1995). SVMs are initially linear and binary, based on maximizing the separation margin between two classes to predict a binary qualitative variable. Kernel functions, however, mean that transformations are now possible for multi-class classification. SVM's basic principle is to divide a dataset into two subsets by a separating hyperplane (Tehrany *et al.* 2014, 2015). The goal is to determine an optimal hyperplane to maximize the distance to the nearest training data (known as support vectors). This distance is the 'margin'.

The accuracy of SVM depends on the setting of C and Gamma. The higher the value of C, the more precise the data point separation. On the other hand, if the value of Gamma is significant, only points near the dividing line will be used in calculation.

Random Forest

RF, developed by (Breiman 2001), can solve both classification and regression problems, and is a combination of a bagging algorithm and random subspace. RF works by constructing many DT following three main steps:

Table 1 | Flood susceptibility variables and meanings used in this study

| Variable | Meaning | Source |
|-------------------|--|--|
| Elevation | The elevation of an area directly affects its ability to accumulate water, and there is an inverse relation between elevation and flooding. Elevations in the province are between 0 and 1,819 m. | Choubin <i>et al.</i> (2019), Nachappa <i>et al.</i> (2020) |
| Slope | Slope controls the flow speed and capacity to accumulate water. In the study area, the slope varies from 0° to 73°. | Yariyan <i>et al.</i> (2020) |
| Aspect | Aspect influences flow rate and soil moisture directly, and the occurrence of flooding indirectly. The aspect varies from 0° to 360° | Mojaddadi <i>et al.</i> (2017) |
| Curvature | Curvature has an essential influence on a region's heterogeneity. The curvature varies from -48 to 43. | Mirzaei <i>et al.</i> (2021) |
| NDVI | NDVI measures the area's vegetation density. Flooding affects regions with low vegetation density more because it influences soil infiltration strongly. In the study area, NDVI varied from -0.16 to 0.19 in 2021 and -0.24 to 0.81 in 2017. | Arabameri <i>et al.</i> (2022) |
| NDBI | NDBI represents a region's building density and is vital in flood susceptibility modeling. Regions with high construction density have low soil permeability, which influences flood occurrence strongly. In this study area, NDBI ranged from 0.3 to 0.58 in 2021 and -0.4 to 0.34 in 2017. | Vilasan & Kapse (2022) |
| Land use | Each type of land use has a different infiltration capacity, and the capacity of agricultural or forest surfaces is higher than construction surfaces. The extent of construction in the study area increased ever faster from 2017 to 2021, and is one of the causes of the increased flood risk. | Paule-Mercado <i>et al.</i> (2017) |
| Distance to river | The floods in the study area are fluvial, so areas near the river tend to be flooded more frequently. Distance to the river ranges from 0 to 3,184 m in the Thua Thien Hue province. | Vojtek & Vojteková (2019) |
| Distance to road | Distance to road has crucial effects on water infiltration capacity. In the study area, national road 1A was critical and led to flooding due to the obstruction of floodwater drainage to the sea. Distance from the road ranges from 0 to 13,818 m. | Nachappa <i>et al.</i> (2020) |
| Rainfall | Heavy rain in a short time leads to high water levels, with the low ground and dense hydrologic network causing flooding. Thua Thien Hue province has the highest mean annual precipitation in Vietnam. Rainfall ranged from 1,032 to 2,032 mm in 2021 and from 1,689 to 2,999 mm in 2017. | Saravanan & Abijith (2022) |
| TWI | TWI represents soil moisture content. In Thua Thien Hue province, the eastern region has the highest TWI | Nguyen (2022a) |

- (1) Using the bootstrap method to build a data subset of similar size to the original dataset.
- (2) Making the DT based on subsets.
- (3) Predicting the outcome by majority voting.

RF can handle multidimensional data and can also work well for missing data (Chen *et al.* 2020). Its performance depends on the number of decision trees (Ntree) and the number of variables used to build the tree (mtry).

Assessment model

Assessing the model's accuracy is a critical step in building a flood susceptibility model. If the model performs well, it can be generalized to assess flood susceptibility elsewhere. In this study, statistical indicators including receiver operating curve (ROC), RMSE, and MAE, were used to assess data learning and validation performance.

The ROC curve shows a model's ability to predict the occurrence of floods correctly and is based on the true positive rate (sensitivity) and false-positive rate (1-specificity). The true positive rate is the proportion of actual positives that are correctly identified as positives by the model, while the false-positive rate is the proportion of actual negatives that are incorrectly identified as positives by the model. The areas under the ROC curve reflect model accuracy.

RMSE and MAE measure the differences between the predicted and observed values (Wang *et al.* 2019; Nguyen 2022a).

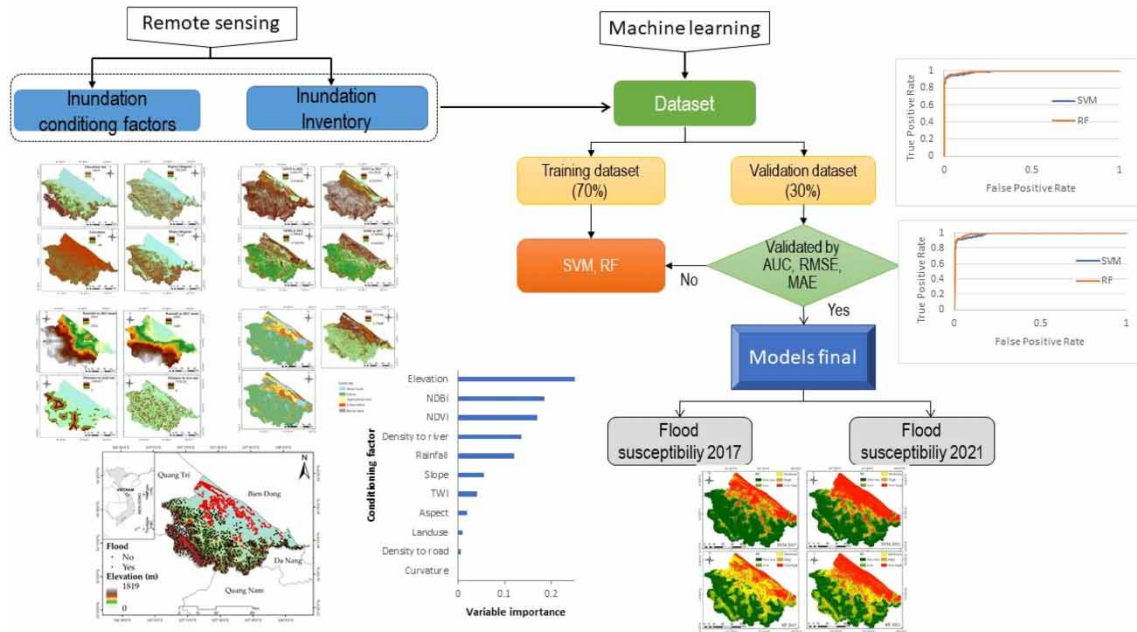


Figure 3 | Flood susceptibility methodology used in the study.

RESULTS

Land use change analysis

Figure 4 shows the land cover change from 2017 to 2021. Forested surfaces decreased by about 100 km² from 2017 to 2021, while urban areas increased by about 70 km², mainly in the city of Hue. At the same time, agricultural areas increased by approximately 50 km². Although land use change is not the root cause of flooding, it is a significant contributing factor that leads to increasingly dangerous floods. During field trips in 2021, residents reported that, since 2010, the speed of floods had increased. One of the reasons for this is believed to be urbanization.

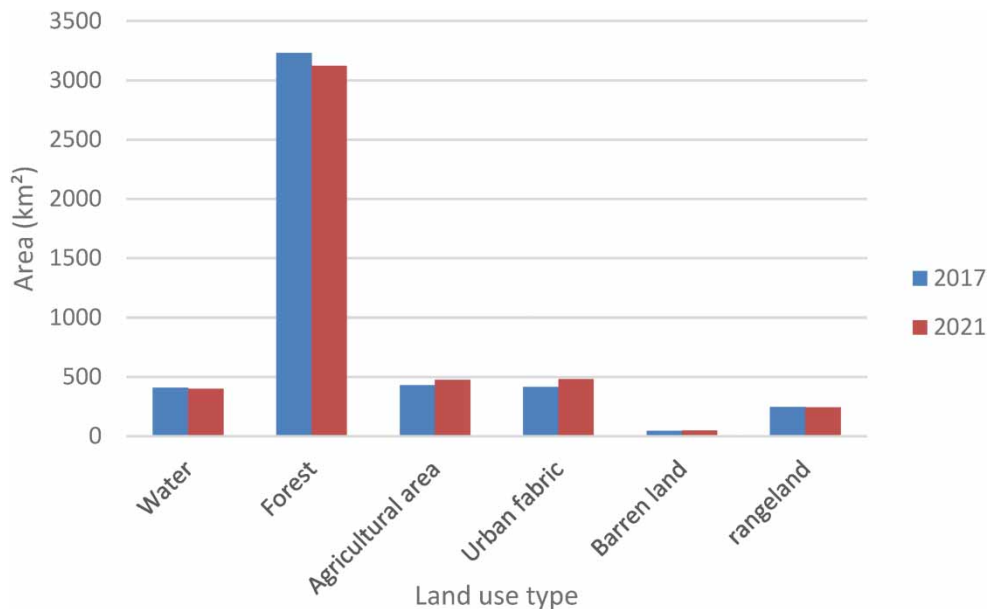


Figure 4 | Land use change from 2017 to 2021 in the study area.

Factor selection

Selecting appropriate factors for flood occurrence is essential as data quality determines the model’s buildability. RF was used in this study to evaluate conditioning factor importance. The results show that elevation, NDBI,

NDVI, distance to river, and rainfall are the most critical factors in the study area, followed by slope, TWI, aspect, land use, and road distance. The curvature factor's importance value is zero with respect to flood occurrence, so it was excluded from the flood susceptibility models (Figure 5).

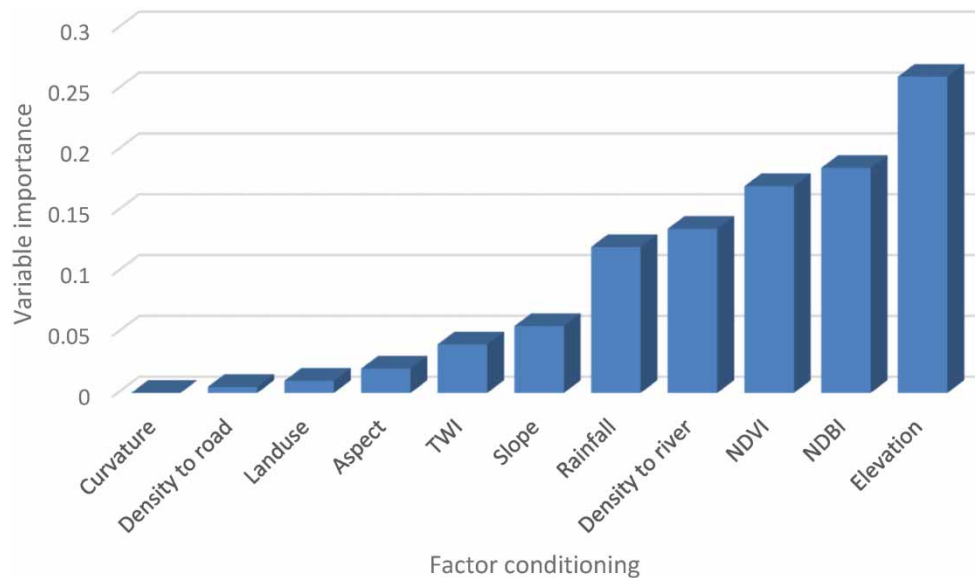


Figure 5 | Conditioning factor importance.

Model assessment

The models' predictive ability for flood susceptibility was evaluated and compared using ROC and AUC. Figure 6 shows the AUC results for the models. The RF model had an AUC value of 0.99 and outperformed the SVM model (AUC = 0.98) in training. In validation, the RF model's predictive capacity (AUC = 0.98) is also superior to that of the SVM model (AUC = 0.97).

Table 2 gives RMSE and MAE for the flood susceptibility models. Clearly, the RF model performs better than the SVM one in terms of both, with values of 0.21 and 0.06, respectively, compared to the SVM model's 0.22 and 0.08 during training. The same was true in validation, the RF model (RMSE = 0.22, MAE = 0.09) outperformed the SVM model (0.23 and 0.12).

Assessment effects of land use and rainfall on flood susceptibility

After validation, the models were used to produce flood susceptibility maps for 2017 and 2021 (Figure 7). Areas with very high flooding susceptibility increased between 2017 and 2021 in both models. The changes are concentrated mostly in the east, with low elevation and building density. The SVM model indicated that the area of very high flood susceptibility increased by about 25% between 2017 and 2021, while RF indicated that the increase was about 50%. On the other hand, the high susceptibility zone decreased by approximately 30% for the SVM model and 20% for the RF model. Areas of moderate susceptibility also decreased by about 25% (SVM) and 30% (RF). The low susceptibility areas increased slightly, by approximately 5–10% for each model (Table 3).

DISCUSSION

Although several different methods have been developed and used to predict the spatial distribution of flood susceptibility, assessment is still considered an obstacle to effective flood risk management (Gharakhanlou & Perez 2023; Wu *et al.* 2023). Floods are caused mainly by the distribution and intensity of precipitation and are often amplified by topographic conditions and landscape characteristics, including water infiltration capacity. Both climate and LULC are expected to influence flood susceptibility in the future.

The results of this study confirmed that altitude, NDBI, NDVI, distance from the river and rainfall are the most critical factors for the probability of flood occurrence in Thua Thien Hue province. Altitude is important because it is directly related to water storage capacity. In Thua Thien Hue province, floods occur mainly in the lower altitude, eastern region. The importance of NDBI and NDVI came second and third for flood susceptibility because

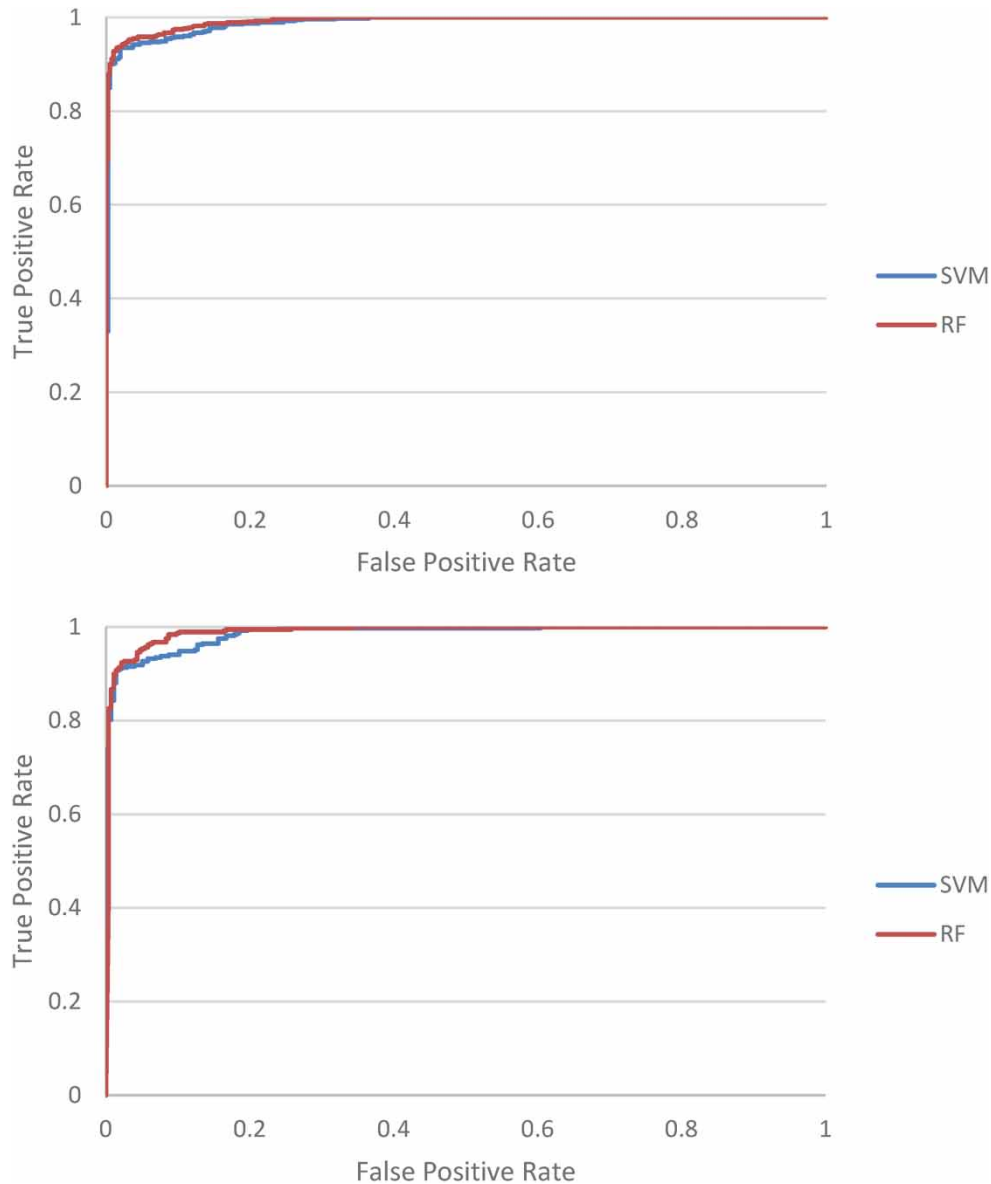


Figure 6 | Model validation with ROC and AUC.

Table 2 | SVM and RF model performance for flood susceptibility

| | Training | | | Validating | | |
|-----|----------|------|------|------------|------|------|
| | AUC | RMSE | MAE | AUC | RMSE | MAE |
| SVM | 0.98 | 0.22 | 0.08 | 0.97 | 0.23 | 0.12 |
| RF | 0.99 | 0.21 | 0.06 | 0.98 | 0.22 | 0.09 |

they influence water infiltration capacity directly. For example, in the study area, the increase in built-up area and decrease area of vegetation are thought to have changed the spatial distribution of floods from 2017 to 2019. Distance from the river was the fourth factor of importance for flood occurrence, because floods in the study area are fluvial, occurring near the river. Rain was only the fifth factor also because floods in the study area are fluvial, occurring at low altitude and due to LULC. Previous studies justify all this. (Nguyen 2022b) found that elevation and land use were critical factors in flood occurrence in Central Vietnam and Ha Tinh province. Ha *et al.* (2022) indicated that elevation, rainfall, and distance to river were the most important factors for flood susceptibility in Thua Thien Hue province, Vietnam.

| | | Very low (km ²) | Low (km ²) | Moderate (km ²) | High (km ²) | Very high (km ²) |
|-----|------|-----------------------------|------------------------|-----------------------------|-------------------------|------------------------------|
| SVM | 2017 | 2636 | 415 | 337 | 393 | 988 |
| | 2021 | 988 | 513 | 287 | 277 | 1436 |
| RF | 2017 | 1673 | 978 | 994 | 551 | 573 |
| | 2021 | 1578 | 915 | 657 | 476 | 1143 |

Figure 7 | Flood susceptibility maps for 2017 and 2021 from SVM and RF.

Table 3 | Distribution of five flood susceptibility classes from SVM and RF

| | | Very low (km ²) | Low (km ²) | Moderate (km ²) | High (km ²) | Very high (km ²) |
|-----|------|-----------------------------|------------------------|-----------------------------|-------------------------|------------------------------|
| SVM | 2017 | 2,636 | 415 | 337 | 393 | 988 |
| | 2021 | 988 | 513 | 287 | 277 | 1,436 |
| RF | 2017 | 1,673 | 978 | 994 | 551 | 573 |
| | 2021 | 1,578 | 915 | 657 | 476 | 1,143 |

In this study, two popular models, SVM and RF, were proposed to assess the effects of climate and LULC change on flood susceptibility in Thua Thien Hue province, Vietnam. The results show that with its AUC value of 0.98, the RF model outperformed the SVM model (AUC = 0.97) with respect to the effects of climate and LULC change on flood susceptibility. RF is one of the most effective and widely used, existing models, because it is the most easily explained (or interpreted) of the models. In addition, it is more transparent in relation to the use made of training data. RF also has the advantage of using all of its initial data more intelligently, in order to limit its errors (Pal 2005; Uddin & Uddiny 2015). While SVM has the advantage in solving overfitting problems, however, it is not suitable for nonlinear problems (Karamizadeh *et al.* 2014).

The models proposed in this study use RMSE as an objective function, because of which, the overfitting problem is seen as an important challenge, as it is for machine learning in general. To reduce the problem's effects, several techniques were applied in this study, such as setting a dropout rate and limiting the search range. However, other methods can also be tried, such as collecting training data from different geographical locations (Nguyen *et al.* 2020).

The study's results also confirmed that the very high flood susceptibility surfaces, concentrated mainly in the east where there were major changes in land use, increased from 2017 to 2021 in the study area. This trend is common worldwide. Nguyen *et al.* (2022a) show that flood susceptibility will increase in the Nhat Le – Kien Giang watershed (Vietnamese central region) will increase from 2005 to 2050 due to land use and climate change. Hagos *et al.* (2022) report that flooding will become more frequent due to climate change and land use change in the Awash River basin in central Ethiopia. Nirupama & Simonovic (2007) show the increase in flood risk in the province of Ontario in Canada due to urban growth. This was confirmed by the study by Hettiarachchi *et al.* (2018), who highlighted the increased flood risk in the South Washington Watershed District (SWWD) in Minnesota, United States, related to climate change and urban growth. This study provides additional evidence for the effects of climate change and change in land use on flood susceptibility.

CONCLUSIONS

The objective of this study was to evaluate flood susceptibility to the impact of climate change and LULC, using machine learning and remote sensing models in relation to Thua Thien Hue province, Vietnam.

The proposed models use 1,864 flood locations and 11 condition factors from 2017 to 2021 as input data. The results show that elevation, NDVI, NDBI, distance to river and rainfall were the most important factors for flood occurrence in the study area.

The combination of machine learning techniques (SVM and RF) with remote sensing successfully assessed the impact of climate change and LULC on flooding, with AUC values exceeding 0.95. The method can be generalized to determine flood susceptibility elsewhere, particularly in data-limited areas.

Very high flood susceptibility areas increased by about 30% from 2017 to 2021 in Thua Thien Hue. The changes are related to the increased construction area – the surfaces are concentrated mainly in the east – and precipitation.

FUNDING STATEMENT

This work is supported by national research program from MOST (Vietnam), project number ĐTDL.CN-93/21.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by V.T.V., H.D.N., P.L.V., M.C.H., V.D.B., T.O.N., V.H.H., T.K.H.N. The first draft of the manuscript was written by V.T.V. and H.D.N. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Afzal, M. A., Ali, S., Nazeer, A., Khan, M. I., Waqas, M. M., Aslam, R. A., Cheema, M. J. M., Nadeem, M., Saddique, N. & Muzammil, M. 2022 Flood inundation modeling by integrating HEC-RAS and satellite imagery: a case study of the Indus River Basin. *Water* **14**(19), 2984.
- Alarifi, S. S., Abdelkareem, M., Abdalla, F. & Alotaibi, M. 2022 Flash flood hazard mapping using remote sensing and GIS techniques in southwestern Saudi Arabia. *Sustainability* **14**(21), 14145.
- Al-Juaidi, A. E., Nassar, A. M. & Al-Juaidi, O. E. 2018 Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences* **11**(24), 1–10.
- Arabameri, A., Seyed Danesh, A., Santosh, M., Cerda, A., Chandra Pal, S., Ghorbanzadeh, O., Roy, P. & Chowdhuri, I. 2022 Flood susceptibility mapping using meta-heuristic algorithms. *Geomatics, Natural Hazards and Risk* **13**(1), 949–974.
- Breiman, L. 2001 Random forests. *Machine Learning* **45**(1), 5–32.
- Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S. & Pradhan, B. 2020 Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment* **701**, 134979.
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. & Mosavi, A. 2019 An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment* **651**, 2087–2096.
- Cortes, C. & Vapnik, V. 1995 Support vector machine. *Machine Learning* **20**(3), 273–297.
- Dewan, A. M., Kumamoto, T. & Nishigaki, M. 2006 Flood hazard delineation in greater Dhaka, Bangladesh using an integrated GIS and remote sensing approach. *Geocarto International* **21**(2), 33–38.
- Gharakhanlou, N. M. & Perez, L. 2023 Flood susceptible prediction through the use of geospatial variables and machine learning methods. *Journal of hydrology* 129121.
- Ghosh, S., Saha, S. & Bera, B. 2022 Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. *Natural Hazards Research*.
- Ha, M. C., Vu, P. L., Nguyen, H. D., Hoang, T. P., Dang, D. D., Dinh, T. B. H., Şerban, G., Rus, I. & Breţcan, P. 2022 Machine learning and remote sensing application for extreme climate evaluation: example of flood susceptibility in the Hue Province, Central Vietnam region. *Water* **14**(10), 1617.
- Hagos, Y. G., Andualem, T. G., Yibeltal, M. & Mengie, M. A. 2022 Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in upper Awash River basin, Ethiopia. *Applied Water Science* **12**(7), 148.
- Hettiarachchi, S., Wasko, C. & Sharma, A. 2018 Increase in flood risk resulting from climate change in a developed urban watershed—the role of storm temporal patterns. *Hydrology and Earth System Sciences* **22**(3), 2041–2056.
- Islam, A. R. M. T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A. & Linh, N. T. T. 2021 Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers* **12**(3), 101075.
- Janizadeh, S., Chandra Pal, S., Saha, A., Chowdhuri, I., Ahmadi, K., Mirzaei, S., Mosavi, A. H. & Tiefenbacher, J. P. 2021 Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. *Journal of Environmental Management* **298**, 113551.

- Junger, L., Hohensinner, S., Schroll, K., Wagner, K. & Seher, W. 2022 Land use in flood-prone areas and its significance for flood risk management – a case study of alpine regions in Austria. *Land* **11**(3), 392.
- Karamizadeh, S., Abdullah, S. M., Halimi, M., Shayan, J. & Javad Rajabi, M. 2014 Advantage and drawback of support vector machine functionality. In: *2014 International Conference on Computer, Communications, and Control Technology (14CT)*. IEEE, pp. 63–65.
- Khoirunisa, N., Ku, C.-Y. & Liu, C.-Y. 2021 A GIS-Based artificial neural network model for flood susceptibility assessment. *International Journal of Environmental Research and Public Health* **18**(3), 1072.
- Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G. & Ho, H. L. 2019 A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology* **573**, 311–323.
- Liu, Y., Xu, Y., Zhao, Y. & Long, Y. 2022 Using SWAT model to assess the impacts of land use and climate changes on flood in the upper Weihe river, China. *Water* **14**(13), 2098.
- Mangukiya, N. K. & Yadav, S. M. 2022 Integrating 1D and 2D hydrodynamic models for semi-arid river basin flood simulation. *International Journal of Hydrology Science and Technology* **14**(2), 206–228.
- Mirzaei, S., Vafakhah, M., Pradhan, B. & Alavi, S. J. 2021 Flood susceptibility assessment using extreme gradient boosting (EGB), Iran. *Earth Science Informatics* **14**, 51–67.
- Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N. & Ghazali, A. H. b. 2017 Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk* **8**(2), 1080–1102.
- Nachappa, T. G., Piralilou, S. T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O. & Blaschke, T. 2020 Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology* **125275**.
- Nandi, A., Mandal, A., Wilson, M. & Smith, D. 2016 Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environmental Earth Sciences* **75**, 1–16.
- Nguyen, H. D. 2022a GIS-based hybrid machine learning for flood susceptibility prediction in the Nhat Le–Kien Giang watershed, Vietnam. *Earth Science Informatics* 1–18.
- Nguyen, H. D. 2022b Spatial modeling of flood hazard using machine learning and GIS in Ha Tinh province, Vietnam. *Journal of Water and Climate Change*.
- Nguyen, H.-D., Pham, V.-D., Nguyen, Q.-H., Pham, V.-M., Pham, M. H., Vu, V. M. & Bui, Q.-T. 2020 An optimal search for neural network parameters using the Salp swarm optimization algorithm: a landslide application. *Remote Sensing Letters* **11**(4), 353–362.
- Nguyen, H. D., Nguyen, Q.-H., Du, Q. V. V., Nguyen, T. H. T., Nguyen, T. G. & Bui, Q.-T. 2021 A novel combination of deep neural network and manta ray foraging optimization for flood susceptibility mapping in Quang Ngai province, Vietnam. *Geocarto International* 1–25.
- Nguyen, H. D., Dang, D. K., Nguyen, Q.-H., Bui, Q.-T. & Petrisor, A.-I. 2022a Evaluating the effects of climate and land use change on the future flood susceptibility in the central region of Vietnam by integrating land change modeler, machine learning methods. *Geocarto International* 1–36.
- Nguyen, H. D., Quang-Thanh, B., Nguyen, Q.-H., Nguyen, T. G., Pham, L. T., Nguyen, X. L., Vu, P. L., Thanh Nguyen, T. H., Nguyen, A. T. & Petrisor, A.-I. 2022b A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. *Hydrological Sciences Journal* **67**(7), 1065–1083.
- Nirupama, N. & Simonovic, S. P. 2007 Increase of flood risk due to urbanisation: a Canadian example. *Natural Hazards* **40**(1), 25.
- Pal, M. 2005 Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* **26**(1), 217–222.
- Pal, S. C., Chowdhuri, I., Das, B., Chakraborty, R., Roy, P., Saha, A. & Shit, M. 2022 Threats of climate change and land use patterns enhance the susceptibility of future floods in India. *Journal of Environmental Management* **305**, 114317.
- Paule-Mercado, M., Lee, B., Memon, S., Umer, S., Salim, I. & Lee, C.-H. 2017 Influence of land development on stormwater runoff from a mixed land use and land cover catchment. *Science of the Total Environment* **599**, 2142–2155.
- Prasad, P., Loveson, V. J., Das, B. & Kotha, M. 2022 Novel ensemble machine learning models in flood susceptibility mapping. *Geocarto International* **37**(16), 4571–4593.
- Rahmati, O., Pourghasemi, H. R. & Zeinivand, H. 2016 Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International* **31**(1), 42–70.
- Roy, P. S., Ramachandran, R. M., Paul, O., Thakur, P. K., Ravan, S., Behera, M. D., Sarangi, C. & Kanawade, V. P. 2022 Anthropogenic land use and land cover changes – a review on its environmental consequences and climate change. *Journal of the Indian Society of Remote Sensing* **50**(8), 1615–1640.
- Saravanan, S. & Abijith, D. 2022 Flood susceptibility mapping of northeast coastal districts of Tamil Nadu India using multi-source geospatial data and machine learning techniques. *Geocarto International* 1–30.
- Sepelhi, M., Malekinezhad, H., Jahanbakhshi, F., Ildoromi, A. R., Chezgi, J., Ghorbanzadeh, O. & Naghipour, E. 2020 Integration of interval rough AHP and fuzzy logic for assessment of flood prone areas at the regional scale. *Acta Geophysica* **68**, 477–493.
- Shah, Z., Saraswat, A., Samal, D. R. & Patel, D. 2022 A single interface for rainfall-runoff simulation and flood assessment – a case of new capability of HEC-RAS for flood assessment and management. *Arabian Journal of Geosciences* **15**(18), 1526.
- Sugianto, S., Deli, A., Miswar, E., Rusdi, M. & Irham, M. 2022 The effect of land use and land cover changes on flood occurrence in Teunom Watershed, Aceh Jaya. *Land* **11**(8), 1271.

- Tehrany, M. S., Pradhan, B. & Jebur, M. N. 2014 Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology* **512**, 332–343.
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. 2015 Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment* **29**, 1149–1165.
- Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A. & Linh, N. T. T. 2021 Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers* **12**(3), 101075.
- Uddin, M. T. & Uddiny, M. A. 2015 A guided random forest based feature selection approach for activity recognition. In: *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. IEEE, pp. 1–6.
- Ullah, K. & Zhang, J. 2020 GIS-based flood hazard mapping using relative frequency ratio method: a case study of Panjkora River Basin, eastern Hindu Kush, Pakistan. *PloS One* **15**(3), e0229153.
- Vilasan, R. T. & Kapse, V. S. 2022 Evaluation of the prediction capability of AHP and F-AHP methods in flood susceptibility mapping of Ernakulam district (India). *Natural Hazards* **112**(2), 1767–1793.
- Vojtek, M. & Vojteková, J. 2019 Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water* **11**(2), 364.
- Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., Shirzadi, A., Shahabi, H., Panahi, S. & Costache, R. 2019 Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. *Journal of Environmental Management* **247**, 712–729.
- Wang, H., Wang, H., Wu, Z. & Zhou, Y. 2021 Using multi-factor analysis to predict urban flood depth based on Naive Bayes. *Water* **13**(4), 432.
- Wu, Y., She, D., Xia, J., Song, J., Xiao, T. & Zhou, Y. 2023 The quantitative assessment of impact of pumping capacity and LID on urban flood susceptibility based on machine learning. *Journal of Hydrology* 129116.
- Yariyan, P., Avand, M., Abbaspour, R. A., Torabi Haghighi, A., Costache, R., Ghorbanzadeh, O., Janizadeh, S. & Blaschke, T. 2020 Flood susceptibility mapping using an improved analytic network process with statistical models. *Geomatics, Natural Hazards and Risk* **11**(1), 2282–2314.
- Yaseen, A., Lu, J. & Chen, X. 2022 Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model. *Stochastic Environmental Research and Risk Assessment* **36**(10), 3041–3061.

First received 26 March 2023; accepted in revised form 18 May 2023. Available online 2 June 2023