Geoscience Frontiers 14 (2023) 101625

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research Paper

Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model

Biswajeet Pradhan [a,b], Saro Lee [c,d,*], Abhirup Dikshit [a], Hyesu Kim [e]

[a] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering & IT, University of Technology Sydney, Sydney, NSW 2007, Australia
[b] Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia
[c] Geoscience Data Center, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124 Gwahang-no, Yuseong-gu, Daejeon 34132, South Korea
[d] Department of Resources Engineering, Korea University of Science and Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, South Korea
[e] Department of Astronomy, Space Science and Geology, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, South Korea

ARTICLE INFO

Article history:
Received 15 December 2022
Revised 6 March 2023
Accepted 22 April 2023
Available online 28 April 2023
Handling Editor: Wengang Zhang

Keywords:
Flood susceptibility
Explainable AI
Deep learning
South Korea

ABSTRACT

Floods are natural hazards that lead to devastating financial losses and large displacements of people. Flood susceptibility maps can improve mitigation measures according to the specific conditions of a study area. The design of flood susceptibility maps has been enhanced through use of hybrid machine learning and deep learning models. Although these models have achieved better accuracy than traditional models, they are not widely used by stakeholders due to their black-box nature. In this study, we propose the application of an explainable artificial intelligence (XAI) model that incorporates the Shapley additive explanation (SHAP) model to interpret the outcomes of convolutional neural network (CNN) deep learning models, and analyze the impact of variables on flood susceptibility mapping. This study was conducted in Jinju Province, South Korea, which has a long history of flood events. Model performance was evaluated using the area under the receiver operating characteristic curve (AUROC), which showed a prediction accuracy of 88.4%. SHAP plots showed that land use and various soil attributes significantly affected flood susceptibility in the study area. In light of these findings, we recommend the use of XAI-based models in future flood susceptibility mapping studies to improve interpretations of model outcomes, and build trust among stakeholders during the flood-related decision-making process.

1. Introduction

Floods are natural hazards that cause devastating financial and socioeconomic damage. In Asian countries, which are among the most flood-prone worldwide, approximately 90% of human losses are caused by natural hazards, principally floods (Dutta and Herath, 2004; Smith, 2013).

In South Korea, flood events occur yearly, mainly due to typhoons and the summer monsoon. Almost 80% of all property damage in South Korea is caused by floods (Kim et al., 2007). Typhoons occur in South Korea 1–3 times per year on average, mainly from August to September. Major recent typhoons include the Rusa and Maemi typhoons, which occurred in 2002 and 2003, respectively.

Flood susceptibility maps have been developed to identify and characterize potential flood-prone areas based on their physical characteristics (Vojtek and Vojteková, 2019). Flood susceptibility maps can help reduce flood-related damage, which is crucial for disaster mitigation (Sahoo and Sreeja, 2015). Recent flood susceptibility mapping studies have used hydrological (Rahman et al., 2019), hydrodynamic (Wagenaar et al., 2020), statistical (Samanta et al., 2018), multi-criteria decision analysis (MCDA) (de Brito and Evers, 2016; Rahman et al., 2019), and machine learning (ML) models (Lee et al., 2018; Darabi et al., 2019) integrated with geographical information system (GIS) software. However, hydrological and hydrodynamic models are time-consuming, and have calibration issues that reduce their accuracy in identifying flood-affected regions (Fenicia et al., 2014). MCDA models are widely used and their accuracy has been demonstrated in several studies (e.g., Danumah et al., 2016; Luu et al., 2018; Rahman

* Corresponding author at: Geoscience Data Center, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124 Gwahang-no, Yuseong-gu, Daejeon 34132, South Korea.
E-mail addresses: Biswajeet.Pradhan@uts.edu.au (B. Pradhan), leesaro@kigam.re.kr (S. Lee).

https://doi.org/10.1016/j.gsf.2023.101625

et al., 2019); among these, the analytic hierarchy process (AHP) model is suitable for complex decision-making based on limited data (Chen et al., 2011; Dikshit et al., 2020c).

ML-based models have been found to be more accurate than other flood susceptibility mapping models. These include artificial neural networks (ANNs) (Rahman et al., 2019), random forest (RF) models (Chen et al., 2020), and support vector machine (SVM) models (Tehrany et al., 2015; Mojaddadi et al., 2017). SVM-based flood susceptibility maps of Malaysia have achieved an accuracy of > 80% (Tehrany et al., 2015), and integrated statistical and ML models have been used for flood susceptibility mapping of Bangladesh, with a prediction accuracy of 86% (Rahman et al., 2019). Chen et al. (2020) found that an RF model produced more accurate susceptibility maps of Jiangxi Province, China than a decision tree approach. Recently, Wagenaar et al. (2020) comprehensively reviewed the use of ML-based models for flood risk and impact assessment.

Deep learning models have exceeded the performance of ML-based models in several fields, such as computer vision and natural language processing (LeCun et al., 2015). A few studies have applied deep learning models to flood susceptibility mapping, including convolutional neural network (CNN)-based models. For example, flood susceptibility maps of Jiangxi Province, China produced using CNN architectures with one, two, or three dimensions (1D, 2D, or 3D, respectively) were compared (Wang et al., 2019). Similarly, a 2D CNN-based model was used to develop susceptibility maps for Iran (Khosravi et al., 2020), and several ML-based models were compared with a deep learning model, the deep belief network (DBF), for flood susceptibility mapping of central Vietnam; the DBF outperformed all ML models (Pham et al., 2021).

However, all of these studies have lacked a key component: model interpretability or explainability. Although hybrid and deep learning models tend to produce more accurate results, they are considered to be "black boxes" and are therefore rarely selected by stakeholders. Thus, models that readily demonstrate how specific outcomes are achieved are in high demand (Dikshit et al., 2020a). One of the most commonly used explainable/interpretable models is the Shapley additive explanation (SHAP) model, which produces different types of plots that illustrate how interdependencies among variables lead to specific model outcomes. SHAP models are increasingly being used for various types of geohazard research, such as mapping building damage after earthquake events (Matin and Pradhan, 2021), vegetation classification (Abdollahi and Pradhan, 2021), and assessment of drought effects (Dikshit and Pradhan, 2021). To our knowledge, this is the first study to apply the explainable artificial intelligence (XAI) model for flood susceptibility mapping. In this study, we used a CNN-based model to develop a flood susceptibility map and then applied SHAP to explain the model outcomes.

2. Materials and methods

2.1. Study area

The region of interest is Jinju Province, in southern South Korea (Fig. 1). The average annual precipitation of Jinju is 1,591 mm, which is mainly concentrated in July, followed by August. In August 2018, a nationwide heavy rainfall event occurred in South Korea, causing two deaths and total property damage of 41.5 billion won. In October of the same year, typhoon Kong-rey produced heavy rainfall throughout Korea, causing two deaths, displacing 2,381 flood victims, and leading to property damage of 54.9 billion won. The total flood damage area within the study region in 2018 was 23 km^2. In September 2019, Typhoon Mitag caused 14 deaths and a total property damage of 167 billion won. The total flooding area of Jinju Province in 2019 was 7.6 km^2 (Ministry of the Interior and Safety (MIS) and Korea, 2019; Ministry of the Interior and Safety (MIS) and Korea, 2020).

2.2. Factors causing flood

In this study, we analyzed 12 factors causing flood based on data from previous studies (Lee et al., 2017, 2018) and the characteristics of the study region.

2.2.1. Geographical factors

Elevation is a key factor influencing the probability of flood occurrence (Lei et al., 2021). Areas with higher elevation experience less flooding, because water flows from higher to lower terrain, where flooding occurs in flatter regions (Botzen et al., 2013). Topographical parameters directly affected by flow extent and runoff speed play important roles in flood occurrence (Kia et al., 2011). Topographical parameters related to flood occurrence are extracted directly from digital elevation models (DEMs) for use in modeling studies. Therefore, highly precise DEMs are essential for flood susceptibility mapping. We divided our study region into seven subregions according to their topographical characteristics (Fig. 2a), including mountains (23%), hills (35.6%), alluvial fans (11.6%), and alluvial plains (11.34%) (See Fig. 3).

Land use/land cover (LULC) is another important factor contributing to floods (Rizeei et al., 2016; Darabi et al., 2019). Areas with healthy vegetation are less prone to floods, as flood events and vegetation density are inversely proportional. Water runoff is a concern in urban areas, which are usually composed of barren lands and impermeable surfaces (Rizeei et al., 2016). Jinju Province comprises urban (6%), agricultural (27%), forest (57.2%), grassland (3.2%), and wetland (0.8%) areas, as well as bare land (1%) and water cover (4.7%) (Fig. 2b).

Lithology refers to the geological characteristics of a region and is a good indicator of past flood events within a given area (He et al., 2007; Arabameri et al., 2019). In this study, we divided the study region into 16 parts according to lithology. The western part of the study area is dominated by metamorphic rocks, including metatectic gneiss, hornblende gneiss, and banded gneiss. Metamorphic rocks are inconsistently covered by Mesozoic sedimentary rocks, with strikes ranging from 10°N to 23°E and dips of approximately 10°E. These sedimentary rocks include Nagdong and Silla series rocks. The Nagdong series consists of the Wonji, Madong, Jinju, and Chilgog formations, distributed from west to east; these rocks are mainly composed of arkose sandstone containing abundant feldspar. The Wonji and Jinju formations are mainly composed of black shale, whereas the Madong and Chilgog formations are composed of purplish sandy shale, which separate the layers. The Wonji formation consists of pebbly sandstone, arkose sandstone, sandstone, shale, and a thin limestone layer from the bottom to the surface. The Madong formation is composed of sandy shale, purplish fine sandstone, and shale. The Jinji formation is composed of arkose sandstone, black sandy shale, and shale. The Chilgog formation is composed of purplish sandy shale and arkose sandstone. The Silla series consists of Silla conglomerate, and Haman and Chindong formation rock. The Silla conglomerate corresponds to the base of the Silla series and is partly composed of agglomerate. The Haman formation is composed of purplish sandy shale, shale, and mudstone, and often includes tuffaceous sandstone. The Chindong formation consists of grayish and black shale. The Haman and Chindong formations alternate with intruded granodiorite. Igneous rocks include pegmatite, porphyrite, acidic dikes, basic dikes, and granodiorite. Pegmatite and porphyrite intrude between metamorphic rocks, and other igneous rocks intrude between sedimentary rocks. The acidic dikes include felsite, and basic dikes include diorite. Quaternary and present-age

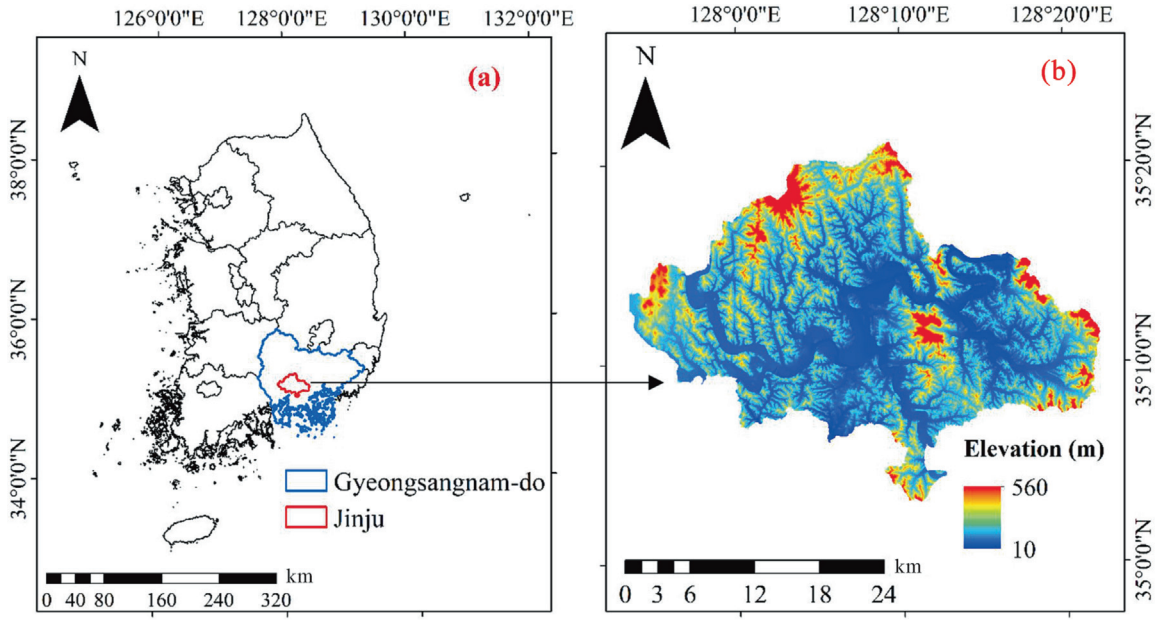

Fig. 1. (a) Location of the study area in South Korea, and (b) Elevation map of Jinju region.

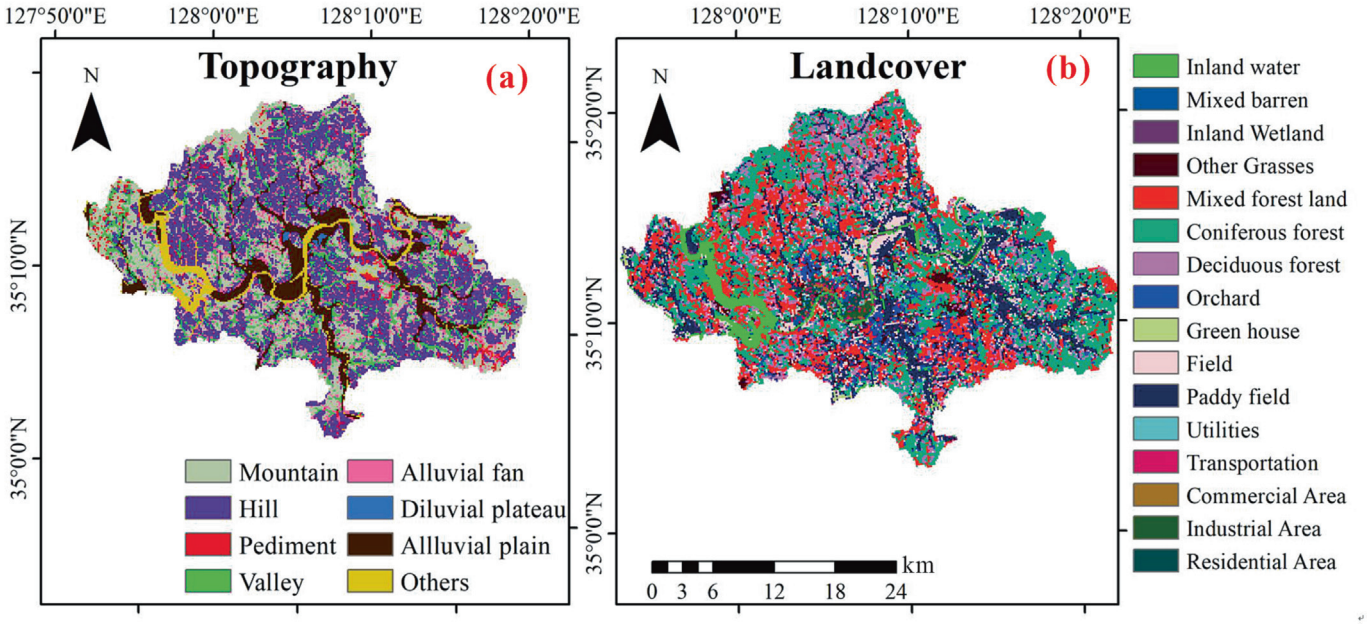

Fig. 2. (a) Topography, and (b) Landcover maps of the study region.

alluvia flow throughout the study area (Choi and Kim, 1963; Choi et al., 1968; Kim et al., 1969).

2.2.2. Soil attributes

Impervious surfaces tend to increase flooding; following heavy precipitation, saturated soil behaves similarly to an impervious surface (Hawley and Bledsoe, 2011; Blum et al., 2020). Water storage potential, which affects the water balance, is determined based on surface soil characteristics (Hong et al., 2018). In this study, we included four soil attributes: soil depth, soil drainage, surface soil texture, and deep soil texture (Fig. 4a–d). Throughout most of the study area, the surface soil texture is silt loam, with some riparian areas showing fine sandy loam and loamy fine sand. The deep soil is approximately 75% clay loam; riparian agricultural land is mainly sandy loam, sand, and silty clay loam, with occasional clay and slit loam. In most urban and agricultural areas, soil thickness reached > 100 cm, whereas in forested areas, soil thickness was approximately 20 cm.

2.2.3. Forest attributes

Forests are important for watershed management because excessive deforestation leads to soil erosion and reduced water retention, which increases flood risk. The forest attributes examined in this study included forest type, density, and composition, and stem diameter (Fig. 5a–d).

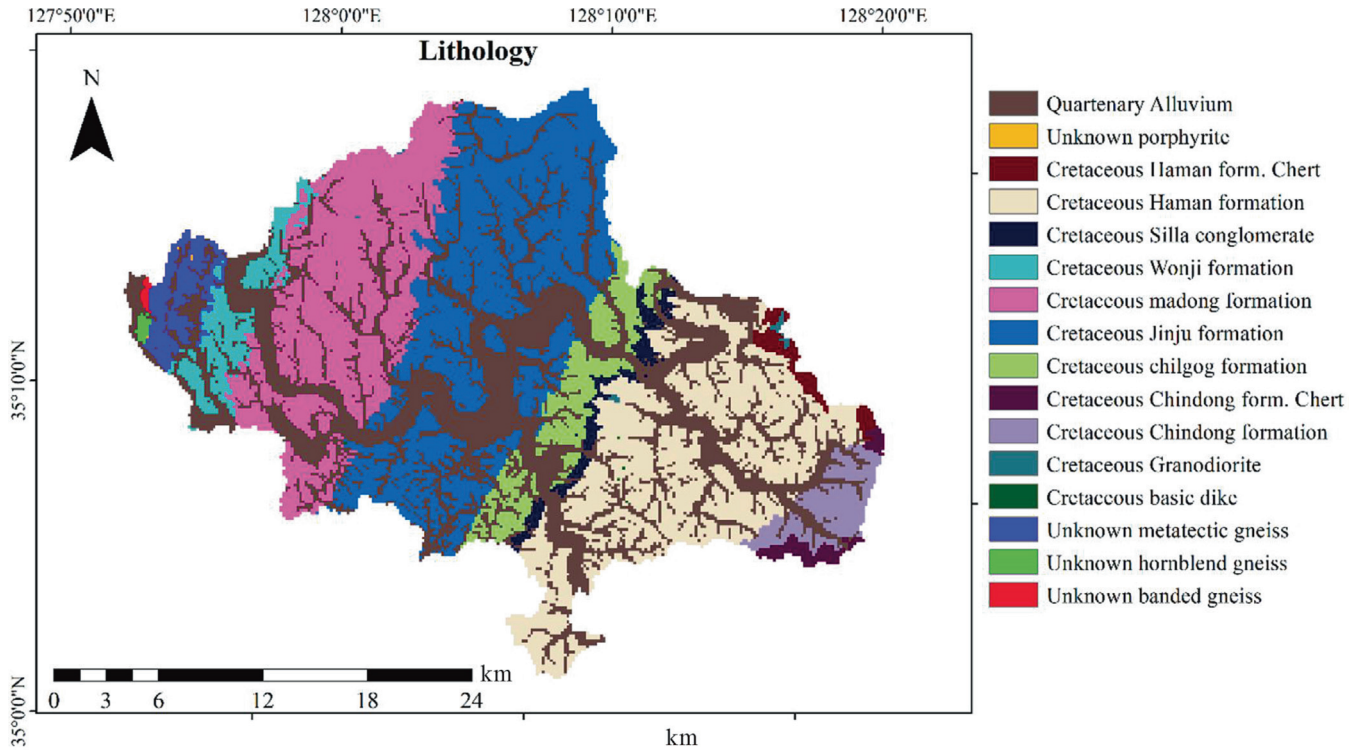

Fig. 3. Lithology map of the study region.

2.3. Flood inventory

Flood inventory maps provide details about inundated regions and provide key information about historical flood event characteristics (Zazo et al., 2018). The flood inventory data used in this study were obtained from LX, the Korea Land Information Corporation, and comprises direct and indirect survey data. The data are presented as a flooding boundary vertex within a plane coordinate system, to delineate the actual flooding area. We used a total of 582 locations of flood events that occurred in 2018 and 2019; among these, 492 occurred in 2018 and 90 occurred in 2019 (Fig. 5). We used 70% of the data for training and the remaining 30% for model testing (See Fig. 6).

3. Methodology

3.1. CNn

The CNN is a hierarchical deep learning algorithm based on local connectivity and shared weights (LeCun et al., 1998). These features allow CNNs to hierarchically extract low-, medium- and high-level image features. CNNs consists of three main layers: convolutional layers, which read input data sequences and automatically extract relevant features, pooling layers, which extract features and identify the most important variables, and fully connected layers, which interpret the internal representations of the data and output a vector (See Fig. 7).

The first layer extracts feature maps related to the target variable; it requires two inputs: an image matrix and filter. The second layer uses an activation layer, which appears after the convolutional layer if it is nonlinear. The choice of activation functions is very important because they help the network learn complex patterns within the data. The final layer conserves the important information and reduces the number of parameters, especially when large images are used as input. Various CNN architectures (e.g., 1D, 2D, or 3D) can be developed based on the data and input type (Wang et al., 2019). In this study, we used a CNN-2D model similar to that of Wang et al. (2019).

3.2. Explainability

There is a fundamental difference between interpretability and explainability (Rudin, 2019). Interpretability does not have a mathematical definition, instead relying on the ability of humans to decipher model outcomes (García and Aznarte, 2020). In contrast, explainability refers to a model-based understanding of the outcomes of a separate black box model (Rudin, 2019). In an ideal scenario, the model should be able to explain the results accurately; this may be true for simpler models, but ML models are more complex and require separate models to examine their outcomes. Rudin (2019) argued that the focus should be on developing interpretable rather than explainable models, but also acknowledged several challenges for interpretable models and that explainable models have considerable value for understanding outcomes in certain applications. For example, recent geohazard studies have shown that explainable models significantly promote understanding of model outcomes (Dikshit and Pradhan, 2021; Matin and Pradhan, 2021).

Explainable models include the local interpretable model-agnostic explanation (LIME) (Ribeiro et al., 2016), neural-backed decision tree (NBDT) (Wan et al., 2020), and SHAP (Lundberg and Lee, 2017) models. SHAP was first introduced as a game theory model to determine the contribution of an individual player in a collaborative game (Shapley, 1953). The idea was to distribute the total gain among players based on their contributions to the outcome; SHAP values provided a solution to the problem of providing a fair reward to every player, assigning a unique value determined by local accuracy, consistency, and null effect (Shapley, 1953). The recent development of ML algorithms by Lundberg and Lee (2017) has opened new avenues for understand-

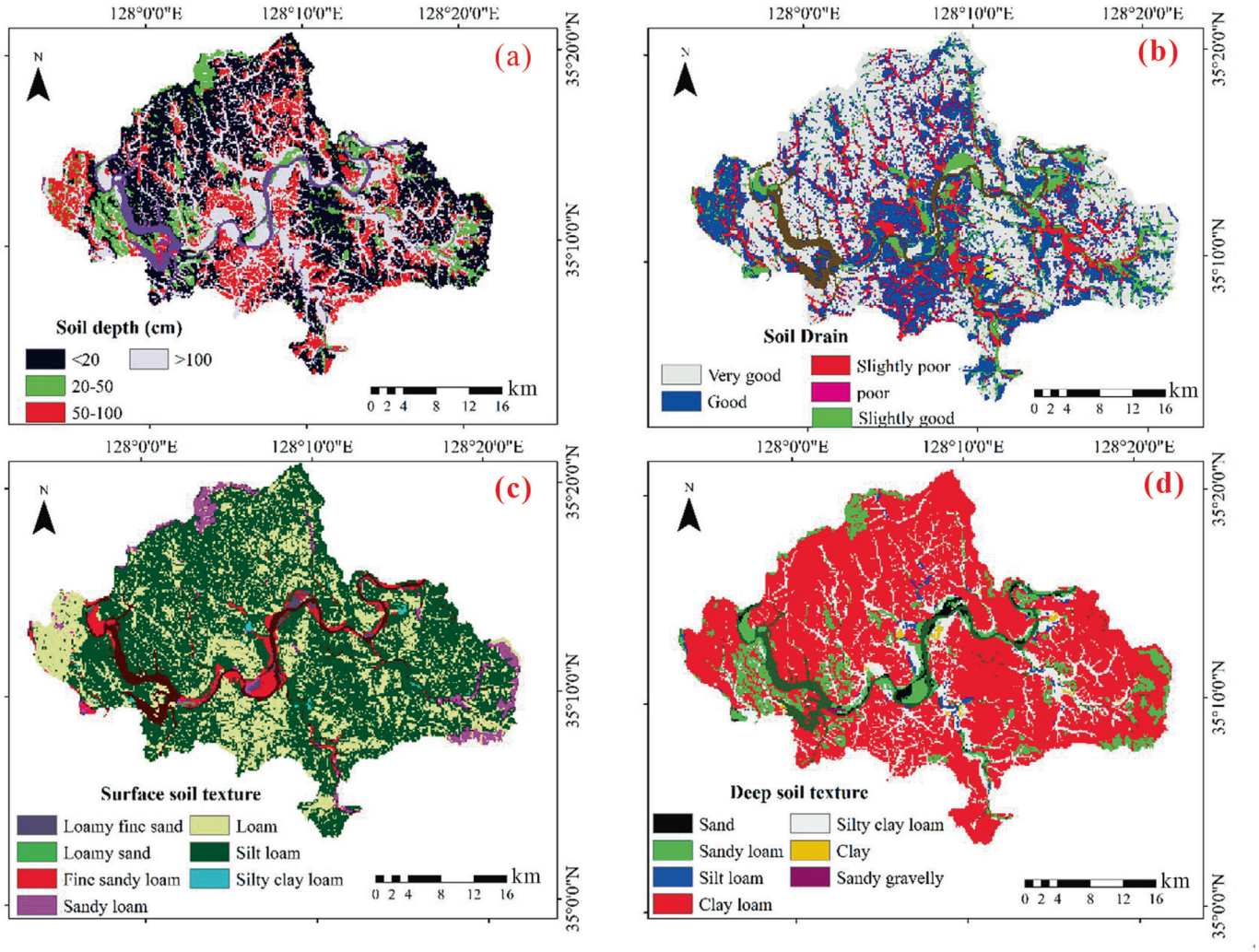

Fig. 4. Soil attributes used as variables. (a) Soil depth; (b) Soil drain; (c) Surface soil Texture; and (d) Deep soil texture.

ing model outputs, providing more transparency for models that are usually considered black boxes.

The Shapley value is calculated based on the average marginal contribution across all possible permutations of the features, as follows Eq. (1).

$$\varnothing_i = \sum_{S \subseteq Nleft\{i\}} \frac{|S|!(n-|S|-1)!}{n!}[v(S \cup \{i\}) - v(S)] \quad (1)$$

where $\varnothing_i$ is the contribution of feature i, N is the set of all features, n is the number of features in N, S is the subset of N containing feature i, and $v(N)$ is the base value, i.e., the predicted outcome for each feature in N without knowing the feature values.

The model outcome for each observation is estimated by summing the SHAP values of all features for that observation. Therefore, the explanatory model is formulated as follows Eq. (2).

$$g(z') = \varnothing_0 + \sum_{i=1}^{M} \varnothing_i z' \quad (2)$$

where $z' \epsilon \{0,1\}^M$, M is the number of features, and $\varnothing_i$ can be obtained from Eq. (2). SHAP provides multiple artificial intelligence (AI) model explainers. Describing the different model explainers is beyond the scope of this study; details can be found in Molnar (2020). The deep explainer was designed specifically for use with deep learning algorithms, which are the focus of this study. A flowchart of the approach used in this study is shown in Fig. 8.

4. Model development

During model development, flood and non-flood pixels were allocated values of 1 and 0, respectively, and triggering factors were overlaid. Thus, all essential data were extracted to flood and non-flood locations; 70% of these data were used for training and 30% for testing, which is the most common split ratio for flood modeling (Tehrany et al., 2014; Wang et al., 2020).

The CNN architecture was comprised of four feature-capturing convolutional layers and a final dense, fully connected layer to learn about feature classification. Hyper-parameterization is a key step when developing a neural network model (LeCun et al., 2015). The sizes of the convolutional and pooling layers are determined according to the scale of its operation (Gowlik et al., 2015). An activation function defines the weighted sum of the input and approximates any nonlinear functions; the rectified linear unit (ReLU) function was used in this study to introduce nonlinearity. A loss function measures inconsistencies between the predicted and observed values; in this study, we used the binary cross-entropy loss function. We also used an Adam optimizer, with standard β values of 0.01 (momentum) and 0.001 (learning rate) for momentumized gradient descent in our back propagation.

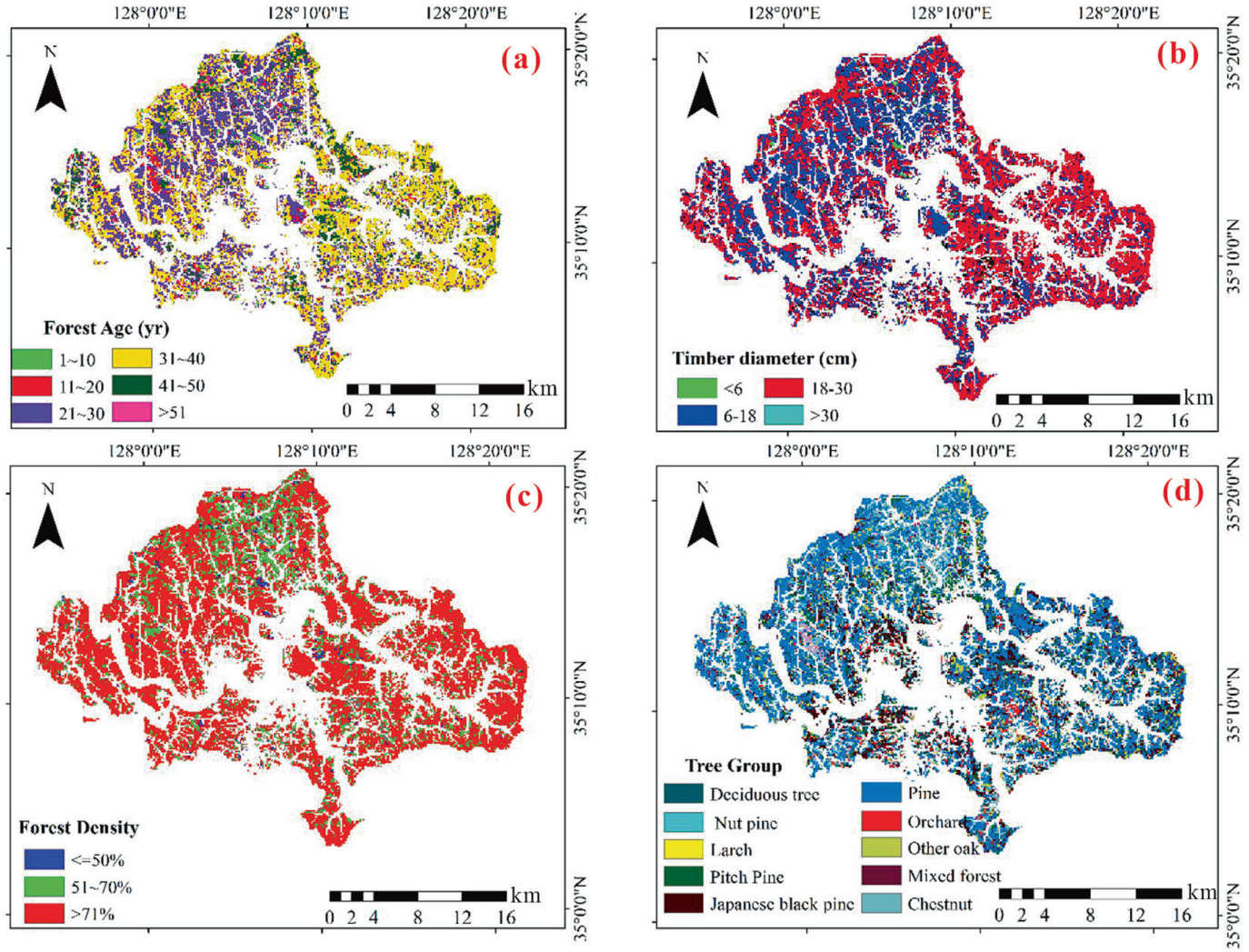

Fig. 5. Forest attributes used in the study. (a) Forest age; (b) Timber diameter; (c) Forest density; and (d) Tree types.

Model performance was evaluated using the area under the receiver operating characteristic curve (AUROC), which is a common approach in geohazard modeling (Fan et al., 2017; Dikshit et al., 2020b) based on the relationship between sensitivity (ordinate axis) and specificity (abscissa), where sensitivity refers to correctly identified flood pixels and specificity refers to correctly identified non-flood pixels using a confusion matrix (Fawcett, 2006). These parameters are defined as follows Eqs. (3) and (4).

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

where *TP* (true-positive) and *TN* (true-negative) are the numbers of correctly classified grid cells, and *FP* (false-positive) and *FN* (false-negative) are the numbers of incorrectly classified grid cells.

The AUROC was used to assess model prediction quality by analyzing its ability to predict the occurrence or non-occurrence of events (Dikshit et al., 2020c). Specifically, an AUROC value of 1 indicates perfect agreement between actual and modeled data, whereas a value of 0.5 indicates the occurrence of an expected outcome by chance, and a value of 0 indicates no agreement (Fawcett, 2006; Choubin et al., 2019). After running the CNN model, a flood susceptibility map was developed and divided into five classes: very high, high, moderate, low, and very low flood susceptibility.

In data-driven modeling, partial dependence plots or bar plots are typically used to show the influence and interactions of each variable. In SHAP-based modeling, dependence plots indicate variable relationships better than conventional approaches (García and Aznarte, 2020; Abdollahi and Pradhan, 2021). Several different types of plots can be constructed based on Shapley values, including the summary plot, which explains the cumulative effect of the variables, the dependence plot, which plots the effect of a single feature on model predictions, the individual force plot, which explains the effects of individual variables on a single observation, and the collective force plot, which is created by combining all force plots, each rotated by 90° and stacked horizontally. In the present study, we used summary and force plots. SHAP summary plots were used instead of conventional bar plots to evaluate global significance, whereas local explanations were obtained based on force plots (García and Aznarte, 2020).

5. Results

CNNs have traditionally been used to capture neighborhood pixel information/features from images. In this study, we utilized this characteristic to capture the geographical neighborhood dur-

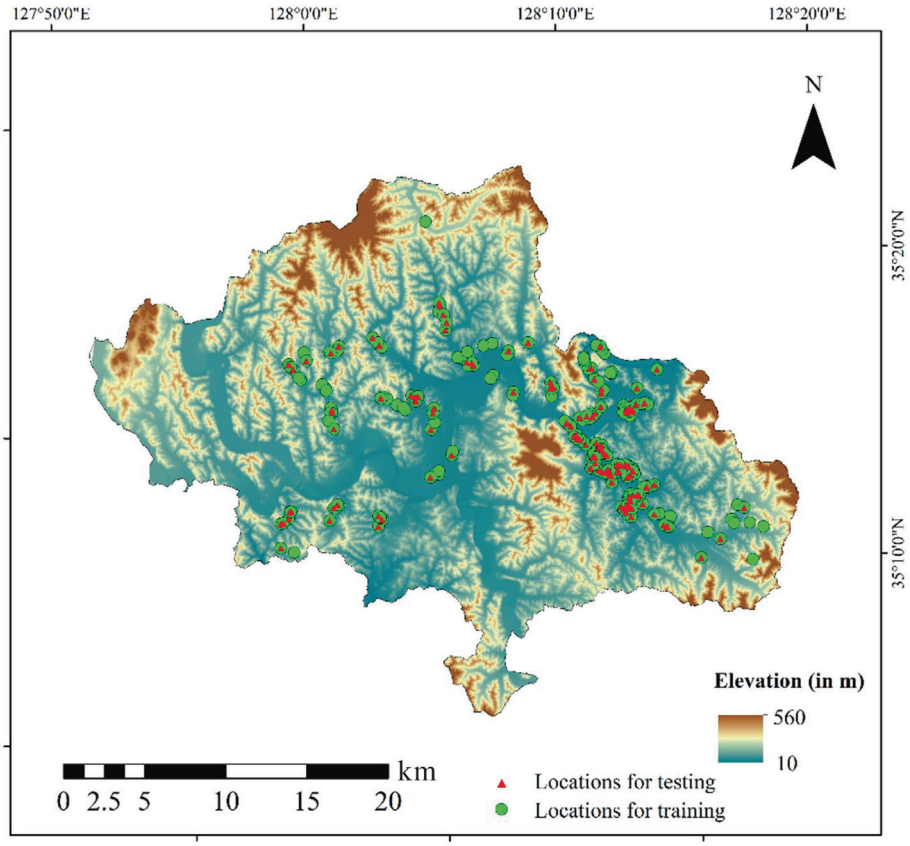

Fig. 6. Flood inventory map used for training and testing the model.

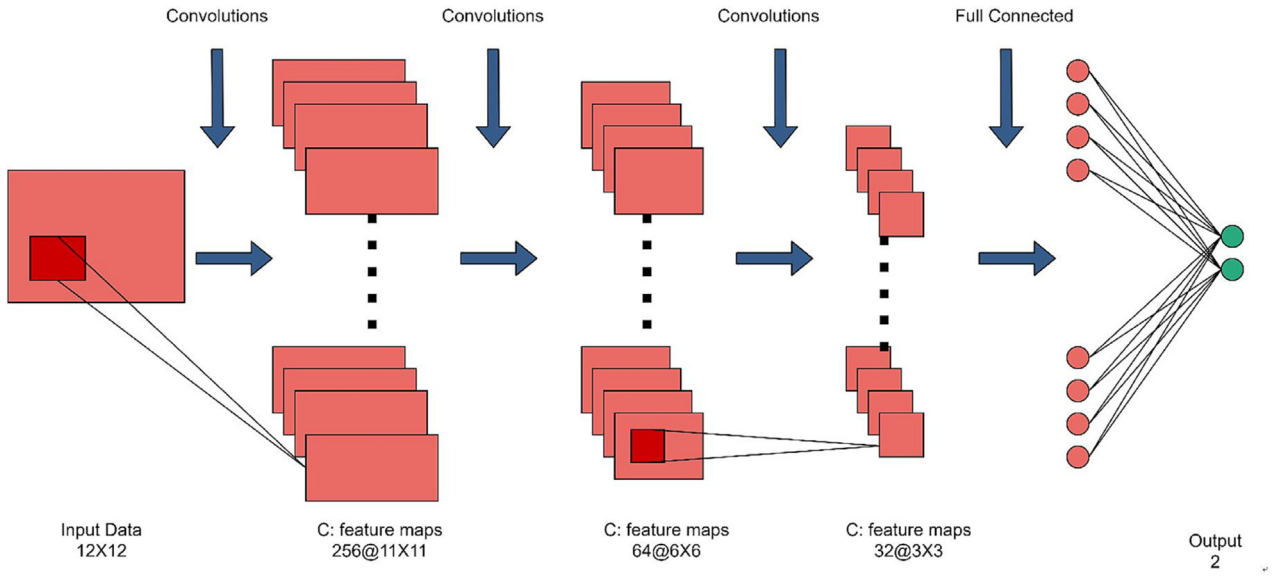

Fig. 7. Schematic illustration of 2D CNN architecture used in the study.

ing flood classification. The AUROC analysis revealed an accuracy of 88.4% for our approach (Fig. 9).

From the model, we developed a map based on five flood susceptibility classes (Fig. 10), which showed that 4.6% and 10.8% of the study region had very high and high flood susceptibility, respectively, whereas 38.6% and 24.8% of the region had very low and low flood susceptibility, respectively; the remaining 21.2% had moderate flood susceptibility.

SHAP models were developed based on a game theory approach, with the properties of different features combined to make a final prediction. The SHAP plots explain the model outputs by considering the importance of neighborhood information. Neighborhoods that promote classification are shown in different colors. For example, a flood class prediction obtained by a CNN model based on a force plot is explained by the plot in Fig. 11. Such explanatory plots demonstrate how multiple variables interact

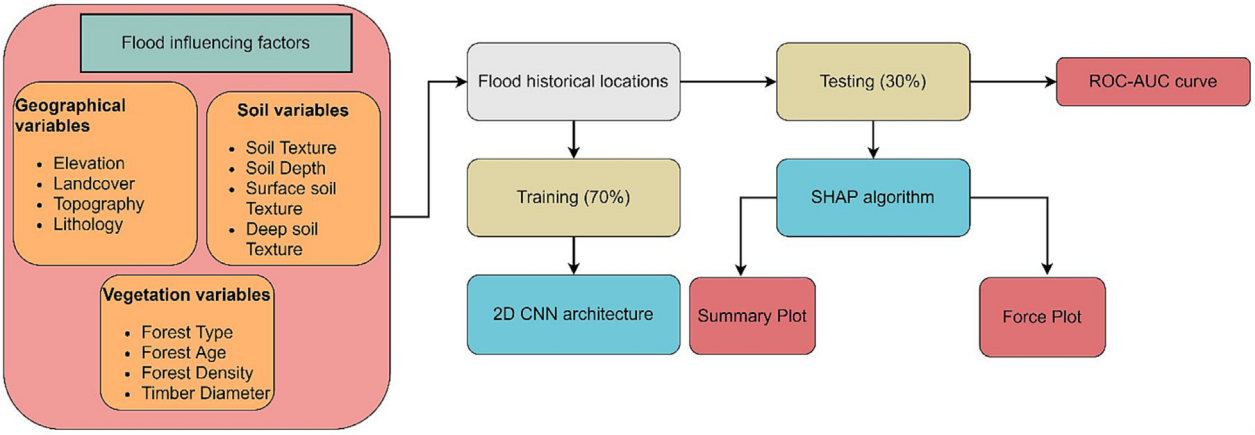

Fig. 8. Flowchart used in the present study.

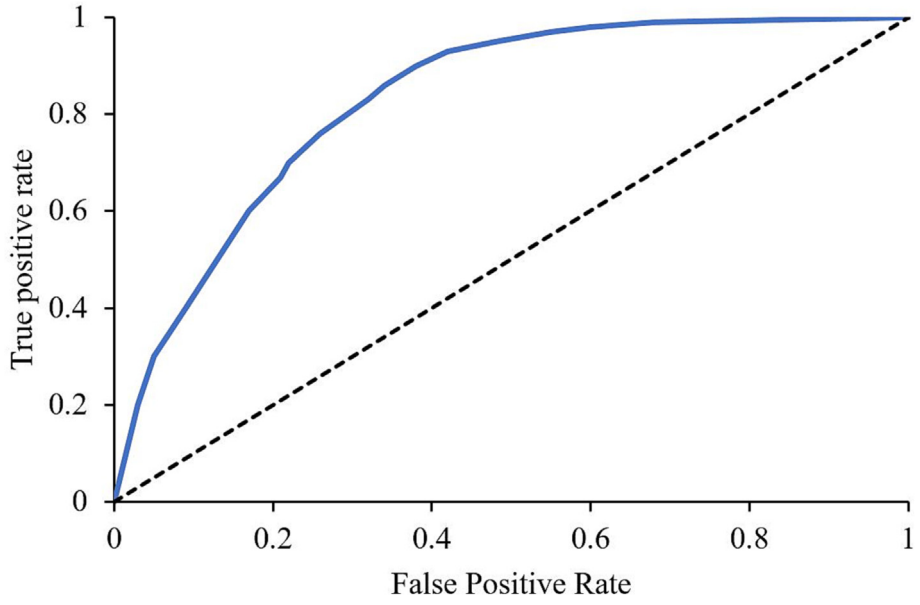

Fig. 9. ROC-AUC curve.

during the production of output (averaged from all predictions) from base data (Abdollahi and Pradhan, 2021). Variables marked in red push the model outcome toward a higher classification, whereas those marked in blue push it toward a lower value. A major benefit of using individual force plots is that they allow the reader to understand the importance of each feature to each pixel, thereby providing a spatial context to output variation. For the output shown in Fig. 11, variables improving classification accuracy for each pixel included land cover, elevation, soil depth, and surface soil texture; those decreasing classification accuracy included forest composition and lithology.

Fig. 12 shows a summary plot, which highlights the low, high, and mean values of each feature among all samples in the training dataset (Matin and Pradhan, 2021). The abscissa represents Shapley values for each observation. Such plots can be used to examine the relationships between a target and variables of interest. In Fig. 12, the most important variable is land use, whereas lithology is the least important variable. The observed importance of land use as the most important variable can be attributed to several potential mechanisms. For example, land use can affect the number of permeable surfaces and vegetation cover, which can impact the rate of infiltration and the amount of runoff. Urbanization and land conversion to impervious surfaces such as buildings and roads can lead to increased runoff and flash flooding. Additionally, land use changes can also affect the hydrological properties of soils, such as their infiltration capacity, which can impact the susceptibility to flooding. On the other hand, lithology, being the least important variable, can be explained by the fact that lithology is a relatively stable property of the Earth's surface and does not change frequently like land use does. Additionally, the study region is mostly composed of the same rock types, thus variations in lithology would not have a significant impact on flood susceptibility. Importantly, this plot summarizes all pixels; it may or may not hold true for individual pixels, as shown in the individual force plot.

6. Discussion

The use of neural networks has greatly advanced the field of flood susceptibility modeling. Several studies have been performed

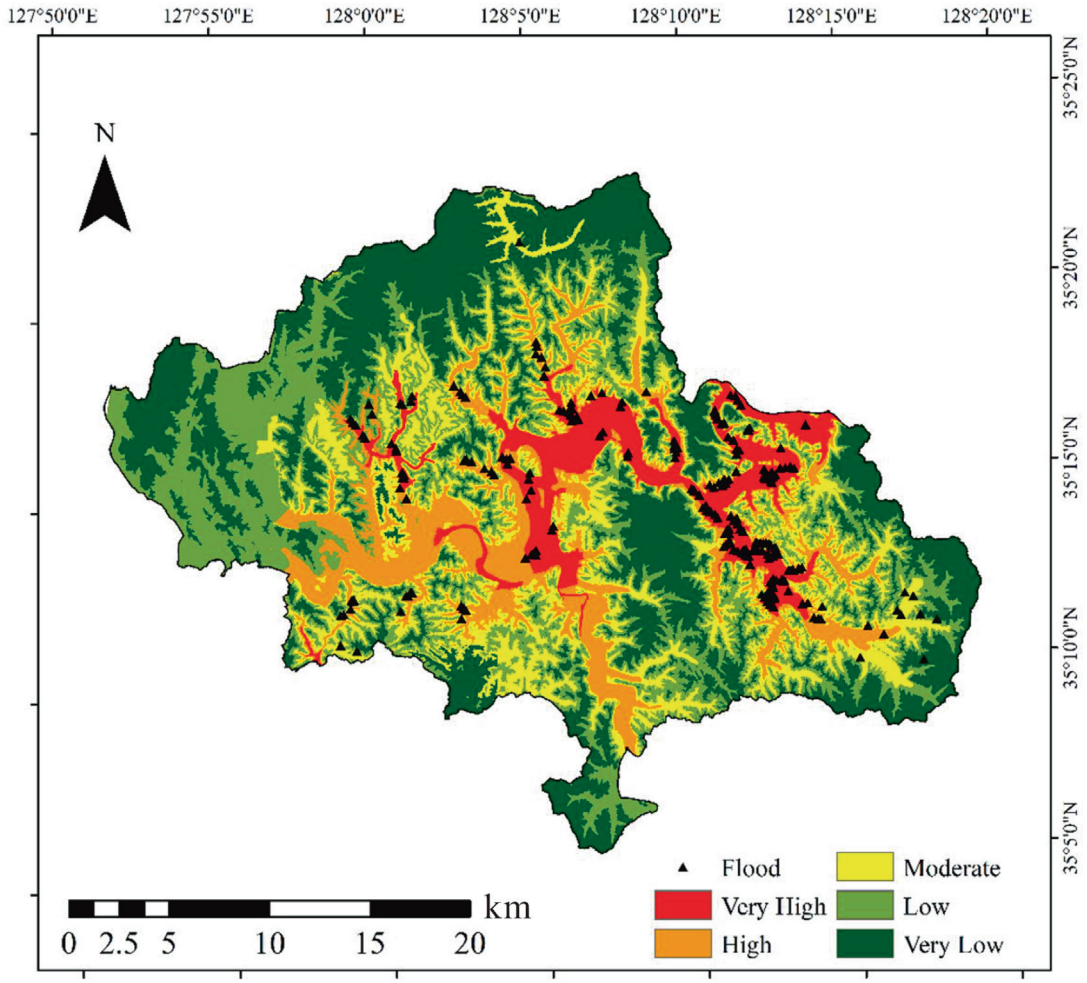

Fig. 10. Flood susceptibility map of the study region.

Fig. 11. SHAP individual force plot (Refer to Table 1 for variable names).

Table 1
Variables used in the present study.

S. No.	Variable	Source
1	Landcover	Ministry of Environment. Korea
2	Elevation	Ministry of Land, Infrastructure and Transport (MOLIT). Korea
3	Soil Depth	National Institute of Agricultural Sciences. Korea
4	Soil Drain	National Institute of Agricultural Sciences. Korea
5	Surface soil texture	National Institute of Agricultural Sciences. Korea
6	Forest Age class	Korea Forest Service
7	Deep soil Texture	National Institute of Agricultural Sciences. Korea
8	Timber diameter	Korea Forest Service
9	Tree Types	Korea Forest Service
10	Forest Density	Korea Forest Service
11	Topography	Ministry of Land, Infrastructure and Transport. Korea
12	Lithology	Korea Institute of Geoscience and Mineral Resources (KIGAM). Korea

using conventional ML-based models for different regions worldwide. A recent surge in the use of deep learning models, such as CNNs, has highlighted their superiority over traditional neural networks. In this study, we used a CNN for flood susceptibility mapping of Jinju Province, South Korea, which has a long history of flood events causing immense damage to infrastructure and loss of life. A total of 12 flood-triggering factors, and 582 historical flood events, were used to develop the model and classify the region. The model achieved an accuracy of 88.4%, demonstrating its reliability.

Previous flood susceptibility studies of South Korea have applied ML models with varying results. For example, Lee et al. (2016) achieved accuracies of 79.1% and 77.2% using RF and boosted tree models, respectively, of Seoul. Similarly, Lee et al. (2018) used frequency ratio and logistic regression models for flood susceptibility mapping of Seoul and achieved an accuracy of > 79% for both models. Lei et al. (2021) compared CNN and a recurrent neural network (RNN) for flood susceptibility assessment in Seoul, and found that CNN provided slightly better results than RNN, with an accuracy of 84%.

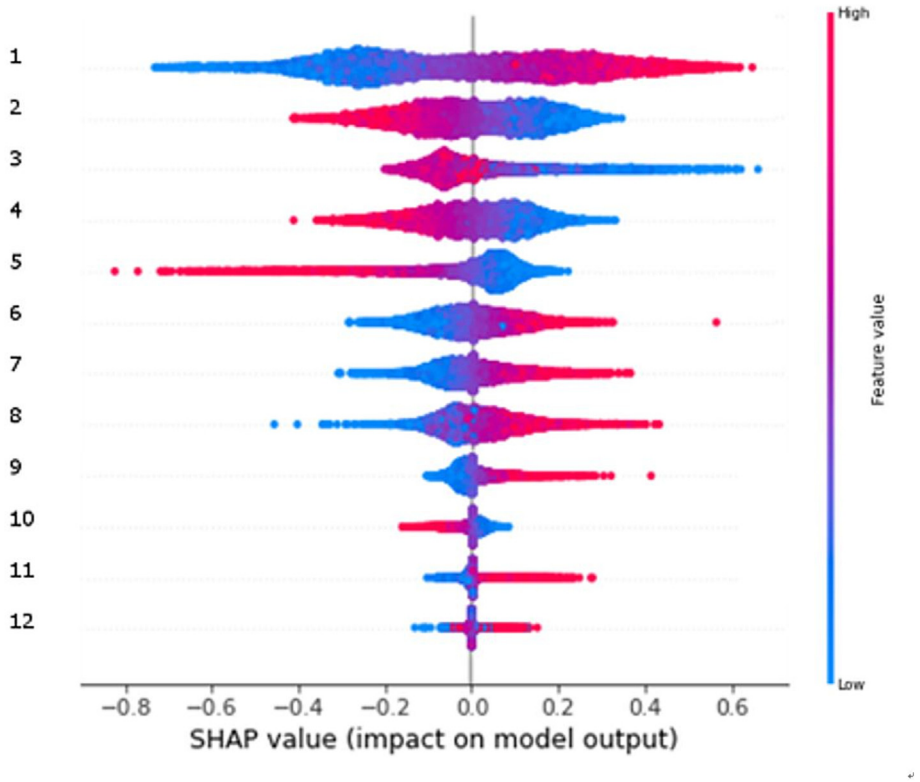

Fig. 12. SHAP summary plot (refer to Table 1 for the variable names).

In this study, we introduced an explainable algorithm to determine how a CNN model achieved a specific result. Two SHAP plots were used for this analysis: an overall summary plot and an individual force plot of flood classification. Based on the summary plot, we determined that land use was the most important variable influencing flood susceptibility, whereas lithology was the least important. It is important to note there exists significant differences between SHAP's summary plot versus traditional feature analysis methods (Wang et al. 2020; Zhang et al. 2021; Zhang et al., 2022). Like, (1) SHAP values are unique for each feature and every data point. Traditional feature importance methods, provide an overall importance score for each feature, which may not be specific to a particular data point. (2) SHAP values are based on the concept of cooperative game theory, which provide a way to fairly distribute a value among a group of individuals. This is a more robust approach as compared to traditional feature importance methods which are typically based on a heuristic or approximation. (3) SHAP values consider the interactions between features, whereas traditional feature importance methods focus on the individual effect of each feature. (4) SHAP values can be used with any model (model-agnostic), whereas traditional feature importance methods may be specific to a certain type of model.

However, single-pixel analysis of flood class showed slightly different results, highlighting the importance of spatial variation. This distinction is particularly important when developing susceptibility maps for large areas, where each region can have different influential variables. Accurate flood susceptibility maps produced using these methods will help stakeholders develop regional mitigation plans and local solutions.

Future studies should also examine the potential of SHAP as a feature identification tool, at both the local (pixel) and regional scales. A recent study used this method to identify factors influencing an earthquake damage mapping study in Palu (Matin and Pradhan, 2021). Another study used Pearson's correlation analysis to identify redundant variables in a flood study in South Korea (Lei et al., 2021). Metrics such as variable importance, information gain ratio, kappa analysis, and spatial autocorrelation have also been applied to analyze variable importance (Meyer et al., 2019). These metrics cannot be evaluated at the pixel level, and provide only an overall sense of variable interaction. The use of SHAP helps remove redundant variables, thus reducing computational cost, which would allow the use of more sophisticated models in developing countries.

7. Conclusion

Floods are among the most destructive recurring natural hazards worldwide. South Korea is greatly affected by flood events; therefore, the development of accurate and interpretable flood susceptibility maps would facilitate the design of effective flood management and mitigation plans. In this study, we used a deep learning CNN model to develop a flood susceptibility map for Jinju Province, South Korea. The main contribution of this work is the application of the SHAP explainable algorithm to determine how the model results were achieved, and the most influential variables. The main finding of this study was that the CNN achieved an AUROC value of 0.88, indicating good accuracy. Moreover, SHAP summary plots showed that, overall, land use was the most influential factor with respect to flood susceptibility; however, this may not hold true for individual flood/non-flood locations, as indicated by the individual force plot. The introduction of XAI models will help unravel the results of black box models and promote a better understanding of variable interactions in geohazard mapping. The utilization of SHAP model in flood susceptibility modelling can lead to a more informed understanding of the underlying mechanisms and factors that drive flood risk. Researchers can use the feature

importance scores provided by the SHAP model to identify key drivers of flood susceptibility and focus their efforts on gaining a deeper understanding of these factors. For practitioners, the SHAP values can be used to prioritize areas for flood mitigation and management efforts, by identifying the areas that are most susceptible to flooding based on the identified key drivers. Additionally, the SHAP model's ability to account for interactions between features, allows practitioners to design more effective and targeted flood management strategies, by considering the complex interactions between various factors that contribute to flood risk.

CRediT authorship contribution statement

Biswajeet Pradhan: Conceptualization, Methodology, Writing – review & editing, Writing – original draft. **Saro Lee:** Supervision, Validation, Visualization, Data curation, Funding acquisition. **Abhirup Dikshit:** Writing – review & editing, Methodology. **Hyesu Kim:** Validation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems, Faculty of Engineering and Information Technology, University of Technology Sydney. Also, this research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) and the National Research Foundation of Korea (NRF) grant funded by Korea government (MSIT) (No. 2023R1A2C1003095).

References

Abdollahi, A., Pradhan, B., 2021. Urban Vegetation Mapping from Aerial Imagery Using Explainable AI (XAI). Sensors 21, 4738.
Arabameri, A., Rezaei, K., Cerdà, C., Conoscenti, C., Kalantari, Z., 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. Sci. Total Environ. 660, 443–458.
Blum, A.G., Ferraro, P.J., Archfield, S.A., Ryberg, K.R., 2020. Causal Effect of Impervious Cover on Annual Flood Magnitude for the United States. Geophys. Res. Lett. 47(5), e2019GL086480.
Botzen, W.J.W., de Boer, J., Terpstra, T., 2013. Framing of risk and preferences for annual and multi-year flood insurance. J. Econ. Psychol. 39, 357–375.
Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., Ahmad, B.B., 2020. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. Sci. Total Environ. 701 134979.
Chen, Y.-R., Yeh, C.-H., Yu, B., 2011. Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan. Nat. Hazards 59 (3), 1261–1276.
Choi, Y.K., Cheong, C.H., Lee, D.S., Kim, S.W., Kim, S.J., 1968. Explanatory text of the geological map of Danseong sheet. Kyeong sang nom do. https://doi.org/10.22747/data.20210701.4263.
Choi, Y.K., Kim, T.Y., 1963. Explanatory text of the geological map of Uiryong sheet. Geological Survey of Korea. https://doi.org/10.22747/data.20210705.4337.
Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci. Total Environ. 651 (2), 2087–2096.
Danumah, J.H., Odai, S.N., Saley, B.M., Szarzynski, J., Thiel, M., Kwaku, A., Kouame, F. K., Akpa, L.Y., 2016. Flood risk assessment and mapping in Abidjan district using multi-criteria analysis (AHP) model and geoinformation techniques (cote d'ivoire). Geoenviron. Disasters 3, 10.
Darabi, H., Choubin, B., Rahmati, O., Torabi Haghighi, A., Pradhan, B., Kløve, B., 2019. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. J. Hydrol. 569, 142–154.
de Brito, M.M., Evers, M., 2016. Multi-criteria decision-making for flood risk management: a survey of the current state of the art. Nat. Hazard Earth Sys. Sci. 16, 1019–1033.
Dikshit, A., Pradhan, B., Alamri, A.M., 2020a. Pathways and challenges of the application of artificial intelligence to geohazards modelling. Gondwana Res. https://doi.org/10.1016/j.gr.2020.08.007.
Dikshit, A., Pradhan, B., Alamri, A.M., 2020b. Short-Term Spatio-Temporal Drought Forecasting Using Random Forests Model at New South Wales. Australia. Appl. Sci. 10, 4254.
Dikshit, A., Pradhan, B., 2021. Interpretable and Explainable AI (XAI) model for spatial drought prediction. Sci. Total Environ. 801 149797.
Dikshit, A., Sarkar, R., Pradhan, B., Jena, R., Drukpa, D., Alamri, A.M., 2020c. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan. Water 12, 267.
Dutta, D., Herath, S.,2004. Trend of floods in Asia and flood risk management with integrated river basin approach,. in: Proceeding of 2nd Asian Pacific Association of Hydrology and Water Resources and Conference, 55–63.
Fan, W., Wei, X.-S., Cao, Y.-B., Zheng, B., 2017. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. J. Mt. Sci. 14, 906–925.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.
Fenicia, F., Kavetski, D., Savenije, H.H., Clark, M.P., Schoups, G., Pfister, L., Freer, J., 2014. Catchment properties, function, and conceptual model representation: is there a correspondence? Hydrol. Process. 28, 2451–2467.
García, M.V., Aznarte, J.L., 2020. Shapley additive explanations for NO_2 forecasting. Ecol. Inform. 56 101039.
Gowlik, P., Tüske, Z., Schlüter, R., Ney, H.,2015. Convolutional Neural Networks for Acoustic Modeling of Raw Time Signal in LVCSR. 16th Annual Conference for the International Speech Communication Association, Germany.
Hawley, R.J., Bledsoe, B.P., 2011. How do flow peaks and durations change in suburbanizing semi-arid watersheds? A southern California case study. J. Hydrol. 405 (1–2), 69–82.
He, B., Xu, Y.-G., Huang, X.-L., Luo, Z.-Y., Shi, Y.-R., Yang, Q.-J., Yu, S.-Y., 2007. Age and duration of the Emeishan flood volcanism, SW China: geochemistry and SHRIMP zircon U-Pb dating of silicic ignimbrites, post-volcanic Xuanwei Formation and clay tuff at the Chaotian section. Earth Planet. Sci. Lett. 255, 306–323.
Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.-X., Chen, W., Kougias, I., Kazakis, N., 2018. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Sci. Total Environ. 621, 1124–1141.
Khosravi, K., Panahi, M., Golkarian, A., Keesstra, S.D., Saco, P.M., Tien Bui, D., Lee, S., 2020. Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. J. Hydrol. 591 125552.
Kia, M.B., Pirasteh, S., Pradhan, B., Mahmud, A.R., Wan, N., Moradi, A., 2011. An artificial neural network model for flood simulation using GIS: Johor River Basin. Malaysia. Environ. Earth Sci. 67, 251–264.
Kim, S., Tachikawa, Y., Takara, K.T.,2007. Recent Flood Disasters and Progress of Disaster Management System in Korea.
Kim, O.J., Yoon, S., Gil, Y.J., 1969. Explanatory text of the geologic map of JinJu sheet. Geological Survey of Korea. https://doi.org/10.22747/data.20210705.4332.
Lecun, Y., Bottou, L., Bengio, Y., Y., Haffner, P.,, 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.
Lee, S., Jeon, S.W., Oh, K.Y., Lee, M.J., 2016. The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: a case study of Inje, Korea. Open Geosci. 8 (1), 117–132.
Lee, S., Kim, J.-C., Jung, H.-S., Lee, M.J., Lee, S., 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomat. Nat. Haz. Risk 8 (2), 1185–1203.
Lee, S., Lee, S., Lee, M.-J., Jung, H.-S., 2018. Spatial Assessment of Urban Flood Susceptibility Using Data Mining and Geographic Information System (GIS) Tools. Sustainability 10 (3), 648.
Lei, X., Chen, W., Panahi, M., Falah, F., Rahmati, O., Uuemaa, E., Kalantari, Z., Ferreira, C.S.S., Rezaie, F., Tiefenbacher, J.P., Lee, S., Bian, H., 2021. Urban flood modeling using deep-learning approaches in Seoul. South Korea. J. Hydrol. 601 126684.
Lundberg, S., Lee, S.-I.,2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
Luu, C., Von Meding, J., Kanjanabootra, S., 2018. Assessing flood hazard using flood marks and analytic hierarchy process approach: a case study for the 2013 flood event in Quang Nam. Vietnam. Nat. Hazards 90, 1031–1050.
Matin, S.S., Pradhan, B., 2021. Earthquake-Induced Building-Damage Mapping Using Explainable AI (XAI). Sensors 21, 4489.
Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Model. 411 108815.
Ministry of the Interior and Safety (MIS), Korea, 2019. Statistical Yearbook of Natural Disaster. Ministry of the Interior and Safety, Korea.
Ministry of the Interior and Safety (MIS), Korea, 2020. Statistical Yearbook of Natural Disaster. Ministry of the Interior and Safety, Korea.
Mojaddadi, H., Pradhan, B., Nampak, H., Ahmed, N., bin Ghazali, A.H., 2017. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomat. Nat. Haz. Risk 8 (2), 1–23.
Molnar, C., 2020. Interpretable machine learning. Lulu Press, Morrisville, USA.

Pham, B.T., Luu, C., Phong, T.V., Trinh, P.T., Shirzadi, A., Renoud, S., Asadi, S., Van Le, H., von Meding, J., Clague, J.J., 2021. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? J. Hydrol. 592 125615.

Rahman, M., Ningsheng, C., Islam, M.M., Dewan, A., Iqbal, J., Washakh, R.A.A., Shufeng, T., 2019. Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. Earth Syst. Environ. 3, 585–601.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?" Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 1135-1144.

Rizeei, H.M., Saharkhiz, M.A., Pradhan, B., Ahmad, N., 2016. Soil erosion prediction based on land cover dynamics at the Semenyih watershed in Malaysia using LTM and USLE models. Geocarto Int. 31, 1158–1177.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215.

Sahoo, S.N., Sreeja, P., 2015. Development of Flood Inundation Maps and quantification of flood risk in an Urban catchment of Brahmaputra River ASCE-ASME. J. Risk Uncertain Eng. Syst. 3, A4015001.

Samanta, S., Pal, D.K., Palsamanta, B., 2018. Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. Appl. Water Sci. 8, 66.

Shapley, L.S., 1953. A value for *n*-person games. Contributions to the Theory of Games. Vol. 2. In: Kuhn, H.W., Tucker, A.W., (Eds.), Annals of Mathematics Studies, No. 28, Princeton University. 307-317.

Smith, K., 2013. Environmental Hazards: Assessing Risk and Reducing Disaster. Routledge.

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. J. Hydrol. 512, 332–343.

Tehrany, M.S., Pradhan, B., Mansor, S., Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125, 91–101.

Vojtek, M., Vojteková, J., 2019. Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. Water 11, 364.

Wan, A., Dunlap, L., Ho, D., Yin, J., Lee, S., Jin, H., Petryk, S., Bargal, S.A., Gonzalez, J.E.,2020. NBDT: Neural-backed decision trees, arXiv preprint arXiv: 2004.00221.

Wagenaar, D., Curran, A., Balbi, M., 2020. Invited perspectives: how machine learning will change flood risk and impact assessmentNat. Hazards Earth Syst. Sci. 20, 1149–1161.

Wang, Y., Fang, Z., Hong, H., 2019. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County. China. Sci. Total Environ. 666, 975–993.

Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., Zhang, W., 2020. Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. Bull. Eng. Geol. Environ. 79, 2763–2775.

Zazo, S., Rodríguez-Gonzálvez, P., Molina, J.-L., González-Aguilera, D., Agudelo-Ruiz, C.A., Hernández-López, D., 2018. Flood Hazard Assessment Supported by Reduced Cost Aerial Precision Photogrammetry. Remote Sens. 10, 1566.

Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L., 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. 12 (1), 469–477.

Zhang, W., Li, H., Han, L., Chen, L., Wang, L., 2022. Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing. China. J. Rock Mech. Geotech. Eng. 14 (4), 1089–1099.