Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: a case study from Slovakia

Matej Vojtek, Jana Vojtekova, Romulus Costache, Quoc Bao Pham, Sunmin Lee, Arfan Arshadh, Satiprasad Sahoo, Nguyen Thi Thuy Linh and Duong Tran Anh

Department of Geography and Regional Development, Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Nitra, Slovakia; Research Institute of the University of Bucharest, Bucharest, Romania; National Institute of Hydrology and Water Management, Bucharest, Romania; Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam; Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam; Department of Geoinformatics, University of Seoul, Seoul, Korea; Center for Environmental Assessment Monitoring, Environmental Assessment Group, Korea Environment Institute (KEI), Sejong, Korea; Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China; School of Ecology & Environment Studies, Nalanda University, Rajgir, Bihar, India; Thuyloi University, Hanoi, Vietnam; Ho Chi Minh City University of Technology (HUTECH) 475A, Ho Chi Minh City, Vietnam

ABSTRACT

Identification of areas susceptible to floods is an important issue which requires an increased attention due to the changing frequency and magnitude of floods, which is mainly a result of the ongoing climate change and increasing anthropic pressure on the landscape. In this study, the aim was to identify the areas susceptible to floods using and comparing two different approaches, namely the multi-criteria decision analysis-analytical hierarchy process (MCDA-AHP) and the machine learning-boosted classification (BCT) and boosted regression (BRT) tree. The study area was represented by the Topla river basin, Slovakia. Altogether, seven relevant flood conditioning factors: elevation, slope, river network density, distance from river, flow accumulation, curve numbers and lithology as well as flood inventory database consisting of 107 flood locations were used. Based on the results, almost 40% of the study area is characterized by high to very high flood susceptibility using the MCDA-AHP. In case of the BCT and BRT models, the share of high and very high flood susceptibility class on the basin area is 45% and 38%, respectively. Validation of the performed flood susceptibility models confirmed generally higher accuracy of the machine learning models. The accuracy of the MCDA-AHP model was 81.33% while the accuracy of the boosted tree models was 87.70% and 91.42%, respectively, for classification and regression. The results of this study can enhance more effective preliminary flood risk assessment according to the EU Floods Directive.

KEYWORDS
Flood susceptibility; multi-criteria decision analysis; analytical hierarchy process; machine learning; boosting; decision trees; GIS; Slovakia

CONTACT Duong Tran Anh ta.duong@hutech.edu.vn

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
1. Introduction

Currently, there are two basic trends indicating that the risk of flooding will increase in the future both in frequency and in magnitude. In addition, the damage caused by floods will also probably have an increasing trend (Hirabayashi et al. 2013). The reasons for the likely increase in the frequency and magnitude of floods are natural variability, the impact of human activities and the associated climate change, which leads to an increase in rainfall intensity (Smith and Ward 1998). The second trend can be seen in an increase in the vulnerability of the society to floods, which is related to the population growth and its economic activities, especially, near rivers or seas (Sofia et al. 2017).

The physical-geographical factors, such as soils, topography and land use/land cover vary in different catchments affecting the transformation of rainfall to runoff, erosion, deposition patterns or sediment delivery and dynamics (Keesstra 2007; Keesstra et al. 2009). However, changes in topography or soils typically occur over longer timescales while watershed hydrologic response, i.e. runoff generation, is strongly influenced by shorter timescale processes like, in particular, the land use/land cover changes (Mohammady et al. 2018; Costache et al. 2020). The interaction and impact of floods and land use/land cover is reciprocal and the effect of floods on land use/land cover as well as on river geomorphology, such as changes in channel width, mobilization of channel sediments or bank erosion mainly in the meandering reaches, can be significant (Yousefi et al. 2018).

In this sense, flood susceptibility maps, which are able to identify areas with high or very high susceptibility to flooding, may represent a useful tool for more effective flood risk management and mitigation. Importance of geographic information systems (GIS) as well as remote sensing data for flood susceptibility mapping is obvious, providing the possibility to work with large datasets of spatial and temporal data.

Based on the literature review, various methods have been developed so far for the identification and assessment of flood susceptibility zones. In addition, these methods were tested and verified at different spatial scales (local, regional, national, supranational or global) and in different regions having diverse predispositions for flooding.

A representative of the knowledge-driven methods is the multi-criteria decision analysis (MCDA). In flood susceptibility mapping, MCDA uses weights which are assigned to the selected flood conditioning factors (Carver 1991; Malczewski 1999). The relative importance and weights of factors are usually being determined by expert judgements (i.e. experts from the research topic). Moreover, different techniques for weighting the factors can be used (Zardari et al. 2015). The most preferred techniques of multi-criteria flood susceptibility assessment include the analytical hierarchy process (Saaty 1980; Elkhrachy 2015; Das 2018; Souissi et al. 2020), weighted linear combination (Kourgialas and Karatzas 2011; Wang et al. 2018) or the analytical network process (Dano et al. 2019; Kanani-Sadat et al. 2019).

Besides the MCDA, many researchers have applied more objective statistical methods for assessing the flood susceptibility, such as frequency ratio (Samanta et al. 2018; Siahkamari et al. 2018) or logistic regression (Pradhan 2010; Lee et al. 2018; Bui et al. 2019). In recent years, however, more advanced and sophisticated methods have been...
applied in flood susceptibility studies including the fuzzy logic (Costache 2019; Ali et al. 2020) or machine learning methods (i.e. data-driven models). The most widely used machine learning algorithms for flood susceptibility mapping are the following (Mosavi et al. 2018): artificial neural networks (Kia et al. 2012), adaptive neuro-fuzzy inference systems (ANFIS) (Bui et al. 2018), support vector machines (SVM) (Tehrany et al. 2015) and decision trees (Lee et al. 2017). Besides the traditional machine learning models, deep learning algorithms have recently gained superiority due to their greater flexibility and predictive accuracy (Khosravi et al. 2020). Decision trees are the machine learning methods which have been applied for flood susceptibility mapping in several studies, such as Tehrany et al. (2015), Lee et al. (2017) or Shafizadeh-Moghadam et al. (2018). Decision trees are based on a tree of decisions starting from branches to the target values of leaves. In this study, boosted classification tree (BCT) and boosted regression tree (BRT) were used. Although there are similarities between these two models, their main difference is in the fact that target values in classification tree model are a discrete set of values and leaves are represented by class labels while in the regression tree model, target values are continuous values as well as the ensemble of trees is used (Mosavi et al. 2018).

The aim of this study is to compare the multi-criteria decision analysis-analytical hierarchy process (MCDA-AHP) technique and machine learning-boosted tree models to identify areas susceptible to flooding based on seven relevant flood conditioning factors and flood inventory database. An element of novelty of this study can be seen in the fact that comparison of the presented approaches for flood susceptibility mapping lacks in literature. In addition, these approaches are applied for the first time in Slovakia (or even in Central Europe) when mapping the flood susceptibility at regional spatial scale. In this study, flood susceptibility is seen as the relatively stable physical-geographical (terrain) as well as more dynamic land cover predispositions of a given area that determine the propensity to flooding, excluding flow or precipitation quantitative information (Jacinto et al. 2015; Santos et al. 2019; Tehrany et al. 2019). The flood conditioning factors were selected in order to reflect mainly fluvial (riverine or regional) flooding that regularly occurs in the study area. This study also partly builds on the results of previous work of Vojtek and Vojteková (2019), especially, when applying the MCDA-AHP based analysis.

2. Study area

The study area is represented by the Topla river basin (Figure 1), which covers an area of 1548 km². The Topla river springs in the Čergov mountain under the Minčol peak approximately at an elevation of 980 m a. s. l. and it mouths into the Ondava river near the Parchovany municipality at an elevation of 105 m a. s. l. The total length of the Topla river is 129 km. The left-sided tributaries include, e.g. Kameneck stream, Radomka stream or Čičava stream while the right-sided tributaries are, e.g. Slatvinec stream, Topola stream, Lomnica stream or Olšava stream.

Geographic coordinates of the study area are as follows: northmost point: 49°27'39"N and 21°16'32"E; southmost point: 48°43'43"N and 21°36'10"E; westmost point: 49°17'03"N and 20°58'22"E; eastmost point: 48°44'38"N and 21°45'22"E.
Figure 1. Study area of the Topla river basin (Slovakia).
According to the geomorphological division of Slovakia (Mazúr and Lukniš 1986), the study area belongs to the following geomorphological units: Lúbovianska vrchovina (highland), Ondavská vrchovina (highland), Busov (mountain), Čergov (mountain), Slanské vrchy (mountain), Beskydské predhorie (foothills), Východoslovenská páhorkatina (upland), Východoslovenská rovina (plain). The highest point has an elevation of 1131 m a. s. l. and it is located on the western boundary of the study area in the Čergov mountain. The lowest point is located in the river mouth.

The Topľa river basin lies in the temperate climate zone. Based on the climatic classification of Slovakia (Lapin et al. 2002), the basin is included in three climatic regions. Lower lying areas, mostly southern parts of the basin, belong to a warm region, which includes one sub-region: warm, moderately humid, with cool winter. This region is characterized with an average annual number of summer days of more than 50 and the daily temperature maximum of more than 25°C. The moderately warm region covers most of the study area and spreads mainly in valleys, uplands and foothills. This region includes two sub-regions: (a) moderately warm, moderately humid, hilly land or highlands and (b) moderately warm, humid, highlands. Furthermore, the region is characterized with the July mean temperature of 16°C or more and an average annual number of summer days of less than 50. The northern and northwestern parts, represented by mountains, are included in a cool region (moderately cool sub-region) with the July mean temperature of less than 16°C. Average annual rainfall in areas with lower elevations ranges from 600 to 700 mm. However, the annual rainfall increases to approximately 1000 mm in the mountainous areas (Bochniček et al. 2015).

Based on the administrative division of Slovakia, the study area is included in the Východné Slovensko (NUTS II), Prešovský kraj (NUTS III) and Košický kraj (NUTS III).

3. Data and methods

3.1. Sources and data

3.1.1. Flood conditioning factors
Flood susceptibility mapping is based on different flood conditioning factors. Data for processing the selected flood conditioning factors were collected from different sources which are reported in Table 1.

Altogether, seven flood conditioning factors were selected, similarly as in the previous work of Vojtek and Vojteková (2019), which were considered to be relevant also for the regional assessment of flood susceptibility: two morphometry factors – elevation and slope; two hydrography factors – river network density and distance from river; two permeability factors – curve number and lithology, represented by the degree of rock permeability; and one hydrology factor – flow accumulation. All flood conditioning factors were processed in ArcGIS software into a raster format with the resolution of 20 × 20 m.

Morphometry factors are represented by the elevation and slope. The digital elevation model (DEM) was interpolated from contour lines and elevation points using the vector topographic map at a scale of 1:50,000 (Figure 2(a)) and the spatial
resolution of 20 × 20 m. The hydrologically correct DEM was created using the ArcGIS software (Hutchinson 1988). As for the slope degree map, it was generated based on the DEM in ArcGIS software (Figure 2(b)).

Hydrography factors of river network density and distance from river were created based on the river network from the vector topographic map at a scale of 1:50,000. The map of river network density (Figure 3(a)) was created using the Line Density tool while the map of distance from river (Figure 3(b)) was generated with the use of the Euclidean Distance tool in ArcGIS software.

Permeability factors are represented by the curve number map which was computed using the SCS-CN method, where the map of hydrological soil groups (Figure

| Primary data | Format       | Source                                           | Derived map                                  |
|--------------|--------------|--------------------------------------------------|----------------------------------------------|
| River network (Topographic map 1:50,000) | Vector (shapefile) | Geodetic and Cartographic Institute Bratislava | Elevation, River network density, Distance from river |
| Contour lines and elevation points (Topographic map 1:50,000) | Vector (shapefile) | Geodetic and Cartographic Institute Bratislava | Elevation, Slope, Flow accumulation |
| Engineering-geological Zoning Map (1:500,000) | Vector (shapefile) | Landscape Atlas of the Slovak Republic (2002) | Lithology |
| Land cover | Vector (shapefile) | CORINE land cover (2018) | Curve number |
| Soil Texture Map (1:500,000) | Vector (shapefile) | Landscape Atlas of the Slovak Republic (2002) | Curve number |

| Sources and primary data used for deriving flood conditioning factors. |

- **Primary data**: Source and primary data used for deriving flood conditioning factors.
- **Format**: Format of the primary data.
- **Source**: Source of the primary data.
- **Derived map**: Derived map from the primary data.

**Figure 2.** Morphometry factors: (a) elevation, (b) slope degree.
4(a)), map of land cover (Figure 4(b)) as well as the official SCS-CN tables (Chow 1964; Cronshey et al. 1986) were the primary inputs. The source data for creating the hydrological soil groups was the vector Soil Texture Map at a scale of 1:500,000 (Čurlík and Šály 2002). Land cover map was created based on the vector layer of CORINE Land Cover from 2018. Using the HEC-GeoHMS extension for ArcGIS software, the resulting raster of curve numbers was computed, as shown in Figure 5(a). The source data for creating the lithology map (Figure 5(b)) was the vector Engineering-geological Zoning Map at a scale of 1:500,000 (Hrašna and Klukanová 2002), which shows the zones of pre-Quaternary and Quaternary rocks. This map was subsequently reclassified to five classes based on the degree of rock permeability according to the work of Hrnčiarová (1993), as shown in Table 2.

Flow accumulation represents the hydrological factor, which was calculated based on the DEM and derived flow direction raster using the ArcGIS software (Figure 6).

3.1.2. Flood locations database

The flood inventory map represents essential data for training as well as testing the selected models. In total, 107 flood points were created based on different sources, such as preliminary flood risk assessment (Ministry of Environment of the Slovak Republic 2011, 2018), databases of the Slovak Water Management Enterprise or Slovak Hydrometeorological Institute, aerial photographs taken from small airplanes, historical documentations and field surveys. In addition, flood points were randomly divided into 70% of training flood locations and 30% of testing flood locations (Figure 1). As for the non-flood locations, they refer mostly to locations with higher
Figure 4. Data inputs for calculating curve numbers: (a) hydrological soil groups, (b) CORINE Land Cover from 2018.

Figure 5. Permeability factors: (a) curve numbers, (b) lithology (degree of permeability).
elevations such as hills or mountain peaks. The non-flood points were mapped based on the large-scale topographic maps of the study area. Since this study focuses on riverine flood susceptibility, the non-flood locations were randomly chosen based on

| Zones of different rock types                  | Degree of permeability |
|------------------------------------------------|------------------------|
| Flyschoid rocks                                | Very Low               |
| Intrusive rocks                                | Low                    |
| Effusive rocks                                 | Moderate               |
| Volcanoclastic rocks                           | Moderate               |
| Undivided volcanic rocks                       | Moderate               |
| Conglomerate-sandstone rocks                   | Moderate               |
| Claystone-siltstone rocks                      | Moderate               |
| Claystone-limestone rocks                      | Moderate               |
| Clay-silty sediments                           | High                   |
| Alternating clay-silty with sandy to gravel sediments | High                 |
| Deluvial deposits                              | High                   |
| Proluvial deposits                             | High                   |
| Floodplain deposits                            | Very High              |
| Loess deposits                                 | Very High              |

Figure 6. Hydrological factor: flow accumulation.
the condition that their distance from river is higher than 100 m. Furthermore, this condition regards predominantly, but not exclusively, the spring areas and the upper basins of the watercourses with low values of flow accumulation, where the occurrence of fluvial floods is the least probable.

3.2. Methodology

The adopted methodology required the application of several methods using various data and specialized software. The main steps of the methodology are the following (Figure 7):

- Data collection and preparation;
- Processing of the flood locations and selected flood conditioning factors applying GIS;
- Flood susceptibility modelling using the MCDA-AHP technique, BCT and BRT models;
- Validation of the resulting flood susceptibility maps.

3.3. Multi-criteria AHP-based analysis

AHP is an expert and mathematical method developed by Saaty (1980) to simplify and improve the decision-making process. The AHP can be defined as a method of
decomposing a complex unstructured situation into simpler parts, i.e. hierarchical system. Using expert pair-wise comparison, this method assigns numerical values to individual factors which express their relative importance. Subsequent synthesis of these evaluations determines the highest priority factor.

To compare the factors more effectively, individual rasters of flood conditioning factors were reclassified into five classes expressing the susceptibility to flooding (5 – very high, 4 – high, 3 – moderate, 2 – low and 1 – very low). Reclassification was carried out based on the frequency of flood occurrences in the Topl‘a river basin in the period 1997–2017, according to the maps of geographical areas and segments of watercourses with historical flood occurrences (Ministry of Environment of the Slovak Republic 2011, 2018), and in the period 1996–2006, according to the work of Solín (2008). For example, the frequency of floods is the highest as well as the most probable at the closest distance to river, therefore, the interval 0–152 m was assigned a value of 5 – very high susceptibility to floods. On the contrary, the interval 1129–2341 m, as the most distant places from the river, was assigned a value of 1 – very low susceptibility to floods due to the lowest occurrence of historical floods.

In case of the river network density map and distance from river map, the classification to intervals was based on the natural breaks (Jenks) grading method. The slope map was classified to intervals according to the work of Demek (1972) while in case of the elevation map, the equal interval grading method was used. The flow accumulation map was manually classified to intervals so that it corresponds to the river network from the topographic map 1:50,000 as much as possible. The curve number map was classified to intervals based on their runoff potential.

Since the expert evaluations were already made for the whole territory of Slovakia in the previous work of Vojtek and Vojteková (2019), the same pair-wise comparisons of the selected flood conditioning factors in a square reciprocal $7 \times 7$ matrix were made also in this study. In addition, the synthesis of the evaluations was similarly performed by normalizing the derived weights and their averaging (Saaty 1980).

Another important step in the MCDA-AHP-based analysis was the calculation of consistency of the pair-wise evaluations. Consistency is recommended to be measured using a consistency ratio (CR), which defines the probability that the matrix ratings were generated randomly. Saaty (1980) suggests that matrices which consistency ratio is greater than 0.10 should be reviewed. The higher the value of the consistency ratio, the higher the possibility of incorrect evaluation of alternatives. The consistency ratio (CR) was calculated using Equation (1):

$$ CR = \frac{CI}{RI} \tag{1} $$

where $CI$ is the consistency index and $RI$ is the random index, which represents the consistency index of a randomly generated pair-wise comparison matrix and depends on the number of criteria (factors) being compared (Saaty 1980).

The consistency index was calculated using Equation (2):

$$ CI = \frac{\lambda - n}{n - 1} \tag{2} $$
where \( n \) is the number of criteria (factors) and \( \lambda \) is the average value of the consistency vector, as suggested by Saaty (1980). The value of the consistency ratio calculated in this study is 0.02, which fulfills the consistency condition and the created matrix evaluations are thus consistent.

The flood susceptibility AHP-based map was created as the weighted sum of the reclassified flood conditioning factors, which were multiplied by their weights according to Equation (3):

\[
FS_{(AHP)} = \sum w_i x_i
\]  

where \( FS_{(AHP)} \) is the AHP-based flood susceptibility, \( w_i \) is the weight of factor \( i \) and \( x_i \) is the reclassified classes of each factor \( i \).

### 3.4. Boosted tree models

Boosted tree models are an ensemble machine learning model of decision trees combined with boosting methods. Similar to other ensemble models based on decision trees, boosted tree model improves the accuracy of the result by fitting several decision trees. Boosted tree models repeatedly choose a random subset for each new tree from the complete dataset, which is consisted of the same number (De’Ath 2007). The difference of boosted trees from other models is the method how the random subset is selected. The boosted tree models use sequential approach for the random subset sampling. It uses a boosting method that weights the incorrectly modelled data from the previous tree to be selected for the new tree. When a subsequent tree is selected, the model tries to select the next tree considering the prediction errors of the previous tree. Since after the first tree is fitted, the model could continuously improve its accuracy by taking into account the suitability of the previous tree (Elith et al. 2008).

Boosted trees show the fast and good performance for training and prediction. They could be also used for response types of various data such as binomial, Gaussian and Poisson, a mix of categorical and continuous data. Especially, they show good performance when there are few observations for the training dataset and they are also robust against missing data and outliers. These characteristics make the boosted tree model suitable for predicting various local environmental issues, such as flood susceptibility.

There are two main parameters in the boosted tree algorithm. First parameter is the complexity (\( c \)), which determines the number of partitions in each tree. The \( c \) value of 1 results in a tree with only one split, regardless of the interaction between the variables in the model. Likewise, if the \( c \) value is 2, a tree with two splits is created. The second parameter is the learning rate (\( lr \)), which determines the contribution of each tree to the final model. In particular, larger \( lr \) values result in fewer trees. Both parameters determine the number of trees needed for optimal prediction. The final goal of the model is to make predictions with the least error. In general, it is important to determine the number of trees by combining the complexity of the tree and the learning rate values, considering the characteristics of the data. The
optimal c and lr values are adjusted based on the size of the total dataset. For datasets with less than 500 training data, a simple tree is considered with a small enough lr value.

The difference between the classification and regression analysis of the boosted tree model lies in the splitting criterion in tree generation. For the classification model, the classification error rate is used through the proportion of training observations in a certain region from each class to calculate the measure of impurity. On the other hand, for the regression model, predictor variable and cut point, which lead to the greatest possible reduction in residual sum of squares, are selected so that they split the predictor space into the regions. The tree with the lowest residual sum of squares is selected. In this respect, the predictor-importance value of each input element is calculated through the node-impurity values for the classification and the resubstitution-estimate for the regression analysis.

Different boosting tree is applied to each class with a value of 1 or 0 and fitted to each class of dependent variables. Finally, a logistic transformation calculates the residuals for subsequent boost steps to predict the classification probabilities for predictions of each value of 1 and 0 (Friedman 2002). BCT calculates the prediction of flood susceptibility through Equation (4):

$$P(y = 1|x) = \frac{1}{1 + \exp(-\sum_{m=1}^{M} h_m(x))}$$

(4)

where $M$ is the number of iterations with an input dataset of $Z = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ and $h_m(x)$ corresponds to single trees.

$$Pf_0(x) = \log\frac{p_1}{1 - p_1}$$

(5)

For the initial approximation $f_0(x)$ of the sigmoid function, $p_1$ is used as an object in the first class.

$$g_i = \frac{dL(y_i, f_m(x_i))}{d f_m(x_i)}$$

(6)

where $g_i$ is the target, historical flooded point data in this study. For the maximum of Equation (6), the gradients are calculated for all objects in the training dataset and adjusted in the gradient direction. A new decision tree $h_m(x)$ is created iteratively and forms the ensemble to maximize likelihood. For the boosted tree models in this study, the parameters were set as follows: learning rate = 0.01, tree complexity = 5 and bag fraction = 0.5.

### 3.5. Validation of flood susceptibility maps

The assessment of the results reliability was performed with the use of the Receiver Operating Characteristic (ROC) curve, which belongs to widely used methods in flood susceptibility studies (Chapi et al. 2017; Ali et al. 2020). The ROC curve
signifies the graphical association of the so-called specificity and sensitivity. The $x$ and $y$ axis in the ROC graph represent the false positive rate (1-specificity) and true positive rate (sensitivity), respectively (Equations (7 and 8)) (Costache et al. 2020):

$$x = 1 - \text{specificity} = 1 - \left(\frac{TN}{TN + FP}\right)$$

(7)

$$y = \text{sensitivity} = \left(\frac{TP}{TP + FN}\right)$$

(8)

where $TP$ (True Positive) is the number of points that are correctly classified as flood, $TN$ (True Negative) is the number of points that are correctly classified as non-flood, $FP$ (False Positive) is the number of non-flooded points classified as flood and $FN$ (False Negative) is the number of flooded points classified as non-flood. The specificity value signifies the relation of accurately predicted non-flood points (true negative) while the sensitivity value represents the accurately predicted flooded points

| Factor                          | Interval             | Reclassification | Share on basin area (%) | Relative importance |
|---------------------------------|----------------------|-------------------|--------------------------|---------------------|
| Slope ($^\circ$)                | 0.0–2.0              | 5                 | 11.60                    | 7                   |
|                                 | 2.1–5.0              | 4                 | 14.03                    |                     |
|                                 | 5.1–15.0             | 3                 | 51.45                    |                     |
|                                 | 15.1–35.0            | 2                 | 22.82                    |                     |
|                                 | 35.1–57.2            | 1                 | 0.09                     |                     |
| River network density (km/km²)  | 0.00–0.76            | 1                 | 10.81                    | 6                   |
|                                 | 0.77–1.16            | 2                 | 24.51                    |                     |
|                                 | 1.17–1.54            | 3                 | 30.44                    |                     |
|                                 | 1.55–1.97            | 4                 | 23.68                    |                     |
|                                 | 1.98–3.19            | 5                 | 10.56                    |                     |
| Distance from river (m)         | 0–152                | 5                 | 41.99                    | 5                   |
|                                 | 153–333              | 4                 | 33.17                    |                     |
|                                 | 334–583              | 3                 | 18.40                    |                     |
|                                 | 584–1129             | 2                 | 5.69                     |                     |
|                                 | 1129–2341            | 1                 | 0.76                     |                     |
| Flow accumulation (number of cells) | 0–200               | 1                 | 96.58                    | 4                   |
|                                 | 201–400              | 2                 | 0.75                     |                     |
|                                 | 401–2000             | 3                 | 1.27                     |                     |
|                                 | 2001–40,000          | 4                 | 1.13                     |                     |
|                                 | 40,001–911,409       | 5                 | 0.27                     |                     |
| Elevation (m a. s. l.)          | 105–300              | 5                 | 34.96                    | 3                   |
|                                 | 301–500              | 4                 | 39.82                    |                     |
|                                 | 501–700              | 3                 | 17.26                    |                     |
|                                 | 701–900              | 2                 | 6.79                     |                     |
|                                 | 901–1131             | 1                 | 1.17                     |                     |
| Curve number                    | 56–59                | 1                 | 0.81                     | 2                   |
|                                 | 60–69                | 2                 | 66.66                    |                     |
|                                 | 70–74                | 3                 | 1.20                     |                     |
|                                 | 75–79                | 4                 | 29.26                    |                     |
|                                 | 80–91                | 5                 | 2.08                     |                     |
| Lithology (degree of permeability) | Very high        | 1                 | 11.82                    | 1                   |
|                                 | High                 | 2                 | 16.34                    |                     |
|                                 | Moderate             | 3                 | 19.40                    |                     |
|                                 | Low                  | 4                 | 0.40                     |                     |
|                                 | Very low             | 5                 | 52.04                    |                     |

Table 3. Flood conditioning factors and their reclassification.
In addition, higher value of 1-specificity represents a high degree of false positives (Khosravi et al. 2019). The value of the ROC ranges between 0 and 1. The model is defined as efficient when the value of ROC is close to 1.

4. Results

4.1. Reclassification and weighting of flood conditioning factors

The result of the reclassification of flood conditioning factors, based on their susceptibility to flooding, is presented in Table 3.

Table 3. The slope factor was given the highest relative importance. The most susceptible, especially to riverine flooding, are slopes from 0 to 2°, which were included in very high reclassification class (5). On the contrary, the lowest reclassification value of 1 was assigned to highest slopes from 35.1 to 57.2°. Regarding the second most important factor of river network density, the highest reclassification value of 5 was assigned to the interval 1.98–3.19 km/km² while the very low reclassification class (1) was represented by the lowest river network density ranging from 0.00 to 0.76 km/km². Distance from river was the third most important factor. The lowest values of this factor (interval 0–152 m) were reclassified as the most susceptible to flooding (5) while the farthest distances from river (interval 1129–2341 m) were assigned the very low reclassification class (1). Flow accumulation was the forth most important factor. It was reclassified in a similar way as the river network density factor, i.e. the lowest (number of cells from 0 to 200) and highest (number of cells from 40,001 to 911,409) values corresponded to very low (1) and very high (5) reclassification classes, respectively. Elevation was the fifth most important factor. The lowest
reclassification value of 1 corresponded to the highest elevations from 901 to 1131 m. The most susceptible to flooding were elevations from 105 to 300 m, which were included in very high reclassification class (5). On the other hand, the lowest values of curve number factor (interval 56–59) were assigned the very low (1) reclassification class while the highest values of this factor (interval 80–91) were included into the very high (5) reclassification class. The least importance was given to the lithology (degree of permeability) factor. The very high degree of rock permeability was included in the very low reclassification class (1) while the very low degree of rock permeability was assigned the very high reclassification class (5).

Table 4 presents the pair-wise comparisons of the flood conditioning factors in a square reciprocal $7 \times 7$ matrix and the normalized weights of factors. The slope factor recorded the highest weight ($w_i$) with the value of 0.35. The other factors had the following weights: river network density ($w_i = 0.24$), distance from river ($w_i = 0.16$), flow accumulation ($w_i = 0.11$), elevation ($w_i = 0.07$), curve number ($w_i = 0.05$) and lithology ($w_i = 0.03$).

Figure 8. Flood susceptibility map using the MCDA-AHP based analysis.
4.2. MCDA-AHP model vs. machine learning-BT models

The resulting flood susceptibility maps were classified according to the quantile classification method. Figure 8 shows the flood susceptibility map performed by the MCDA-AHP based analysis. The computed flood susceptibility map based on the boosted classification tree model is shown in Figure 9 and the result of the boosted regression tree model is shown in Figure 10.

Regarding the MCDA-AHP model (Figures 8 and 11), the largest share (22.1%) on the basin area can be seen in the moderate flood susceptibility class. This class is represented mainly by the transition areas between valleys and uplands with mostly moderate slopes. The low flood susceptibility class covers 19.3% of the basin area and it corresponds mainly to upland and mountainous areas with higher slopes. On the contrary, the high flood susceptibility class accounts for 19.8% of the basin area and it is represented mainly by valley slopes with moderate to low values of slope degree. The share of very high flood susceptibility class on the basin area is 20.1%, which spatially corresponds to valleys, especially their bottoms, and river floodplains with low slopes and elevations. To a large extent, the built-up areas and arable land are
included in the very high flood susceptibility class. These areas have been affected several times by the historical riverine floods causing severe damage especially to infrastructure. The lowest percentage share on the basin area was recorded by the very low flood susceptibility class (18.7%), which is predominantly represented by mountains with higher elevations, slopes and forest cover. As it can be seen in Figures 8 and 11, 40% of the Topla river basin is characterized by high to very high flood susceptibility and almost two thirds of the study area is included in the moderate to very high flood susceptibility. Conversely, very low to low flood susceptibility accounts for more than a third of the basin area.

As shown in Figures 9 and 11, the boosted classification tree model underestimated the very low flood susceptibility class with only 1.1% share on the basin area, which is less by 17.7% and 19.3% than in case of the MCDA-AHP and BRT models, respectively. This class occurs only as fragmented areas within the study area. As a result, the rest of the flood susceptibility classes (from low to very high) overestimate the percentage of flood susceptibility classes of the other two models (i.e. MCDA-AHP and BRT). In particular, their share is as follows: low (27.5%), moderate (26.7%), high (23.3%) and very high (21.5%). The low flood susceptibility class
corresponds mainly to mountainous areas with higher elevations, slopes and forest cover. As for the areas included in the moderate flood susceptibility class, their spatial dominance can be seen mainly in the northwestern part of the basin, where they are represented by upland to mountainous terrain with moderate to higher slopes and high share of forest cover. The high and very high flood susceptibility classes correspond mainly to valleys and river floodplains with low slopes and elevations.

Difference in the spatial distribution of high and very high flood susceptibility classes between the MCDA-AHP and BCT model is clearly visible especially in the southeastern part of the basin, where the very high class dominates in case of the MCDA-AHP model while the proportion of high and very high class in case of the BCT model is more or less the same. Applying the boosted classification tree model, 45% of Topla river basin is characterized by high to very high flood susceptibility, which is higher share by 5% and 7% than in case of the MCDA-AHP and BRT models, respectively.

When comparing the share of flood susceptibility classes on the basin area for the MCDA-AHP and BRT models (Figure 11), it can be seen that their percentages are similar. The very low flood susceptibility class of the BRT model recorded lower share only by 1.7%, as compared to the MCDA-AHP. In case of the low and moderate flood susceptibility classes, the difference is 2.7% and 2.4%, respectively. The very low and low flood susceptibility classes in case of the BRT model are mostly represented by mountain ridges and upland areas with high slopes and elevations and high share of forests. The moderate flood susceptibility class spatially binds mainly to transition areas between valleys and uplands with mostly moderate slopes. The lowest difference
was recorded in the high flood susceptibility class (only 0.8%). In case of the BRT model, this class is represented mainly by the valley positions. As for the very high flood susceptibility class, the BRT model recorded higher share by 1.2%, as compared to the MCDA-AHP model. The spatial occurrence of the very high flood susceptibility class is similar to the MCDA-AHP and BCT models, i.e. it corresponds to valley and river floodplain areas. However, when the three models are compared, for example, in the southeastern part of the basin, it can be seen that the occurrence and share of the very high flood susceptibility class decreases in case of the MCDA-AHP, BCT and BRT models, respectively. Based on Figures 10 and 11, 38% of the study area is characterized by high to very high flood susceptibility when applying the BRT model.

4.3. Predictor-importance of flood conditioning factors

In this study, predictor-importance values of each factor were calculated by summing up the drop in node impurity for BCT and for the resubstitution estimate for BRT over all nodes and trees; the value is determined relative to the largest sum over all factors, which is the most important factor in the process. It computes the sum for all factors so that the predictor-importance value of the factor that was not used in the modelling could be also computed.

Table 5 shows the predictor-importance values of each factor in the boosted tree modelling. The predictor importance was expressed as a distribution of values from 0 to 1, based on the maximum value, whereby the node-impurity values in the classification and the resubstitution-estimate in the regression were summed, respectively. The factors can be differentiated to decide the contribution of the factors to the prediction of the dependent variable of interest. To sum up, a factor with a predictor-importance value closer to 1 means that it is relatively more related to the occurrence of floods.

In Table 5, the top three variables were slope (1.0), curve number (0.9168) and distance from river (0.9092) in case of the BCT model. As for the BRT model, the curve number (1.0) was the most influential variable. The flow accumulation (0.7935) was the second most influential variable followed by slope (0.7253). The least influential factors in case of both BCT and BRT models were the river network density and lithology.
4.4. Validation of flood susceptibility maps

Validation of the performed flood susceptibility models was based on 30% of flood points referred to as testing points. Figure 12(a) shows the accuracy of the performed MCDA-AHP-based analysis, which recorded the value of 81.33%. The accuracy rate of the BCT model was 87.70%, as shown in Figure 12(b). The accuracy of the BRT model recorded the value of 91.42%, as shown in Figure 12(c). Based on the validation results, the best predictive capability was achieved by the BRT model while the predictive accuracy of the MCDA-AHP model was the lowest. However, it should be noted that all three models used in this study achieved acceptable accuracy for regional flood susceptibility mapping in the Topla river basin.

5. Discussion

The results achieved were interpreted also in perspective of previous similar studies. For instance, Toosi et al. (2019) applied the modified MCDA incorporating the hydrological model SWAT as well as using seven flood conditioning factors. Their results showed that 97% of flood events corresponds with moderate to very high flood susceptibility class, which is very similar finding as in this study, where 99% of flood events occurred in moderate to very high flood susceptibility class. Another study by Kanani-Sadat et al. (2019) suggests that fuzzy-based MCDA-analytical network process provides better accuracy (93.8%) than the AHP model (91.8%). Moreover, Khosravi et al. (2019) compared three MCDA techniques (VIKOR, TOPSIS and SAW) with two machine learning models (NBT and NB). The accuracy of all models was more than 95% while the NBT model achieved the accuracy of 98%. As a result, the machine learning methods were more accurate in predicting the flood susceptible areas, which is also in accordance with the results achieved in this study.

Tang et al. (2018) used six flood conditioning factors, the MCDA-local ordered weighted averaging (OWA) technique and Monte Carlo simulation to assess flood susceptible areas. Their results suggest that the MCDA should be improved by using the mentioned methods in order to provide more objective flood susceptibility evaluation. This is true also in this study because the applied MCDA is a knowledge-driven method, where it is necessary to know the solved problem very-well as well as

Figure 12. Accuracy rate of the applied flood susceptibility models: (a) MCDA-AHP, (b) BCT, (c) BRT.
the importance and impact of the criteria, which evaluate the achieved result. Regarding the MCDA, it is obvious that there is a strong subjective influence in determining factor weights as well as in selecting the experts. On the other hand, objectivity should be an essential characteristic of each expert. In addition, due to a certain level of subjectivity used in the MCDA, they generally achieve lower accuracy, as compared to the machine learning methods.

In case of the decision tree models, Lee et al. (2017) applied random-forest as well as boosted tree model for both variants (classification and regression). The random-forest models (regression 78.78% and classification 79.18%) resulted in higher accuracy than the boosted tree models (regression 77.55% and classification 77.26%). However, compared to the boosted tree models used in this study, they achieved lower predictive capability by more than 10%. Furthermore, Shafizadeh-Moghadam et al. (2018) compared several stand-alone and ensemble models. According to their findings, the boosted regression tree model recorded the highest accuracy (97.5%) among nine stand-alone machine learning models. However, the ensemble model to estimate the median achieved even slightly better accuracy (97.6%). In a similar study by Tehrany et al. (2019), two robust machine learning models, namely decision tree and support vector machine, were compared. The validation resulted in accuracies of 85.52% and 88.47% for support vector machine and decision tree model, respectively. Based on the aforementioned studies as well as on this study, it can be claimed that decision tree models belong to the most accurate machine learning models and, therefore, also to one of the most accurate methods for flood susceptibility mapping.

Furthermore, the flood susceptibility in the study area was analyzed by Ali et al. (2020), who compared four hybrid and standalone models, namely the Decision-making trial and evaluation laboratory (DEMATEL)-Analytic network process (ANP), Naïve Bayes tree (NBT)-Frequency ratio (FR), NBT-Statistical index (SI) and Logistic regression (LR), based on 14 conditioning factors and the same flood inventory map. When comparing the results of the ROC curves, the presented models achieved lower values by approximately 7%, 11% and 17% in case of the MCDA-AHP, BCT and BRT models, respectively, as compared to the DEMATEL-ANP model, which reached the lowest ROC value of 98.7%. The study by Ali et al. (2020) revealed that the share of very high and high flood susceptibility classes on the basin area varied between 20% and 44% for the presented models (NBT-FR = 19.89%, LR = 32.35%, DEMATEL-ANP = 32.52%, NBT-SI = 43.85%). In the presented study, however, this share has lower range and is more balanced: BRT = 38%, BCT = 40% and MCDA-AHP = 45%. Differences in the presented study and the study by Ali et al. (2020) were mainly caused by the use of different models as well as flood conditioning factors, which also influenced the spatial distribution of flood susceptibility zones in the Topl’a river basin. On the other hand, few similarities can be seen, in particular, for the multi-criteria models applied (MCDA-AHP and DEMATEL-ANP), where the spatial distribution of very high and high susceptibility is quite similar mainly in the southeastern part of the study area. Overall, the results of both studies agree on the fact that the southern and southeastern parts of the study area as well as valleys and river floodplains in central or northern parts are highly susceptible to fluvial floods. In general, these parts of the basin have low elevations, low slope angles, high flow
accumulation, very high curve numbers and are covered mainly by arable land and built-up areas.

An important part of the discussion section is to present the limitations of the methods as well as data used, which can be summarized in the following comments:

i. The first limitation arises from the selection and number of flood conditioning factors. It should be noted that there is no consensus among researchers on how many factors should be used to identify areas susceptible to floods (Das 2019). On the other hand, some studies provide recommendations in this respect. According to Mahmoud and Gan (2018), more than six factors should be used in MCDA-based flood susceptibility studies to avoid the overestimation of some of the factors. However, in literature there are several studies which used less than six factors, for example, Rahmati et al. (2016) or Samanta et al. (2016). On the other hand, Tehrany et al. (2019) claim that using more flood conditioning factors in modelling do not necessarily ensure that the model will be more accurate. Consequently, they emphasize that the results can be significantly influenced by the modelling process rather than the number of factors.

ii. Another limitation can be seen in the data sources used for processing flood conditioning factors. Having input data with the same original map scale or spatial resolution would be the optimal case. However, this is difficult to achieve in practice and thus the produced maps of flood conditioning factors can be generalized to a certain extent depending on the accuracy of the input data.

iii. Regarding the flood inventory database, it would be suitable to extend the database of flood events for the study area including more flood events, which would allow better comparison and validation of the resulting models.

iv. As already mentioned in this section, the limitation of the MCDA-AHP based analysis is that it is the knowledge-driven method, where the degree of subjectivity used in weighting the factors by the experts from the research topic is high, as compared to more ‘objective’ data-driven models, such as decision trees.

v. Boosting method is processed sequentially and the current weak learner is affected by the previous weak learner, i.e. the weighting of the error is raised if an error occurs. Therefore, the boosted tree models are suitable to model the data scarce areas, but they cannot be free from overfitting problems which require considering the characteristics of the data.

vi. A limitation of the presented methodology can be seen in its applicability mostly to potential fluvial (riverine) flooding. In order to count with other types of flood, such as flash flood, pluvial flood or sheetflood, the methodology (e.g. selection of factors and their reclassification) should be modified and adapted.

The future research can be directed towards the following issues:

- Minimizing the mentioned limitations (i)–(vi) of the methods and data used as much as possible, i.e. resolving the issues of model complexity, accuracy (quality) of input data and number of factors.
Comparing more MCDA techniques, statistical methods and machine learning algorithms as well as their ensembles to find the most acceptable and accurate method for flood susceptibility mapping in the Topla river basin.

All in all, the results suggest that effective approaches to flood mitigation should include preventing the urban expansion and new construction in high to very high flood susceptible zones as well as applying the flood risk management plans and flood control solutions based on the topographical characteristics of the basin (Khosravi et al. 2020). In this regard, the potential of nature-based solutions (Kalantari et al. 2018; Keesstra et al. 2018) to mitigate the flood risk and problems in urban development under climate change should be considered as well as interconnected with spatial planning and management strategies (Vojtek and Vojteková 2018). This study presents an approach which can be easily replicated in other basins as well as continuously updated in case of changes in flood conditioning factors over the time, thus providing a sustainable non-structural measure, which could be incorporated into integrated basin management plans (Kundzewicz 1999). The paradigm shift in flood protection from an engineering approach to an integrated, holistic and sustainable approach, which focuses on the whole river basin, increases the importance of flood maps (Werrity 2006). In this sense, the identification (mapping) of flood susceptibility/hazard/risk can provide the essential information for the selection of appropriate strategies and measures to avoid, mitigate, share or accept the risk (Plate 2002).

6. Conclusions

Flooding and the minimization of flood damage is highly topical issue which requires attention due to the increasing frequency and magnitude of flood situations and the increasing extent of flood damage, which is associated mainly with the ongoing climate change and increasing anthropic pressure on the landscape. Different measures are used in flood protection, whether structural or non-structural, each of which plays an important role. As long as the absolute flood protection is not possible, mainly flood prevention, management and mitigation should be prioritized, including the identification and assessment of areas susceptible to floods.

In this study, two different approaches, i.e. MCDA-AHP technique and machine learning-boosted tree models, were used for identification of areas susceptible to flooding based on seven flood conditioning factors and flood inventory database consisting of 107 flood locations. Based on the results, 40%, 45% and 38% of the study area, respectively, for the applied MCDA-AHP, BCT and BRT models, is characterized by high to very high flood susceptibility. Validation of the performed flood susceptibility models confirmed higher accuracy of the machine learning models, as respect to the multi-criteria-based model. In particular, the accuracy of the MCDA-AHP model was 81.33% while the accuracy of the boosted tree models was 87.70% and 91.42%, respectively, for classification and regression. Despite this fact, the accuracy of all three models is acceptable for regional flood susceptibility modelling.

The results achieved in this study may be useful for the next revision of the preliminary flood risk assessment and the associated identification of geographical areas
with potentially significant flood risk according to the Directive 2007/60/EC and Act no. 7/2010. Moreover, the identification of areas with high and/or very high flood susceptibility is important for the subsequent and more detailed determination of flood hazard, defined by the flood extent, flow depth or velocity, using sophisticated hydraulic (1D or 2D) models.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-18-0185.

**ORCID**

Matej Vojtek http://orcid.org/0000-0001-9369-3173  
Jana Vojteková http://orcid.org/0000-0002-8904-9673  
Quoc Bao Pham http://orcid.org/0000-0002-0468-5962  
Duong Tran Anh http://orcid.org/0000-0002-6775-0055

**References**

Act no. 7/2010 Coll. of 2 December 2009 on the flood protection. [accessed 2020 Nov. 23].  
https://www.enviroportal.sk/uploads/2012/04/page/environmental-temy/rizi/zakon_7_2010.pdf

Ali SA, Parvin F, Pham QB, Vojtek M, Vojteková J, Costache R, Linh NTT, Nguyen HQ, Ahmad A, Ghorbani MA. 2020. GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision making approach, naïve Bayes tree, bivariate statistics and logistic regression: a case of Topla basin, Slovakia. Ecol Indic. 117:106620.

Bochniček O, Borsányi P, Čepčeková E, Faško P, Chmelík M, Jančovičová L, Kapolková H, Labudová I, Mikulová K, Mišaga O, et al. 2015. Climate Atlas of Slovakia. Bratislava: SHMU.

Bui TD, Khosravi K, Li S, Shahabi H, Panahi M, Singh VP, Chapi K, Shirzadi A, Panahi S, Chen W, et al. 2018. New hybrids of ANFIS with several optimization algorithms for flood susceptibility modeling. Water. 10(9):1210.

Bui DT, Khosravi K, Shahabi H, Daggupati P, Adamowski JF, Melesse AM, Pham BT, Pourghasemi HR, Mahmoudi M, Bahrami S, et al. 2019. Flood spatial modeling in Northern Iran using remote sensing and GIS: a comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. Remote Sens. 11(13):1589.

Carver SJ. 1991. Integrated multi-criteria evaluation with geographical information systems. Int J Geogr Inf Sys. 5(3):321–339.

Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ Model Soft. 95:229–245.

Chow VT. 1964. Handbook of applied hydrology. New York: McGraw-Hill Book Company.

Costache R. 2019. Flood susceptibility assessment by using bivariate statistics and machine learning models - a useful tool for flood risk management. Water Resour Manage. 33(9):3239–3256.
Costache R, Pham QB, Corodescu-Roșca E, Cimpianu C, Hong H, Linh NTT, Fai CM, Ahmed AN, Vojtek M, Pandhiani SM, et al. 2020. Using GIS, remote sensing, and machine learning to highlight the correlation between the land-use/land-cover changes and flash-flood potential. Remote Sens. 12(9):1422.

Cronshey R, McCuen RH, Miller N, Rawls W, Robbins S, Woodward D. 1986. Urban hydrology for small watersheds (TR-55). Washington: Natural Resources Conservation Service.

Čurlík J, Šály R. 2002. Zrnitost pôdy. In: Hrnčiarová T, editor. Atlas krajiny Slovenskej republiky. Banská Bystrica: MŽP SR, Bratislava and SAŽP; p. 111.

Dano UL, Balogun AL, Matori AN, Yusouf KW, Abubakar IR, Mohamed MAS, Aina YA, Pradhan B. 2019. Flood susceptibility mapping using GIS-based analytic network process: a case study of Perlis, Malaysia. Water. 11(3):615.

Das S. 2018. Geographic information system and AHP-based flood hazard zonation of Vaitarna basin, Maharashtra, India. Arab J Geosci. 11(19):56–74.

De’Ath G. 2007. Boosted trees for ecological modeling and prediction. Ecology. 88(1):243–251.2.0.CO;2

Demek J. 1972. Manual of detailed geomorphological mapping. Prague: Academia.

Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. [accessed 2020 Nov. 23]. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:288:0027:0034:SK:PDF.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. J Anim Ecol. 77(4):802–813.

Elkhrachy I. 2015. Flash flood hazard mapping using satellite images and GIS tools: a case study of Najran City, Kingdom of Saudi Arabia (KSA). Egyp J Remote Sens Space Sci. 18(2):261–278.

Friedman JH. 2002. Stochastic gradient boosting. Comput Stat Data Anal. 38(4):367–378.

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H, Kanae S. 2013. Global flood risk under climate change. Nat Clim Change. 3(9):816–821.

Hrasna M, Klukanová A. 2002. Inžinierskogeologická rajonizácia. In: Hrnčiarová T, editor. Atlas krajiny Slovenskej republiky. Banská Bystrica: MŽP SR, Bratislava and SAŽP; p. 83.

Hrnčiarová T. 1993. Hodnotenie potenciálnej zraniteľnosti zásob podzemných vôd pre ekologické plánovanie krajiny. Zivotné Prostredie. 27:311–314.

Hutchinson MF. 1988. Calculation of hydrologically sound digital elevation models. In Third International Symposium on Spatial Data Handling; Sydney: International Geographical Union, Columbus.

Jacinto R, Grosso N, Reis E, Dias L, Santos FD, Garrett P. 2015. Continental Portuguese territory flood susceptibility index – contribution to a vulnerability index. Nat Hazards Earth Syst Sci. 15(8):1907–1919.

Kalantari Z, Ferreira CSS, Keesstra S, Destouni G. 2018. Nature-based solutions for flood-drought risk mitigation in vulnerable urbanizing parts of East-Africa. Curr Opin Environ Sci Health. 5:73–78.

Kanani-Sadat Y, Arabsheibani R, Karimipour F, Nasseri M. 2019. A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. J Hydrol. 572:17–31.

Keesstra SD. 2007. Impact of natural reforestation on floodplain sedimentation in the Dragonja basin, SW Slovenia. Earth Surf Process Landforms. 32(1):49–65.

Keesstra S, Nunes J, Novara A, Finger D, Avelar D, Kalantari Z, Cerda A. 2018. The superior effect of nature based solutions in land management for enhancing ecosystem services. Sci Total Environ. 610–611:997–1009.

Keesstra SD, Van Dam O, Verstraeten GV, Van Huissteden J. 2009. Changing sediment dynamics due to natural reforestation in the Dragonja catchment, SW Slovenia. Catena. 78(1):60–71.
Khosravi K, Panahi M, Golkarian A, Keesstra SD, Saco PM, Bui DT, Lee S. 2020. Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. J Hydrol. 591:125552.

Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly HB, Gróf G, Ho HL, et al. 2019. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol. 573:311–323.

Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A. 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci. 67(1):251–264.

Kourgialas NN, Karatzas GP. 2011. Flood management and a GIS modelling method to assess flood-hazard areas – a case study. Hydrol Sci J. 56(2):212–225.

Kundzewicz ZW. 1999. Flood protection—sustainability issues. Hydrol Sci J. 44(4):559–571.

Lapin M, Faško P, Melo M, Šťastný P, Tomlайн J. 2002. Climatic regions. In: Hrnčiarová T, editor. Atlas krajiny Slovenskej republiky. Banská Bystrica: MŽP SR, Bratislava and SAŽP; p. 95.

Lee S, Kim JC, Jung HS, Lee MJ, Lee S. 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics Nat Hazards Risk. 8(2):1185–1203.

Lee S, Lee S, Lee MJ, Jung HS. 2018. Spatial assessment of urban flood susceptibility using data mining and Geographic Information System (GIS) tools. Sustainability. 10(3):648.

Mohammady M, Moradi HR, Zeinivand H, Temme AJ, Yazdani MR, Pourghasemi HR. 2018. Modeling and assessing the effects of land use changes on runoff generation with the CLUE-s and WetSpa models. Theor Appl Climatol. 133(1-2):459–471.

Mosavi A, Ozturk P, Chau K. 2018. Flood prediction using machine learning models. Literature review. Water. 10(11):1536.

Plate E. 2002. Flood risk and flood management. J Hydrol. 267(1-2):2–11.

Pradhan B. 2010. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spatial Hydrol. 9(2):1–18.

Rahmati O, Zeinivand H, Besharat M. 2016. Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomatics Nat Hazards Risk. 7(3):1000–1017.

Saaty TL. 1980. The analytic hierarchy process. New York: McGraw Hill Company.
Siahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O. 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. Geocarto Int. 33(9):1–15.

Smith K, Ward R. 1998. Floods: physical processes and human impacts. Chichester: John Wiley & Sons.

Sofia G, Roder G, Dalla Fontana G, Tarolli P. 2017. Flood dynamics in urbanised landscapes: 100 years of climate and humans’ interaction. Sci Rep. 7:40527.

Solín L. 2008. Analýza výskytu povodňových situácií na Slovensku v období rokov 1996-2006. J Hydrol Hydromech. 56(2):95–115.

Souissi D, Zouhri L, Hammami S, Msaddek MH, Zghibi A, Dlala M. 2020. GIS-based MCDM-AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia. Geocarto Int. 35(9):991–1017.

Tang Z, Zhang H, Yi S, Xiao Y. 2018. Assessment of flood susceptible areas using spatially explicit, probabilistic multi-criteria decision analysis. J Hydrol. 558:144–158.

Tehrany MS, Jones S, Shabani F. 2019. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. Catena. 175:174–192.

Tehrany MS, Pradhan B, Jebur MN. 2015. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch Environ Res Risk Assess. 29(4):1149–1165.

Toosi AS, Calbimonte GH, Nouri H, Alaghmand S. 2019. River basin-scale flood hazard assessment using a modified multi-criteria decision analysis approach: a case study. J Hydrol. 574:660–671.

Vojtek M, Vojteková J. 2018. Flood maps and their potential role in local spatial planning: a case study from Slovakia. Water Policy. 20(5):1042–1058.

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, Pamučar D, Gigović L, Drobnjak S, Bui DT, Duan H. 2018. A hybrid GIS multi-criteria decision-making method for flood susceptibility mapping at Shangyou, China. Remote Sens. 11(1):62.

Werrity A. 2006. Sustainable flood management: oxymoron or new paradigm? Area. 38:16–23.

Yousefi S, Mirzaei S, Keesstra S, Surian N, Pourghasemi HR, Zakizadeh HR, Tabibian S. 2018. Effects of an extreme flood on river morphology (case study: Karoon River, Iran). Geomorphology. 304:30–39.

Zardari H, Ahmed K, Shirazi SM, Yusop ZB. 2015. Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management. London: Springer.