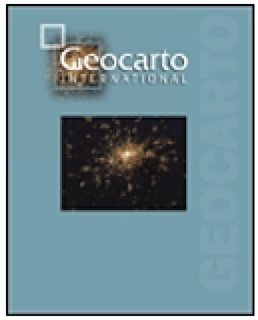
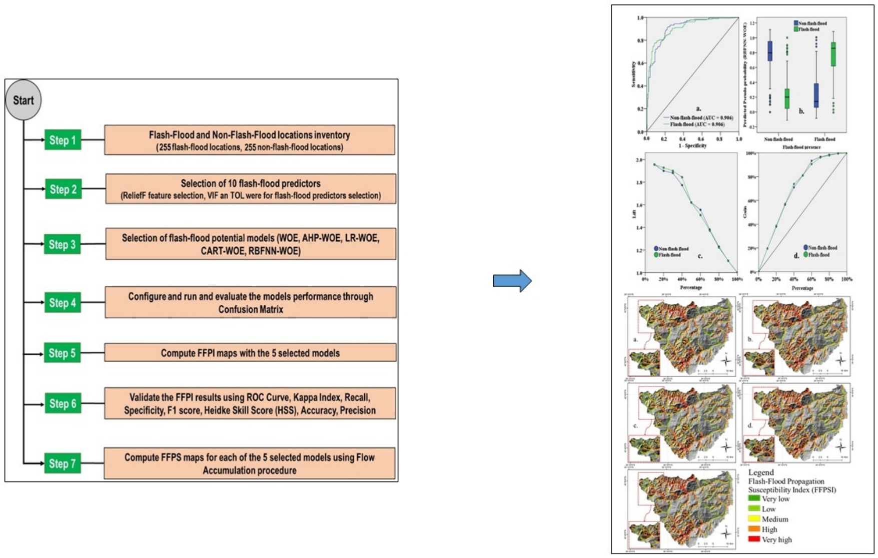


Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning

Journal:	*Geocarto International*
Manuscript ID	TGEI-2021-0559.R1
Manuscript Type:	Research Article
Keywords:	flash-floods propagation susceptibility, bivariate statistics, multicriteria decision-making, machine learning, Romania



338x190mm (96 x 96 DPI)

- Flash-Flood Propagation Susceptibility Index (FFPSI) was proposed and calculated;
- A number of 255 flash-flood locations were used for modelling;
- A number of 10 flash-flood predictors were used to estimate the susceptibility;
- One stand-alone and four ensembles were used to derive the Flash-Flood Potential;
- Flow Accumulation was used to derive the final FFPSI maps.

Ms. Ref. No.: TGEI-2021-0559 (218569821)

Title: “Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning”

Journal: Geocarto International

Dear Dr Rundquist
Regional Editor
Geocarto International,

Thank you very much for giving us a chance to revise the manuscript. We would like to express our gratitude towards the Editor and Reviewers for their valuable comments and suggestions that helped to improve the original submitted manuscript.

In this revised version, we carefully considered all the comments from Editor and the three reviewers, point by point. The following paragraphs include a point – to – point response to reviewers’ comments and suggestions. We hope that the revised manuscript will satisfy the Reviewers and Editor. We kindly request you to consider this manuscript for publication in your esteemed journal.

Best regards,

The authors

List of changes in the revised paper:

This document explains the changes made in the revised manuscript while dealing with the comments raised by the reviewers. Reviewers’ comments are marked in **black**; authors’ response is shown in **blue;** in **green,** we provide the revised text, while the changes in the revised manuscript are marked in **red**.

Response to Reviewers Reviewer #1

General reviewer comment: **“**The current research uses some methods for flood susceptibility mapping. There are many papers that used more robust machine learning models in previous studies. My main criticism is about novelty of paper.”

Authors: We would like to express our sincere gratitude to the reviewer for providing rewarding and constructive feedback. We have carefully read and addressed all comments, point by point, below. We also mentioned which are the main elements of novelty. We copy the text below:

“The main element of novelty that characterizes this study is represented by the use and computation for the first time in the literature of Flash-Flood Propagation Susceptibility Index (FFPSI), which is of a real help to create a complete overview regarding the flash-flood susceptibility at the level of a river catchment. Another element of novelty is represented by the use for the first time in the literature of the following ensemble models in order to determine the flash-flood susceptibility: AHP-WOE and RBFNN-WOE.” – line 694

Specific reviewer comments:

Reviewer comment 1: “What is the difference of your study with previous studies such as: "Towards a flood vulnerability assessment of watershed using integration of decision-making trial and evaluation laboratory, analytical network process, and fuzzy theories."; "Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins."; "Incorporating multi-criteria decision-making and fuzzy-value functions for flood susceptibility assessment."; "Integrated machine learning methods with resampling algorithms for flood susceptibility prediction.""

Author response: We thank the reviewer’s comment. We tried to explain which are the main differences between our study and the mentioned studies in the Discussion section. We copy the text below:

“It should be noted that the previous studies regarding the estimation of flash-flood susceptibility by machine learning techniques, carried out so far, did not include the study of the susceptibility of the valleys to the propagation of flash-flood waves (Anquetin et al., 2010; Janizadeh et al., 2019). Moreover, many researchers were focused, in their previous works, only on the evaluation of normal flood susceptibility (Azareh et al., 2019; Dodangeh et al., 2020; Hosseini et al., 2021; Mosavi et al., 2020), without taking into account the flash-flood phenomenon particularities. Besides of the previous research works which take into account only the local flood susceptibility given by the punctual conditions and rainfall, this article proposes a new and complete approach regarding the study of slopes susceptibility to runoff and, also, regarding the susceptibility of valleys to the propagation of the flash-floods. Therefore, through FFPSI, for each valley across the study area are highlighted the characteristics of the upslope catchment that could determine a high exposure to flash-flood. This new approach was conducted with the help of bivariate statistics and machine learning and also using the Flow Accumulation procedure. In fact, the propagation of the flash-flood wave is the element that generates the most significant material damage and loss of human life (Mujumdar, 2001)." – line 626

Reviewer comment 2: "The return period of flood locations (y-variable) is not clear.".

Author response: We thank the reviewer for the comment. We add an explanation within the text. Please find the text below:

"It should be noted that the majority of identified flash-floods were determined by the river discharge values with a return period of 10 years. Though, it is important to mention that the return period couldn't be established for each flash-flood event because the phenomena occurred on river sectors without hydrometric measurements." – line 150

Reviewer comment 3: "L 207: How did you collect the flood locations?"

Author response: We thank the reviewer for the comment. We copy below the explanation:

"Thus, in the present case, in order to evaluate the susceptibility of the surfaces to the genesis of the flash-floods, data regarding the places where these phenomena occurred in the past were collected. In this regard, the damage reports provided by the General Inspectorate for Emergency Situation (GIES) of Romania and the information from mass-media were used. Totally, a number of 255 de flash-flood locations were collected." – line 145

Response to Reviewers Reviewer #2

Reviewer comment 1: "The authors discussed the research gap (novelty) in the Discussion and Conclusion sections. However, a research gap statement should be written in the Introduction section to succinctly inform your audience what holes in the literature your research is trying to fill."

Author response: We thank reviewer for taking his time to review our manuscript and for the valuable suggestions. We added a statement in the Introduction in order to inform the audience about the novelty of our study. We copy the text below:

"In this context, the present study wants to propose a methodology for estimating the susceptibility to flash-flood propagation, this topic not being addressed so far in the literature." – line 112

Reviewer comment 2: "Spatial modelling using AHP elicits the experts' opinions in the area because the experts' rankings are used to calculate the factors' weights. So, the data used in the AHP computations were assumed to be obtained through a questionnaire survey filled in by several well-versed experts in the field. However, the author(s) didn't mention how many experts were interviewed, their expertise, and their years of experience. Moreover, the authors should give justification for using a specific number of experts. See the study below as an example:

Dano, U. L. (2021). An AHP-based assessment of flood triggering factors to enhance resiliency in Dammam, Saudi Arabia. GeoJournal, 1-16.

Dano, U. L. (2020). Flash Flood Impact Assessment in Jeddah City: An Analytic Hierarchy Process Approach. Hydrology, 7(1), 10."

Author response: We thank the reviewer for the comment. We added new explanations within the text and also we cited the mentioned research papers. We copy the text below:

"It should be mentioned that for the present study, the data necessary for the application of AHP method was obtained through an expert-based questionnaire survey administered to a number of 19 experts from the National Institute of Hydrology and Water Management of Romania, with a high expertise in flash-flood risk assessment. The number of interviewed experts is very close to the number that was also used in previous works from the literature (Dano, 2021, 2020)."- line 316

Reviewer comment 3: “A separate table containing spatial and non-spatial data and their sources with dates and scales where applicable should be provided.”

Author response: We thank the reviewer for the valuable suggestion. We added a new table in order to meet your requirement.

Table 1 Data used, source, resolution, scale and type

Data	Source	Resolution	Scale	Type
Digital Elevation Model (DEM)	Shuttle Radar Topography Mission (SRTM)	30 m	-	Spatial
Flash-Flood points	General Inspectorate for Emergency Situation (GIES) of Romania; mass-media	-	-	Spatial
Non-Flash-Flood points	Aerial imagery; field survey	-	-	Spatial
Rainfall (mm/year)	Worldclim v2	-	-	Spatial
Land use/cover	Corine Land Cover, 2018	1 km	-	Spatial
Hydrological Soil Groups	Digital Soil Map of Romania	-	1:200000	Spatial
Lithology	Digital Geological Map of Romania	-	1:200000	Spatial

Reviewer comment 4: “The author(s) didn’t compare their findings with prior studies in the field. The Discussion section should highlight important discoveries and how they support/corroborate or differ from previous studies and likely explanations by citing recent literature.”

Author response: We thank the reviewer for his valuable suggestion. We added new paragraph in which we compare our study and discoveries with the previous research works and also we explain how our results differ from the previous studies. We copy the text below:

“It should be noted that the previous studies regarding the estimation of flash-flood susceptibility by machine learning techniques, carried out so far, did not include the study of the susceptibility of the valleys to the propagation of flash-flood waves (Anquetin et al., 2010; Janizadeh et al., 2019). Moreover, many researchers were focused, in their previous works, only on the evaluation of normal flood susceptibility (Azareh et al., 2019; Dodangeh et al., 2020; Hosseini et al., 2021; Mosavi et al., 2020), without taking into account the flash-flood phenomenon particularities. Besides of the previous research works which take into account only the local flood susceptibility given by the punctual conditions and rainfall, this article proposes a new and complete approach regarding the study of slopes susceptibility to runoff and, also, regarding the susceptibility of valleys to the propagation of the flash-floods. Therefore, through FFPSI, for each valley across the study area are highlighted the characteristics of the upslope catchment that could determine a high exposure to flash-flood. This new approach was conducted with the help of bivariate statistics and machine learning and also using the Flow Accumulation procedure. In fact, the propagation of the flash-flood wave is the element that generates the most significant material damage and loss of human life (Mujumdar, 2001).” – line 626

Reviewer comment 5: “Figure 1 source should be provided.”

Author response: We added the source of Figure 1.

Response to Reviewers Reviewer #3

General reviewer comment: “The paper " Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning” covers an interesting and actual topic. However, the paper can't be accepted for publication before some changes. Below are listed specific comments:”

Authors: We would like to express our sincere gratitude to the reviewer for providing rewarding and constructive feedback. We have carefully read and addressed all comments, point by point, below.

Specific reviewer comments:

Reviewer comment 1: “According to international procedures, abbreviations should be introduced the first time they are mentioned in brackets.”

Author response: We thank the reviewer for valuable suggestions. We introduced the abbreviations their first time mention in text.

Reviewer comment 2: “line 108-109 “Another sample of 255 points were placed in areas where the flash-floods did 109 not occur in the past.” / How can we make sure that these 255 points are non-flash-flood locations or in other words on which basis the authors considered them as non-flash-flood locations maybe through field trips or by processing satellite images, please precise?"

Author response: We thank the reviewer for the comment. We added new explanations within the text. We copy the text below:

"It is worth to note that, along the information from the governmental authorities that didn't mention the occurrence of flash-flood events, the non-flash-flood locations were also placed based on the analysis satellite images and field surveys." – line 155

Reviewer comment 3: "line 631-633 "The main element of novelty that characterizes this study is represented by the use and computation for the first time in the literature of Flash-Flood Propagation Susceptibility Index (FFPSI)" / we can find in the literature many studies which introduced various flood Susceptibility Indices"

Author response: We thank the reviewer for the comment. Indeed, there are many studies which introduced various flood Susceptibility Indices. Though, this article proposes an approach that include both the susceptibility to surface runoff at the slope level, and then, with the help of Flow Accumulation procedure, the susceptibility of valleys to the flash-flood propagation was also evaluated. We added new explanations in the text:

"Besides of the previous research works which take into account only the local flood susceptibility given by the punctual conditions and rainfall, this article proposes a new and complete approach regarding the study of slopes susceptibility to runoff and, also, regarding the susceptibility of valleys to the propagation of the flash-floods. Therefore, through FFPSI, for each valley across the study area are highlighted the characteristics of the upslope catchment that could determine a high exposure to flash-flood. This new approach was conducted with the help of bivariate statistics and machine learning and also using the Flow Accumulation procedure. In fact, the propagation of the flash-flood wave is the element that generates the most significant material damage and loss of human life (Mujumdar, 2001)." – line 632

Reviewer comment 4: "The FFPSI introduced in this study using the flow accumulation map classified the degree of susceptibility at river level not for the entire basin. What is the interest of this index? It seems like a kind of clipping of the maps shown in fig 7 by the Basin Rivers because we know that the flow accumulation map shows a very high contrast between the watercourses and the other lands. In fact, upon visual inspection, it is obvious that the values of both indices (FFPI and FFPSI) are the same or at least proportional."

Author response: We thank the reviewer for the comment. In Fig. 7 there is represented the surface runoff potential at the slope level which in many parts of the study area has a high value. Nevertheless, the FFPI represented in Fig. 7 didn't show the same high values on the valleys that are near the slopes with a high susceptibility. This is the reason for which the Flow Accumulation was applied because the valleys will have the FFPI values that were weighted on the upslope catchment area and the new FFPSI (Fig. 12) will better show the potential power of a flash-flood event at the river valley level. The Flow Accumulation procedure was applied through an workflow developed in ArcGIS software in which the input data was represented by the Flow Accumulation derived from Digital Elevation Model and the raster of FFPI. We would like to ensure the reviewer that we didn't simply clip the FFPI Raster along the river valleys.

Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning

Abstract: The present study aims to enrich the specialized literature by proposing and calculating a new flash-flood propagation susceptibility index (FFPSI). Thus, firstly the Flash-Flood Potential Index (FFPI) using the ensembles of the next models was calculated: Weights of Evidence (WOE), Analytical Hierarchy Process (AHP), Logistic Regression (LR), Classification and Regression Trees (CART), and Radial Basis Function Neural Network-Weights of Evidence (RBFN-WOE). A number of 255 flash-flood locations, split into training (70%) and validating (30%) samples, along with 10 predictors were used as input in the five models. The Receiver Operating Characteristics (ROC) Curve and several statistical metrics were used to evaluate the Flash-Flood Potential Index results. LR-WOE and AHP-WOE were the most performant models. Nevertheless, all the applied models performed very well (AUC > 0.85). Further, the FFPSI was determined by integrating the FFPI results into a Flow Accumulation procedure. Over 55% of the valleys identified are characterized by high and very high values of FFPSI.

Keywords: flash-floods propagation susceptibility; bivariate statistics; multicriteria decision-making; machine learning; Romania

1. Introduction

The current context marked by the imminent transition from the moderate meteorological phenomena to the meteorological phenomena characterized by extreme severity, brings into the discussion the necessity of some urgent adaptation measures to combat the extreme weather negative effects. This transformation of meteorological phenomena at the planetary scale, mainly due to the global climate change (Markolf et al., 2019), entails an exponential increase in the frequency of risk hydrological phenomena such as flash-floods and floods generated by them (Antronico et al., 2019). Currently, according to Hofman and Schüttrumpf (2019) flash-floods are considered among the most devastating natural hazards. Globally, the total number of victims annually caused by these phenomena between 1996 and 2015 is estimated at 150061 (Costache et al., 2020c). This is due to the very high speed of appearance and manifestation, which varies from a few tens of minutes to a maximum of 6 hours (Lee and Kim, 2019), as well as the violence with which the mechanical action of water affects the socio-economic and environmental elements during such a phenomenon. Flash-floods are generally characteristic for river basins with a medium to high relief slope and a small surface. These 2 elements determine a very short time of water concentration from the slopes towards the river channels (Costache, 2014a; Prăvălie and Costache, 2013). It is obvious that where the flash-flood wave along a river valley meets an area with a lower slope will eventually generates a devastating flood. One of the most effective non-structural measures taken to mitigate the flash-flood effects is represented by the exact identification of the surfaces on which the surface runoff on the slopes is manifested. Additionally, it is mandatory to detect the river valleys along which a high potential for flash-flood propagation exists.

The accelerated development of computerized techniques has created favourable premises for the application of modern methodologies that can allow the rapid and high accurate assessment of the above-mentioned surfaces. In this regard, GIS techniques are widely used to map the areas susceptible to flash-floods (Costache, 2014b; Zaharia et al., 2017, 2015). An increasing number of researchers are trying to integrate these GIS techniques with advanced computational algorithms that are specific to bivariate statistics, artificial intelligence and multi-criteria decision-making (Al-Abadi, 2018; Ali et al., 2020; Arabameri et al., 2020; Costache, 2019). Among the most used bivariate statistical techniques found in studies focused on assessing the susceptibility to natural hazards are: Frequency Ratio (Cao et al., 2016; Costache and Zaharia, 2017), Weights of Evidence (Chen et al., 2018), Certainty Factor (Z. Chen et al., 2019), Evidential Belief Function (Omar F Althuwaynee et al., 2014), Statistical Index (Chen et al., 2015), and Index of Entropy (Al-Abadi and Shahid, 2015). It should also be mentioned that the application of bivariate statistics in the field of susceptibility to natural hazards requires as input data the points or areas where the analyzed phenomena were recorded in the past (Arabameri et al., 2019). In fact, these input data are mandatory to be used also in the case of machine learning or artificial intelligence algorithms. The most well-known machine learning models applied in the study of natural hazards are: Multilayer Perceptron (Ngo et al., 2018), Support Vector Machine (Choubin et al., 2019), Decision Trees (Omar F. Althuwaynee et al., 2014), k-Nearest Neighbor (Avand et al., 2019), Logistic Regression (Bui et al., 2011), Naïve Bayes (Hosseini et al., 2020), Bagging (W. Chen et al., 2019), Dagging (Yariyan et al., 2020), Decorate (Zhang et al., 2012), Adaptive Neuro-Fuzzy Inference System (Ahmadlou et al., 2019). It is also a common practice to generate ensembles between machine learning and bivariate statistics (Costache and Bui, 2019), or between several machine learning models taken together (Pham et al., 2018). Also, noteworthy is the application of optimization algorithms such as: Particle Swarm Optimization (Bui et al., 2017), Harris Hawk Optimization (Bui et al., 2019a) and Biogeography based-Optimization (Wang et al., 2019). Widely used in determining the susceptibility to natural hazards are also the specific multicriteria decision-making methods as well: Analytical Hierarchy Process (Dahri and Abida, 2017, Sajedi-Hosseini et al., 2018), DEMATEL (Kanani-Sadat et al., 2019) and VIKOR (Ameri et al., 2018).

In this context, the present study wants to propose a methodology for estimating the susceptibility to flash-flood propagation, this topic not being addressed so far in the literature. The flash-flood propagation susceptibility will be computed by completing two major stages. The first stage will consist in determining the flash-flood susceptibility by applying the bivariate Weights of Evidence (WOE) method, as well as their novel ensembles with Analytical Hierarchy Process (AHP), Logistic Regression (LR), Classification and Regression Trees (CART) and Radial Basis Function Neural Network (RBFNN). The evaluation of the accuracy of flash-flood susceptibility results, provided by the 5 models, will be done through ROC Curve method and several statistical metrics. The second stage will consist in the actual calculation of the flash-flood propagation susceptibility by using the results of the first stage and the Flow Accumulation method.

2. Study area

The present study is focused on the Zăbala river basin, located in the mountainous area of the central-south-eastern part of Romania. The study area represents a small to medium-sized basin with a total surface of 600 km^2. The altitude of the study area varies from 312 m to 1786 m (Fig. 1). This high amplitude of elevation across a relatively small area, creates the premises for the genesis and propagation of flash-floods from the upper to the lower part of the river basin. In fact, the river basin is characterized by a relatively high average slope of 12.7°, this being another indicator of the high potential for the flash-flood genesis. According to the existing information, the afforestation degree of the river basin is around 60%. The genesis of flash-floods is also favoured by the hard rocks in the substrate of the study area, as well as by the presence of pasture vegetation on relatively compact surfaces. Important damages to the socio-human elements were generated by flash-floods during the years: 2010, 2016, 2017 and 2019.

3.2. Data

3.1. Flash-flood inventory

Any natural phenomenon has a higher occurrence probability over the areas where it has already occurred and where the environmental elements favour its genesis (Dottori et al., 2018).

Therefore, in the natural hazards susceptibility studies, it is very important to identify the locations that have already been affected by that phenomenon, and then to establish the spatial relationship between the presence/absence of the hazard and the characteristics of geographical factors (Yariyan et al., 2020). Thus, in the present case, in order to evaluate the susceptibility of the surfaces to the genesis of the flash-floods, data regarding the places where these phenomena occurred in the past were collected. In this regard, the damage reports provided by the General Inspectorate for Emergency Situation (GIES) of Romania and the information from mass-media were used. Totally, a number of 255 de flash-flood locations were collected. It should be noted that the majority of identified flash-floods were determined by the river discharge values with a return period of 10 years. Though, it is important to mention that the return period couldn't be established for each flash-flood event because the phenomena occurred on river sectors without hydrometric measurements. Another sample of 255 points were placed in areas where the flash-floods did not occur in the past. These points were considered as non-flash-flood locations. It is worth to note that, along the information from the governmental authorities that didn't mention the occurrence of flash-flood events, the non-flash-flood locations were also placed based on the analysis satellite images and field surveys. Both of the samples were split in training (70%) and validating (30%) datasets. The training datasets will be used exclusively for running the models, while the validating dataset will be used to validate the flash-flood susceptibility results.

3.2. Flash-Flood Predictors

According to the above section, the characteristics of geographical factors are those that influence the genesis and manifestation of flash-floods. Thus, in order to identify as accurately as possible, the surfaces favourable to the flash-floods genesis, a number of 10 conditioning factors were taken into account. Six morphometrical predictors were derived from the Digital Elevation Model (DEM). The DEM at a spatial resolution of 30 meters, was extracted from SRTM, 30 databases. Another 3 flash-flood predictors were extracted or derived from vector databases as follows: land use/cover was extracted from Corine Land

Cover, 2018 database; hydrological soil groups din Digital Soil Map of Romania, 1:200,000; lithology din Digital Geological Map of Romania, 1:200,000. Another predictor, represented by Modified Fournier Index (MFI), was achieved by the processing of Worldclim v2 database in raster format. Below, each factor was described from the perspective of their influence on flash-flood phenomena.

Slope is the essential factor that creates favourable conditions for both flash-flood genesis and propagation (Antonetti et al., 2019). Thus, areas with steep slopes will favour the occurrence of rapid surface runoff and the formation of flash-floods (Fontanine and Costache, 2013; Hapciuc et al., 2016). In the case of the study area, the slope of the relief was derived from the DEM at a cell size of 30 meters. As can be seen in Fig. 2a, the slope of the relief has values between 0 ° and 48 °. This interval was divided into 5 classes, taking into account the literature (Costache, 2014c).

Land use / cover is another important geographic element with a major contribution in the genesis of flash-floods (Zhao et al., 2019). The lands where the pastures predominate or which are totally devoid of vegetation will determine the appearance of runoff on the slopes, while the forested regions protect the surface of the land against torrential phenomena (Hosseini et al., 2020). In total, a number of 5 use classes were identified, over 60% of the total river basin being covered by forest (Fig. 2b).

Lithology is an essential parameter in defining the degree of impermeability of a surface. This degree of impermeability contributes decisively to the potential for rapid runoff manifestation on the slopes (Talukdar et al., 2020). The conglomerates, breccias, sandy flysch and marls shale are predominant in the study area (Fig. 2c).

Hydrological Soil Groups, influence in an indirect manner the flash-floods genesis. Thus, runoff will be favoured above soils with a high clay content, such as those in hydrological group D, while more active infiltration in soils with a high sand content will cause a decrease in flash-flood potential (Gessesse et al., 2015). Within the study area the largest surfaces are occupied by the hydrological soil group B (Fig. 2d).

Plan curvature is described by the line generated at the intersection of terrain surface and a horizontal plane. This morphometric indicator highlights the difference between the convergent and divergent runoff manifested at the ground surface. The following three classes were defined for plan curvature (Fig. 2e): -2.36 - -0.1; -0.09 – 0.1; 0.1 – 2.19.

Profile curvature is another morphometric factor obtained from DEM. In terms of flash-flood susceptibility, the importance of this factor is given by the fact that its negative values indicate the areas where surface runoff is accelerated, while its positive values show the areas where surface runoff is diminished (Ali et al., 2020). According to the scientific literature, the profile curvature values were grouped into the following 3 classes: -3.08 - -0.05; -0.04 – 0.05; 0.05 – 3.65 (Fig. 2f).

Convergence Index (CI) is a morphometric factor derived from DEM at the same spatial resolution as the slope of the relief. The values of this index show the degree of concentration of all fluvial and torrential organisms in a given region. A high hydrographic convergence is highlighted by negative CI values, close to -100, while the interfluvial surfaces have associated positive values. In the study area, the CI values are between -78 and 95 (Fig. 3a). The range of value was divided into 5 classes according to the literature (Prăvălie and Costache, 2014).

The Modified Fournier Index (MFI) highlights the spatial distribution of the rainfall intensity (Costache et al., 2020a). For this reason, the consideration of this indicator for estimating the flash-flood potential has a higher degree of representativeness than the consideration of multiannual average precipitation values. MFI is determined through the following mathematical relation:

$$MFI = \sum_{i=1}^{12} \frac{P_i^2}{P} \tag{1}$$

where: MFI – Modified Fournier Index, Pi - being the monthly precipitation at month i, Pt - the annual precipitation. For the Zăbala river basin, MFI was determined by processing the precipitation data from Worldclim v2 database. In the case of the study area, the following 4 MFI classes were delineated: <60, 60 – 90, 90 – 120, >120 (Fig. 3b).

Aspect predictor derived from DEM, is a real help for the evaluation of susceptibility to flash-floods because the differentiation of the surfaces on the 9 orientation groups can indicate in a clear way which is the potential of humidity that exists at the level of each group (Chapi et al., 2017). In the case of the present research area, the largest areas are covered by the North-East exposed surfaces (Fig. 3c).

Topographic Wetness Index (TWI) was obtained in SAGA GIS 2.0.2 software by DEM processing. TWI values are calculated by dividing the upslope catchment area to the slope angle (Hong et al., 2018). The values of this indicator, within the study area, range from -7.35 to 24.66. Following the recommendation from the previous scientific works (Lei et al., 2020), the entire range of TWI values was divided into five classes using the Natural Breaks method (Fig. 3d).

A succinct presentation of the data used in the present research is included also in Table 1.

4. Methods

The workflow applied in the present research is synthetically presented in the Figure 4. The methods, the software used and their training procedure are described in the following rows.

4.1. Multicollinearity assessment and feature selection

Variance Inflation (VIF) and Tolerance (TOL) are 2 of the most popular indices used to evaluate the multicollinearity among the variables that are used as input in a mathematical model (Miles, 2014). In fact, the assessment of multicollinearity among flash-flood predictors is mandatory to reduce redundant information and bias within models (Wheeler and Tiefelsdorf, 2005). Thus, in this paper VIF and TOL will be estimated through the SPSS 21 software. It should be noted that TOL values less than 0.2 and VIF higher than 4, may indicate the presence of multicollinearity (Dou et al., 2018).

The ReliefF method will ensure the initial evaluation of the predictive ability of variable used to estimate the flash-flood potential. Thus, feature selection process can: i) help to reduce the time of models training; ii) made the models less complex and also easier to analyse, iii) help to select the best variables in order to increase the models accuracy; iv) can decrease the overfitting. ReliefF Attribute is able to deal with multiclass problems (Urbanowicz et al., 2018), and therefore, was also selected to be used in the present research. Also, it is worth to admit that the ReliefF algorithm is able to operate with continuous and discrete data. This method consider the given attribute value associated to the closest instance of different or the same class (Urbanowicz et al., 2018). In the present study, the ReliefF will be run using Weka 3.9 software.

4.2. Weights of Evidence (WoE)

WoE is a bivariate statistical method which is based on Bayes theory. This algorithm is very popular in research works focused on natural risk susceptibility evaluation (Khosravi et al., 2016). In the present research *WoE* was applied as stand-alone model for flash-flood susceptibility assessment and at the same time the *WoE* coefficients were also used as input in the following models in order to create a number of four ensemble: Analytical Hierarchy Process, Logistic Regression, Classification and Regression Trees, and Radial Basis Function Neural Network. The estimation of *WoE* coefficients was based on the spatial overlapping of flash-flood pixels with factor classes/categories. The mathematical relations used in this regard are written below (Costache and Bui, 2019):

$$W^+ = ln\frac{P(B|S)}{P(B|\bar{S})} \quad (2)$$

$$W^- = ln\frac{P(\bar{B}|S)}{P(\bar{B}|\bar{S})} \quad (3)$$

where: W^+ - positive weight, W^- - negative weight, P – the probability, B – the presence of flash-flood predictor, $\bar{B}$ - the absence of flash-flood predictor, S – the presence of flash-flood phenomena, $\bar{S}$ - the absence of flash-flood phenomena.

In order to be implemented in GIS environment, the above relations can be transformed into (Mohammady et al., 2019):

236

$$W^{+} = \ln\left(\left[\frac{Npix_1}{Npix_1 + Npix_2}\right] / \left[\frac{Npix_3}{Npix_3 + Npix_4}\right]\right) \quad (4),$$

$$W^{-} = \ln\left(\left[\frac{Npix_2}{Npix_1 + Npix_2}\right] / \left[\frac{Npix_4}{Npix_3 + Npix_4}\right]\right) \quad (5),$$

237

238 where: $Npix_1$ - number of flash-flood pixels within a predictor class; $Npix_2$ number of
239 flash-flood pixels outside of the predictor class; $Npix_3$ – number of pixels without flash-flood
240 phenomena in the predictor class; $Npix_4$ - number of flash-flood pixels without flash-flood
241 phenomena outside of the predictor class; W^+ - positive weight, W^- - negative weight.

242 The final value of a *WoE* coefficient was achieved using the following formula (Costache
243 and Bui, 2019):

244 $$Wf = Wplus + Wmin\,total - Wmin \quad (6),$$

245 where: *Wplus* – is the positive weight of a class factor, *Wmin* – is the negative weight of a
246 class factor, *Wmintotal* – is the total of all negative weights in a multiclass map.

247 **4.3. Analytical Hierarchy Process (AHP)**

248 AHP is a multicriteria decision-making model which is frequently involved in the research
249 works whose main purpose is the identification of regions susceptible to natural risks (Akıncı
250 et al., 2013; Ghosh and Kar, 2018; Pourghasemi et al., 2016). An important aspect which
251 should be exposed is that the AHP represents a semi-quantitative method in which a very
252 important weight is allocated to the expert judgment. Thus, by applying this model a problem
253 could be solved by an active involvement of the experts in the research workflow. Proposed
254 by Saaty (1980), AHP algorithm applied in the present study will consists of six major steps

through which the problem can be break down into several components. The steps are briefly described below:

i) Establish the objectives and split the problem into many components;

ii) Defining the criteria and the alternatives;

iii) Generating the AHP pair-wise comparison matrix using the expert judgement. More details regarding the construction of pair-wise comparison matrix can be found in (Costache, R. and Tien Bui, D., 2020);

iv) Using the eigenvalue method to calculate the relative importance of each flash-flood predictor;

v) Assessing the quality of pair-wise comparison using the Consistency Ratio (CR). CR is estimated as follows:

$$CI = \frac{\lambda_{max} - n}{\mathrm{n} - 1} \tag{7}$$

where CI represents the value of consistency index; λ is the eigenvalue with the highest value within the entire matrix, which can be computed using the eq. 8; n is number of flash-flood predictors.

$$CR = \frac{CI}{RI} \tag{8}$$

where RI represents the value of random consistency index which can be found in literature (Agarwal et al., 2013).

A consistent pair-wise comparison is highlighted by a CR under 0.1 (Sun, 2010).

vi) Calculate the flash-flood potential index (FFPI*)* by integrating the AHP weights with WoE coefficients in GIS environment as follows:

$$FFPI_{AHP-WOE} = \sum_{j=1}^{n} AHP_j Wf_{ij} \qquad (9)$$

274 where AHP_j is the importance of a flash-flood predictor j, Wf_{ij} is the weights of evidence
275 coefficient associated to a class i of predictor j, and n is the number of predictors.
276 It should be mentioned that for the present study, the data necessary for the application of
277 AHP method was obtained through an expert-based questionnaire survey administered to a
278 number of 19 experts from the National Institute of Hydrology and Water Management of
279 Romania, with a high expertise in flash-flood risk assessment. The number of interviewed
280 experts is very close to the number that was also used in previous works from the literature
281 (Dano, 2021, 2020).

282 4.4. Logistic Regression

283 Logistic Regression (LR) model aims to identify the best relation between a set of predictors
284 and a binary variable (Kavzoglu et al., 2014; Pradhan, 2010). Therefore, it can be admitted
285 that the logistic regression method is especially used to predict the absence and the presence
286 of a specific process, based on the characteristics of the spatial relationship between certain
287 predictors and dependent variable. Logistic Regression model is able to work with both
288 continuous and discrete variables or with a combination of both. In the present research, the
289 dependent variable is represented by the flash-flood and non-flash-flood locations, while the
290 explanatory/independent variables are represented by the flash-flood predictors. It is worth
291 to note that the flash-flood points were encoded with 1, while non-flash-flood points were
292 encoded with 0.

By adapting to the present research the next equation represent the mathematical expression of the logistic regression linear model (Bui et al., 2011):

$$p = \frac{1}{1 + e^{-Z}} \quad (10)$$

where p is the probability of a flash-flood event, Z is a value from $-\infty$ to $+\infty$, calculated with the next relation:

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (11)$$

where b_0 is the model intercept value, the β_i (i=0, 1, 2, ..., n) are the values of the logistic regression slope coefficients, and the x_i (i=0, 1, 2, ..., n) represent the flash-flood predictors having assigned the WOE coefficients.

Within Logistic Regression model, the multicollinearity, which will be assessed according to *4.1.* section, can induce some inaccuracies which could affect also the model hypothesis (Midi et al., 2010). The application of Logistic Regression model was possible through SPSS software in which the data were imported in tabular format. The accuracy of the classification done in LR depends the selection of optimal cut-off classification. In this regard, a trial process was carried out with the next cut-off classification values indicated in the literature (Soureshjani and Kimiagari, 2013): 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.99. Finally, the classification associated with the cut-off value that provides the highest accuracy, will be selected. Following the training procedure, the LR coefficients (βi) will be computed. In fact, these coefficients will be equal to the weight of each flash-flood predictor. Thus, the eq. 10 will be implemented in GIS environment to determine the flash-flood potential.

4.5. Classification and Regression Tree

Classification and Regression Tree (CART) is a popular machine learning model used in research works focused on natural hazards susceptibility computation (Costache and Bui, 2019; Hong et al., 2015; Yeon et al., 2010; Youssef et al., 2016). CART algorithm could be run using the following type of variables: categorical, binary and number. This characteristic represents an advantage of this model. Within the CART model, the selection of predictors is carried out so that the data error to be diminished. The entropy in CART model, represents the measure to which a predictor is preferred against to another. It should be noted that if a predictor has a missing value, it will not be involved in the construction of the tree optimal ramification. In this case the missing values are substituted by surrogates (Breiman et al., 1984). A terminal node within the CART structure is equal to the average response in that specific node (Breiman et al., 1984). The best sampling rule in CART model training procedure consists of the direct association between the target attribute in two child nodes and is described by the next relation (Costache et al., 2020b):

$$I(Split) = \left[0.25(q(1-q))^{u}\Sigma_{k}|PL(k) - PR(k)|\right]^{2} \quad (12)$$

where: k is the index of the target classes, PL(k) and PR(k) represent the probability distributions of the target in the left and right child nodes, respectively, and the power term u embeds a user-trollable penalty on splits that create child nodes with unequal sizes (Wu et al., 2008).

In the present study, CART-WOE ensemble was applied by using SPSS software. The optimization of CART-WOE model was made by adjusting their parameters in order to achieved the highest accuracy. Finally, the pruned decision tree will keep only the most important information.

4.6. Radial Basis Function Neural Network (RBFNN)

RBFNN is a type of neural network that consists of the following three layers: i) input layer; ii) hidden layer; iii) output layer (Zare et al., 2013). From the present research perspective, the input layer will contain as input data the flash-flood predictors with WOE coefficients assigned, the hidden layer will help to process and translate the information from the input to the output layer and backward, while the output layer will consist of two neurons represented by the flash-flood and non-flash-flood points. According to the literature (Pham et al., 2020), through the hidden layer, the RBF non-linear activation function will be applied to train the neural network. More specifically, the RBFNN consists of the computation of Euclidean Distance from the evaluated points towards the neurons centre. Further, the RBF will be applied to the distance in order to quantify the influence of the neurons. Usually, in this regard, the Gaussian function is used to express the mathematical form of RBF(Pham et al., 2020):

$$\phi_j = exp\left(-\frac{(X-c_j)^2}{2\sigma_j^2}\right) \tag{13}$$

where ϕ_j is the RBF of the j^{th} RBF neuron, $X = (x_1, x_{2,\dots,} x_n)^T$ is the input vector with d input variables, $c_j = (c_{1j}, c_{2j,\dots,} c_{dj})^T$ is the center vector, and σ_j^2 is the spread.

In terms of RBFNN output, for each class will be calculated the probability and the class with the highest probability will receive the input data. The RBFNN output can be derived with the following a weighted sum (Qasem and Shamsuddin, 2011):

$$Y = \sum_{j=1}^{p} w_j \phi_j \tag{14}$$

Where Y represents the RBFNN output, p represents the sum of neurons, w_j is the weight assigned from the j^{th} RBF neuron to the output layer.

In the present research, the RBFNN-WOE ensemble was applied using SPSS software. One of the crucial steps of the training procedure was the establishment of the optimal hidden neurons number. Thus, the optimal number of neurons in the hidden layer was established according to the highest accuracy achieved by the model and which was measured with the help of confusion matrix.

4.7. Flash-Flood Potential results validation methods

4.7.1. ROC Curve

The receiver operating characteristic (ROC) curve is the most popular method involved in the validation of the results of studies related to the susceptibility to floods and flash-floods (Arabameri et al., 2020; Bui et al., 2019b; Ngo et al., 2018). The graphic of ROC Curve is associated to the representation of the sensitivity on Y axis against the 1-Specificity on X axis (Aguilar et al., 2013). From the present research point of view, this method indicates the capacity of a model to correctly estimate the occurrence of flash-flood hazard. Within ROC Curve model, the most valuable quantitative information is provided by the Area Under Curve (AUC) which range from 0 to 1. The values near to 1 highlight a high performance of the applied models (Vakhshoori and Zare, 2018). The following relation is used to calculate the AUC:

$$\mathrm{AUC} = \frac{(\sum TP + \sum TN)}{(\mathrm{P} + \mathrm{N})} \tag{15}$$

where P is equal to the sum of flash-flood locations, N is equal to the sum of non-flash-flood locations, TP (true positive), TN (true negative) are the sums of flash-flood and non-flash-flood correctly classified locations.

4.7.2. Statistical metrics

Along with ROC Curve method, the following 7 statistical indices were involved in the results validation procedure: Kappa Index, Sensitivity, Specificity, F1 score, Accuracy, Precision. The significance of the statistical indices is represented by the agreement between the observed flash-flood and non-flash-flood locations and the predicted flash-flood susceptibility values (Costache, 2019). The aforementioned metrics can be computed with the next equations (Canbek et al., 2017; Costache et al., 2020c):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{17}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{18}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

(19)

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

$$\text{k} = \frac{\text{p}_\text{o} - \text{p}_\text{e}}{1 - \text{p}_\text{e}} \tag{21}$$

where P is the number of flash-flood pixels, N is the number of non-flash-flood pixels, FP (false positive) and FN (false negative) are the sums of flash-flood and non-flash-flood erroneously classified locations, k is kappa coefficient, p_o is the observed flash-flood locations, and p_e is the estimated flash-flood susceptibility pixels."

4.8. Flash-Flood Propagation Susceptibility Index

The Flash-Flood Propagation Susceptibility Index (FFPSI) is a novel concept and indicator proposed and defined for the first time in the literature in the present paper. FFPSI can be defined as the potential of the river valleys across a specific area to propagate the flash-floods from the upper part of a catchment toward its lower zone. In order to estimate the FFPSI within the present research territory, the Flash-Flood Potential Index values, calculated through the above-described models, were integrated in a *Flow Accumulation* procedure. The *Flow Accumulation* generates, in GIS environment, a raster in which each cell value is equal to the weighted sum of all cells in the raster that drain to that cell (O'Callaghan and Mark, 1984). Therefore, in order to calculate the FFPS the *Flow Accumulation* will be used to weight the *FFPI* values on the hydrographic network within the study area. ArcGIS 10.3 software was used to implement the workflow intended to compute the FFPS value. Thus, in a first stage, the *Flow Direction* across the study area was computed. Then, the Simple Flow Accumulation (SFA) and the FFPI Weighted Flow Accumulation ($FFPI_{WFA}$) rasters were derived. Finally, the FFPS values were achieved by dividing the $FFPI_{WFA}$ to SFA as suggested in the next equation:

$$FFPSI = \frac{FFPI_{WFA}}{S_{FA}} \quad (22).$$

5. Results

5.1. Multicollinearity assessment and feature selection

In the present study, TOL values, one of the multicollinearity indicators, range from 0.411 for lithology to 0.976 for Aspect. The minimum VIF value was 1.025, and corresponds to the Aspect factor, while the maximum one was 2.432 and corresponds to the Lithology (Table

2). Given the fact that TOL has values higher than 0.2 and VIF has values below 4, we can admit that among the 10 flash-flood predictors there is no serious multicollinearity.

In terms of ReliefF attribute, used to evaluate the predictive ability of flash-flood conditioning factors, the highest values was achieved by Slope (0.215), followed by Plan curvature (0.039), MFI (0.036), Convergence Index (0.026), TWI (0.019), Profile curvature (0.016), Land use (0.015), HSG (0.014), Aspect (0.006) and Lithology (0.003). Given the fact that all the ReliefF scores were higher than 0, we can consider that all the flash-flood predictors contribute in a specific measure to the genesis of this phenomenon. Therefore, all the predictors will be used in the analysis.

5.2. Results of Weights of Evidence (WOE)

The application of Weights of Evidence method revealed that WOE coefficients range from -4.01 for slopes lower than 3° to 3.09 for hydrological soil group B. The lowest value achieved by slopes lower than 3° is explained by the impossibility of flash-flood genesis on surfaces which are almost flats. The areas covered by flysch, marls shale, sandstones, clays and schists are also characterized by very low WOE values (-2.5) (Table 3). Instead, the hydrological soil group C (2.09), negative profile curvature (0.56), pastures (0.36) and north-eastern slopes (0.3) have high WOE coefficients.

Using the WOE coefficients inserted in eq. 6, the $FFPI_{WOE}$ was calculated (Fig. 7a). The $FFPI_{WOE}$ values were standardized between 0 and 1 and after that were classified into 5 classes using *Natural Breaks* method. According to GIS modelling, the very low values, between 0 and 0.38, cover 6.54% of Zăbala river catchment (Fig. 10). Another 20.31% of the research area represents low $FFPI_{WOE}$ values and are limited between 0.39 and 0.51. It can be observed that the very and low flash-flood potential are mainly presented in the eastern region, at lowers altitudes. The medium values, ranging from 0.52 and 0.62, are spread on 28.14%, while high and very high values are encountered on approximately 45.01% of the entire territory.

5.3. Results of Analytical Hierarchy Process - Weights of Evidence (AHP-WOE)

The first step in the computation of $FFPI_{AHP\text{-}WOE}$ was the construction of the pair-wise comparison matrix with the help of Microsoft Excel 2016 software. Thus, through the assignment of a relative dominant value, each of the 10 flash-flood predictors was rated against every other (Table 4). Following the steps written at *4.3*, the normalized weight of each flash-flood predictor was determined and the quality of comparisons was evaluated. The highest weight was assigned to the slope (0.26), followed by land use (0.179), lithology (0.111), profile curvature (0.11), plan curvature (0.085), MFI (0.07), convergence index (0.05), TWI (0.05), hydrological soil group (0.045) and aspect (0.04). The good quality of the comparisons is attested by the Consistency Ratio (CR) value equal to 0.03. Finally, by implementing the equation 9 in ArcGIS software the $FFPI_{AHP\text{-}WOE}$ values were derived (Fig. 7b). The values were standardized between 0 and 1, and were classified into five classes using *Natural Breaks* method. The very low values of flash-flood potential index, ranging from 0 to 0.27 appear on around 2.38% of the study area and are mainly present in the extreme eastern part of the study area. The low $FFPI_{AHP\text{-}WOE}$ values, between 0.28 and 0.44, occupy 28.49% of Zăbala river catchment and are distributed especially in the median part of the research area. Medium values of the same index account 26.81% of the total territory, and are mainly spread in the eastern half of the research perimeter. The high slopes and altitudes from the southern and northern parts of Zăbala river catchment are covered by high and very high flash-flood potential which span on 42.32% of the research area.

5.4. Results of Logistic Regression - Weights of Evidence (LR-WOE)

Following the trial process, the highest accuracy of the model was obtained by the application of a cut-off value of 0.5. Thus, according to Fig. 5 and Table 5, it can be observed that the best accuracy of 86.87% was associated to a Sensitivity of 83.8% and a Specificity of 89.94%. The good performance of the classification performed through LR-WOE model is also indicated by the classification plot (Fig. 6) in which the observed and predicted probabilities are distributed according to their frequencies. More specifically, this plot displays the frequency in which the model would predict a flash-flood outcome, encoded with '1', using the computed predicted probability in the case in which the outcome was 'non-flash-flood'. Therefore, the distribution of observation cases, predominantly in the extreme left and right of the plot, indicate the very good performance of the model. This situation is associated with the absence of the cases in the middle region of the plot.

One of the most important output of the LR-WOE training process is represented by the logistic regression coefficients values (β). Thus, it can be noted that the land use achieved the highest coefficient (1.603), followed by slope (1.274), aspect (1.253), MFI (0.835), TWI (0.431), plan curvature (0.377), convergence index (0.102), lithology (0.099), hydrological soil groups (0.098) and profile curvature (-0.215) (Table 6). Further, by using these coefficients in equation 11 implemented in Map Algebra of ArcGIS 10.3 software, the flash-

485 flood potential was calculated. The $FFPI_{LR-WOE}$ values were standardized between 0 and 1
486 and their values were reclassified into 5 classes using *Natural Breaks* method (Fig. 7c).
487 The very low values, between 0 and 0.47, are spread on 7.95% of the study area and are
488 mainly present in the eastern half of Zăbala catchment. The low $FFPI_{LR-WOE}$, between 0.48 –
489 0.57, can be found especially in the western half of the research zone and occupy 25.45% of
490 the territory. The medium $FFPI_{LR-WOE}$, ranging from 0.58 to 0.69, cover 25.2% of the research
491 zone and are randomly distributed over the Zăbala river catchment. The high and very high
492 flash-flood potential, with $FFPI_{LR-WOE}$ higher than 0.7, span on around 41.4% of the entire
493 study area and can be found mainly in the northern and southern halfs.

494 **5.4. Results of Classification and Regression Trees - Weights of Evidence (CART-WOE)**

495 The training process of CART-WOE was done by optimizing the number of parent and
496 terminal nodes of the best pruned tree. The optimization was done according to the accuracy
497 value presented in Table 5. Thus, the highest accuracy (86.31%) was achieved with a tree
498 characterized by a number 4 terminal nodes (Fig. 8a).

499 According to the training process, the highest importance was assigned to Slope (0.559),
500 followed by Convergence Index (0.081), MFI (0.059), TWI (0.037), plan curvature (0.034),
501 hydrological soil groups (0.029), lithology (0.015), aspect (0.003), profile curvature (0.002)
502 and land use (0.001). The computation of $FFPI_{CART-WOE}$, assumed the use of flash-flood
503 predictors relative weights, in GIS map algebra.

504 The normalized values of $FFPI_{CART-WOE}$ were reclassified into 5 groups using the *Natural*
505 *Breaks* method (Fig. 7d). The very low class of flash-flood potential, between 0 and 0.1,

covers a small area equal to 2.36% of the Zăbala river catchments, and is spread only in the eastern part. The low values span on 7.39% and range between 0.11 and 0.4. The medium class of $FFPI_{CART-WOE}$ accounts approximately 47.3% of the study area and is randomly and covers large areas in the eastern half of Zăbala river catchment. The high and very high values have 42.4% of the perimeter of Zăbala river catchment and are characterized by $FFPI_{CART-WOE}$ values higher than 0.5. These critical areas are mainly located in the western side of the research territory.

5.5. Results of Radial Basis Function Nueral Network - Weights of Evidence (RBFNN-WOE)

The optimal RBFNN-WOE architecture was established in concordance with the highest performances achieved during the model training. A first indicator of the performance is the confusion matrix (Table 5), that revealed a highest accuracy of 84.91% associated to an architecture with a number of 14 hidden neurons within the hidden layer (Fig. 8b). The very good performance of RBFNN-WOE classification is revealed also by the AUC (0.906) ROC Curve constructed for both flash-flood and non-flash-flood sample (Fig. 9a). At the same time the pseudoprobability plot (Fig. 9b), in which the values above 0.5 of y-axis highlight represent the correct classification, attest that the flash-flood and non-flash-flood locations are correctly classified. Moreover, the high performance of RBFNN-WOE ensemble classification is indicated also by Lift (Fig. 9c) and Gain (Fig. 9d) charts.

After the training procedure, the importance of flash-flood predictors was derived. Thus, the highest importance was achieved by slope (0.251), followed by land use (0.122), MFI

(0.102), TWI (0.091), lithology (0.078), plan curvature (0.078), convergence index (0.074), aspect (0.073), profile curvature (0.069) and hydrological soil groups (0.062). Using these values, the $FFPI_{RBFNN\text{-}WOE}$ was calculated. Further the standardized range of $FFPI_{RBFNN\text{-}WOE}$ was grouped into 5 classes using *Natural Break* method. The first class, between 0 and 0.3, cover 2.59% of the study area and belongs to the surfaces characterized by a very low flash-flood potential (Fig. 7e). Approximately 17.74% of Zăbala river catchment has a low $FFPI_{RBFNN\text{-}WOE}$ with values between 0.31 and 0.45 which are located mainly in the eastern half of the research zone. Around 37.38% of the river basin has a medium flash-flood potential which can be found especially in the western part of the research area. Together, the high and very high values of $FFPI_{RBFNN\text{-}WOE}$ account 41.29% of the entire territory.

5.6. Validation of FFPI results

The first step in FFPI the results validation is the application of ROC Curve with their 2 plots represented by Success Rate, constructed with the training sample, and Prediction Rate, constructed with the validating sample. Thus, the Fig. 11a indicates that, in terms of Success Rate, the highest performance was achieved by $FFPI_{LR\text{-}WOE}$ with an AUC of 0.923, being followed by $FFPI_{RBFNN\text{-}WOE}$ (AUC = 0.911), $FFPI_{AHP\text{-}WOE}$ (AUC = 0.903), $FFPI_{CART\text{-}WOE}$ (AUC = 0.901) and $FFPI_{WOE}$ (AUC = 0.865). Instead, the highest performance in terms of Prediction Rate was achieved by $FFPI_{AHP\text{-}WOE}$ (AUC = 0.894), followed by $FFPI_{CART\text{-}WOE}$ (AUC = 0.891), $FFPI_{RBFNN\text{-}WOE}$ (AUC = 0.88), $FFPI_{LR\text{-}WOE}$ (AUC = 0.875) and $FFPI_{WOE}$ (AUC = 0.854).

The second stage of results validation consisted in the computation of several statistical metrics. As can be observed in Table 7, the use of training sample, highlights as most accurate results the $FFPI_{LR\text{-}WOE}$ (Accuracy = 0.877), followed by $FFPI_{RBFNN\text{-}WOE}$ (Accuracy = 0.866), $FFPI_{AHP\text{-}WOE}$ (Accuracy = 0.86), $FFPI_{CART\text{-}WOE}$ (Accuracy = 0.846) and $FFPI_{WOE}$ (Accuracy = 0.793). It can be observed that the hierarchy of the values of the other statistical metrics followed the same pattern as accuracy indicator in terms of training dataset. Instead, in terms of validating dataset, the most accurate results is $FFPI_{AHP\text{-}WOE}$ (Accuracy = 0.882), followed by $FFPI_{LR\text{-}WOE}$ (Accuracy = 0.868), $FFPI_{RBFNN\text{-}WOE}$ (Accuracy = 0.862), $FFPI_{CART\text{-}WOE}$ (Accuracy = 0.855) and $FFPI_{WOE}$ (Accuracy = 0.803).

5.7 Flash-Flood Propagation Susceptibility Index (FFPSI)

The novel FFPSI was calculated for each model according to the methodology described at sub-section 4.8. It should be noted that FFPSI values were classified using Natural Break algorithm. Thus, in terms of WOE method (Fig. 12a), the spatiall modelling of FFPSI revealed that a percentage of 5.59% of identified valleys have a very low flash-flood propagation susceptibility. There valleys are located especially in the eastern part of Zăbala river catchment. Another percentage of 15.68% is represented by the valleys with a low flash-flood propagation susceptibility which are situated on the median part of study area. The medium $FFPSI_{WOE}$ represents 25.19%, while the high and very high potential characterize 53.55% of the identified river valleys (Fig. 13).

Fig. 12b indicates that only 0.66% of the identified river valleys are characterized by a very low flash-flood propagation susceptibility according to the AHP-WOE ensemble. The very low flash-flood propagation susceptibility is encountered on around of 19.55% of the river valleys, while medium $FFPSI_{AHP\text{-}WOE}$ characterizes 22.31% (Fig. 12b). The high and very high potential for flash-flood propagation is characteristic for 57.48% of the analyzed river valleys. In terms of LR-WOE ensemble the classes of flash-flood propagation susceptibility have the following spatial distribution: very low – 7.92%, low – 18.92%, medium – 22.66%, high – 30.25% and very high – 20.25% (Fig. 12c).

Following the application of CART-WOE ensemble, the very low flash-flood propagation susceptibility appears on 0.73% of the river valleys, the low susceptibility is present on 20.78%, medium susceptibility on 22.68%, while the high and very high susceptibility has 55.81% of the total analyzed valleys (Fig. 12d). In terms of RBFNN-WOE, the highest percentage is represented by the valleys with a high flash-flood propagation susceptibility (28.81%), followed by the very high propagation susceptibility (26.69%), medium propagation susceptibility (20.62%), low propagation susceptibility (18.58%) and very low flash-flood propagation susceptibility (5.3%) (Fig. 12d).

6. Discussions

This study is conducted in the undeniable context of the global climate change and its effects on the inevitable multiplication of hydrological risk phenomena such as flash-floods (Fowler and Wilby, 2010). It should be noted that the previous studies regarding the estimation of flash-flood susceptibility by machine learning techniques, carried out so far, did not include the study of the susceptibility of the valleys to the propagation of flash-flood waves (Anquetin et al., 2010; Janizadeh et al., 2019). Moreover, many researchers were focused, in their previous works, only on the evaluation of normal flood susceptibility (Azareh et al., 2019;

Dodangeh et al., 2020; Hosseini et al., 2021; Mosavi et al., 2020), without taking into account the flash-flood phenomenon particularities. Besides of the previous research works which take into account only the local flood susceptibility given by the punctual conditions and rainfall, this article proposes a new and complete approach regarding the study of slopes susceptibility to runoff and, also, regarding the susceptibility of valleys to the propagation of the flash-floods. Therefore, through FFPSI, for each valley across the study area are highlighted the characteristics of the upslope catchment that could determine a high exposure to flash-flood. This new approach was conducted with the help of bivariate statistics and machine learning and also using the Flow Accumulation procedure. In fact, the propagation of the flash-flood wave is the element that generates the most significant material damage and loss of human life (Mujumdar, 2001). Therefore, the integrated study of the surface runoff potential on the slopes and the susceptibility of the valleys to flash-flood waves propagation provides the clearest overview of the areas along the rivers that are at risk of being affected. Usually, the flash-flood waves propagation is simualted with the help of 1D (Leandro et al., 2011) or 2D (Abderrezzak et al., 2009) models, these approaches having the disadvantage of the fact that, unlike the workflow proposed in the present study, the realization of such a modeling at the level of a hydrographic basin of over 500 km^2 is time consuming and requires a very large volume of data which are often very expensive (Dewals et al., 2008).

This research paper includes a first part in which the Flash-Flood Potential Index was calculated and spatialized through 5 models, and the second part in which the Flash-Flood

Propagation Susceptibility Index was proposed, calculated and spatialized for the first time in the literature, taking into account the results of the first part and applying the Flow Accumulation method. The results regarding FFPI reveal a high performance of the applied models, these being characterized by AUC-ROC Curve values higher than 0.854.

It is also observed that the ensemble models obtained significantly better results compared to the stand-alone WOE model. Thus, in the case of training sample, WOE obtained an AUC equal to 0.865, this being clearly smaller than the weakest ensemble model, CART-WOE, which had an AUC equal to 0.901. The same aspect is true for the validating sample, where the WOE model obtained an AUC of 0.854, significantly lower than the AUC of 0.875 which was achieved by the LR-WOE ensemble. The higher performance of the ensemble models compared to the stand-alone ones, within the evaluation of flash-flood susceptibility, was also highlighted in the previous studies. Thus, according to Arabameri et al. (2020), the hybrid models are used to enhance the prediction ability of the algorithms used to map the spatial distribution of natural phenomena likelihood. Moreover, Pham et al. (2016) indicate that the ensemble models are superior to the stand-alone ones. Additionally, Costache et al. (2020c) highlight the superiority of *k*-Nearest Neighbor and K-Star ensembles with Analytical Hierarchy Process comparing to the stand-alone models, in terms of flash-flood susceptibility assessment.

The second part of the study, in which the FFPSI is spatialized, shows that the most exposed valleys to the propagation of flash-floods are those in the immediate vicinity of the slopes located in the central-southern and central-northern areas of the Zabala river basin. Also, it can be observed the predominance in the study area, in percentages higher than 50%, of the valleys having a high and very high potential for the flash-flood propagation. This fact is another indication of the existence of a high exposure of socio-economic elements in the study area to flash-floods.

7. Conclusions

In the present study a complex methodological workflow was developed to estimate the susceptibility to flash-flood propagation in the Zabala river basin. In this regard, a number of 10 flash-flood predictors and 255 flash-flood and 255 non-flash-flood locations were used as input data in the following models: WOE, AHP-WOE, LR-WOE, CART-WOE and RBFNN-WOE. These models were used in order to estimate the flash-flood potential index across the study area. The training process and, after that, the validation of the results achieved, required the split of flash-flood and flood datasets into training and validating samples. In order to map the FFPI, the *Natural Break* classification method was used for the results of all applied models. According to the results provided, a surface between 41% and 55% of the study area is covered by a high and very high flash-flood potential. The results validation, which is mandatory in this type of studies, revealed that LR-WOE, in terms of training sample (AUC = 0.923), and AHP-WOE, in terms of validating sample (0.894), were the most perfomant models. In order to estimate the flash-flood propagation susceptibility index (FFPSI), the results provided by the 5 models were integrated in the Flow Accumualtion procedure. Thus, it reavealed that around 56% of the river valleys identified within the study area are characterized by a high and very high FFPSI values.

The main element of novelty that characterizes this study is represented by the use and computation for the first time in the literature of Flash-Flood Propagation Susceptibility Index (FFPSI), which is of a real help to create a complete overview regarding the flash-flood susceptibility at the level of a river catchment. Another element of novelty is represented by the use for the first time in the literature of the following ensemble models in order to determine the flash-flood susceptibility: AHP-WOE and RBFNN-WOE.

The accurate results, atested by the results validation procedure, make from this study a benchmark for future studies related to the assessment of susceptibility to flash-floods in other study areas. Also, given the accuracy of the results, this study can be used by government authorities to mitigate the negative effects of flash-flood phenomena.

References

Abderrezzak, K.E.K., Paquier, A., Mignot, E., 2009. Modelling flash flood propagation in urban areas using a two-dimensional numerical model. Nat. Hazards 50, 433–460.

Agarwal, E., Agarwal, R., Garg, R., Garg, P., 2013. Delineation of groundwater potential zone: An AHP/ANP approach. J. Earth Syst. Sci. 122, 887–898.

Aguilar, M.A., del Mar Saldaña, M., Aguilar, F.J., 2013. Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images. Int. J. Appl. Earth Obs. Geoinformation 21, 427–435.

Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., Panahi, M., 2019. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). Geocarto Int. 34, 1252–1272.

Akıncı, H., Özalp, A.Y., Turgut, B., 2013. Agricultural land use suitability analysis using GIS and AHP technique. Comput. Electron. Agric. 97, 71–82.

Al-Abadi, A.M., 2018. Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. Arab. J. Geosci. 11, 218.

Al-Abadi, A.M., Shahid, S., 2015. A comparison between index of entropy and catastrophe theory methods for mapping groundwater potential in an arid region. Environ. Monit. Assess. 187, 576.

Ali, S.A., Parvin, F., Pham, Q.B., Vojtek, M., Vojteková, J., Costache, R., Linh, N.T.T., Nguyen, H.Q., Ahmad, A., Ghorbani, M.A., 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 Topľa basin, Slovakia. Ecol. Indic. 117, 106620. https://doi.org/10.1016/j.ecolind.2020.106620

Althuwaynee, Omar F, Pradhan, B., Park, H.-J., Lee, J.H., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. Catena 114, 21–36.

Althuwaynee, Omar F., Pradhan, B., Park, H.-J., Lee, J.H., 2014. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides 11, 1063–1078.

Ameri, A.A., Pourghasemi, H.R., Cerda, A., 2018. Erodibility prioritization of sub-watersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. Sci. Total Environ. 613, 1385–1400.

Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., Manus, C., 2010. Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event. J. Hydrol. 394, 134–147.

Antonetti, M., Horat, C., Sideris, I.V., Zappa, M., 2019. Ensemble flood forecasting considering dominant runoff processes–Part 1: Set-up and application to nested basins (Emme, Switzerland). Nat. Hazards Earth Syst. Sci. 19, 19–40.

Antronico, L., Coscarelli, R., De Pascale, F., Condino, F., 2019. Social Perception of Geo-Hydrological Risk in the Context of Urban Disaster Risk Reduction: A Comparison between Experts and Population in an Area of Southern Italy. Sustainability 11, 2061.

Arabameri, A., Pradhan, B., Rezaei, K., Sohrabi, M., Kalantari, Z., 2019. GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. J. Mt. Sci. 16, 595–618.

Arabameri, A., Saha, S., Chen, W., Roy, J., Pradhan, B., Bui, D.T., 2020. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. J. Hydrol. 125007.

Avand, M., Janizadeh, S., Naghibi, S.A., Pourghasemi, H.R., Khosrobeigi Bozchaloei, S., Blaschke, T., 2019. A Comparative Assessment of Random Forest and k-Nearest Neighbor Classifiers for Gully Erosion Susceptibility Mapping. Water 11, 2076.

Azareh, A., Rafiei Sardooi, E., Choubin, B., Barkhori, S., Shahdadi, A., Adamowski, J., Shamshirband, S., 2019. Incorporating multi-criteria decision-making and fuzzy-value functions for flood susceptibility assessment. Geocarto Int. 1–21.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Belmont, CA: Wadsworth. Int. Group 432, 151–166.

Bui, D.T., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., Trinh, P.T., 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric. For. Meteorol. 233, 32–44.

Bui, D.T., Lofman, O., Revhaug, I., Dick, O., 2011. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat. Hazards 59, 1413.

Bui, D.T., Moayedi, H., Kalantar, B., Osouli, A., Pradhan, B., Nguyen, H., Rashid, A.S.A., 2019a. A novel swarm intelligence—Harris hawks optimization for spatial assessment of landslide susceptibility. Sensors 19, 3590.

Bui, D.T., Tsangaratos, P., Ngo, P.-T.T., Pham, T.D., Pham, B.T., 2019b. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Sci. Total Environ. 668, 1038–1054.

Canbek, G., Sagiroglu, S., Temizel, T.T., Baykal, N., 2017. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. Presented at the 2017 International Conference on Computer Science and Engineering (UBMK), IEEE, pp. 821–826.

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L., Niu, C., 2016. Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. Sustainability 8, 948.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Khosravi, K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ. Model. Softw. 95, 229–245.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., 2018. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. Sci. Total Environ. 634, 853–867.

Chen, W., Li, W., Hou, E., Bai, H., Chai, H., Wang, D., Cui, X., Wang, Q., 2015. Application of frequency ratio, statistical index, and index of entropy models and their comparison in landslide susceptibility mapping for the Baozhong Region of Baoji, China. Arab. J. Geosci. 8, 1829–1841.

Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H.M., Hou, E., Wang, S., 2019. Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. Nat. Resour. Res. 1–20.

Chen, Z., Liang, S., Ke, Y., Yang, Z., Zhao, H., 2019. Landslide susceptibility assessment using evidential belief function, certainty factor and frequency ratio model at Baxie River basin, NW China. Geocarto Int. 34, 348–367.

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, 2087–2096.

Costache, R., 2019. Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. Sci. Total Environ. 659, 1115–1134.

Costache, R., 2014a. Using GIS techniques for assessing lag time and concentration time in small river basins. Case study: Pecineaga river basin, Romania. Geogr. Tech. 9, 31–38.

Costache, R., 2014b. Estimating multiannual average runoff depth in the middle and upper sectors of Buzău River Basin. Geogr. Tech. 9, 21–29.

Costache, R., 2014c. Assessing monthly average runoff depth in Sărățel river basin, Romania. Analele Stiintifice Ale Univ. Alexandru Ioan Cuza Din Iasi-Ser. Geogr. 60, 97–110.

Costache, R., Bao Pham, Q., Corodescu-Roşca, E., Cîmpianu, C., Hong, H., Thi Thuy Linh, N., Ming Fai, C., Najah Ahmed, A., Vojtek, M., Muhammed Pandhiani, S., 2020a. 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, 1422.

Costache, R., Bui, D.T., 2019. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. Sci. Total Environ. 691, 1098–1118.

Costache, R., Hong, H., Pham, Q.B., 2020b. Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. Sci. Total Environ. 711, 134514. https://doi.org/10.1016/j.scitotenv.2019.134514

Costache, R., Pham, Q.B., Sharifi, E., Linh, N.T.T., Abba, S., Vojtek, M., Vojteková, J., Nhi, P.T.T., Khoi, D.N., 2020c. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. Remote Sens. 12, 106.

Costache, R., Tien Bui, D., 2020. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. Sci. Total Environ. 712C, 136492. https://doi.org/10.1016/j.scitotenv.2019.136492

Costache, R., Zaharia, L., 2017. Flash-flood potential assessment and mapping by integrating the weights-of-evidence and frequency ratio statistical methods in GIS environment– case study: Bâsca Chiojdului River catchment (Romania). J. Earth Syst. Sci. 126, 59.

Dahri, N., Abida, H., 2017. Monte Carlo simulation-aided analytical hierarchy process (AHP) for flood susceptibility mapping in Gabes Basin (southeastern Tunisia). Environ. Earth Sci. 76, 302.

Dano, U.L., 2021. An AHP-based assessment of flood triggering factors to enhance resiliency in Dammam, Saudi Arabia. GeoJournal 1–16.

Dano, U.L., 2020. Flash flood impact assessment in Jeddah City: An analytic hierarchy process approach. Hydrology 7, 10.

Dewals, B., Giron, E., Ernst, J., Hecq, W., Pirotton, M., 2008. Integrated assessment of flood protection measures in the context of climate change: hydraulic modelling and economic approach. WIT Trans. Ecol. Environ. 108, 149–159.

Dodangeh, E., Choubin, B., Eigdir, A.N., Nabipour, N., Panahi, M., Shamshirband, S., Mosavi, A., 2020. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Sci. Total Environ. 705, 135983.

Dottori, F., Martina, M.L.V., Figueiredo, R., 2018. A methodology for flood susceptibility and vulnerability analysis in complex flood scenarios. J. Flood Risk Manag. 11, S632–S645.

Dou, J., Yamagishi, H., Zhu, Z., Yunus, A.P., Chen, C.W., 2018. TXT-tool 1.081-6.1 A comparative study of the binary logistic regression (BLR) and artificial neural network (ANN) models for GIS-based spatial predicting landslides at a regional scale, in: Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools. Springer, pp. 139–151.

Fontanine, I., Costache, R., 2013. Using GIS techniques for surface runoff potential analysis in the Subcarpathian area between Buzău and Slănic rivers, in Romania. Cinq Cont. 3, 47–57.

Fowler, H., Wilby, R., 2010. Detecting changes in seasonal precipitation extremes using regional climate model projections: Implications for managing fluvial flood risk. Water Resour. Res. 46.

Gessesse, B., Bewket, W., Bräuning, A., 2015. Model-based characterization and monitoring of runoff and soil erosion in response to land use/land cover changes in the Modjo watershed, Ethiopia. Land Degrad. Dev. 26, 711–724.

Ghosh, A., Kar, S.K., 2018. Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. Nat. Hazards 94, 349–368.

Hapciuc, O.-E., Romanescu, G., Minea, I., Iosub, M., Enea, A., Sandu, I., 2016. Flood susceptibility analysis of the cultural heritage in the Sucevita catchment (Romania). Int. J. Conserv. Sci. 7.

Hofmann, J., Schüttrumpf, H., 2019. Risk-based early warning system for pluvial flash floods: Approaches and foundations. Geosciences 9, 127.

Hong, H., Pradhan, B., Xu, C., Bui, D.T., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena 133, 266–281.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X., Chen, W., 2018. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. Sci. Total Environ. 625, 575–588.

Hosseini, F.S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., Haghighi, A.T., 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci. Total Environ. 711, 135161.

Hosseini, F.S., Sigaroodi, S.K., Salajegheh, A., Moghaddamnia, A., Choubin, B., 2021. Towards a flood vulnerability assessment of watershed using integration of decision-making trial and evaluation laboratory, analytical network process, and fuzzy theories. Environ. Sci. Pollut. Res. 1–12.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T.V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B.T., Lee, S., 2019. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. Sustainability 11, 5426.

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.

Kavzoglu, T., Sahin, E.K., Colkesen, I., 2014. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11, 425–439.

Khosravi, K., Nohani, E., Maroufinia, E., Pourghasemi, H.R., 2016. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Nat. Hazards 83, 947–987.

Leandro, J., Djordjević, S., Chen, A., Savić, D., Stanić, M., 2011. Calibration of a 1D/1D urban flood model using 1D/2D model results in the absence of field data. Water Sci. Technol. 64, 1016–1024.

Lee, B.-J., Kim, S., 2019. Gridded flash flood risk index coupling statistical approaches and TOPLATS land surface model for mountainous areas. Water 11, 504.

Lei, X., Chen, W., Avand, M., Janizadeh, S., Kariminejad, N., Shahabi, Hejar, Costache, R., Shahabi, Himan, Shirzadi, A., Mosavi, A., 2020. GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sens. 12, 2478.

Markolf, S.A., Hoehne, C., Fraser, A., Chester, M.V., Underwood, B.S., 2019. Transportation resilience to climate change and extreme weather events–Beyond risk and robustness. Transp. Policy 74, 174–186.

Midi, H., Sarkar, S.K., Rana, S., 2010. Collinearity diagnostics of binary logistic regression model. J. Interdiscip. Math. 13, 253–267.

Miles, J., 2014. Tolerance and variance inflation factor. Wiley StatsRef Stat. Ref. Online.

Mohammady, M., Pourghasemi, H.R., Amiri, M., 2019. Assessment of land subsidence susceptibility in Semnan plain (Iran): a comparison of support vector machine and weights of evidence data mining algorithms. Nat. Hazards 1–21.

Mosavi, A., Golshan, M., Janizadeh, S., Choubin, B., Melesse, A.M., Dineva, A.A., 2020. Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. Geocarto Int. 1–20.

Mujumdar, P., 2001. Flood wave propagation. Resonance 6, 66–73.

Ngo, P.-T., Hoang, N.-D., Pradhan, B., Nguyen, Q., Tran, X., Nguyen, V., Samui, P., Tien Bui, D., 2018. A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using Sentinel-1 SAR imagery and geospatial data. Sensors 18, 3704.

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Comput. Vis. Graph. Image Process. 28, 323–344.

Pham, B.T., Nguyen-Thoi, T., Qi, C., Van Phong, T., Dou, J., Ho, L.S., Van Le, H., Prakash, I., 2020. Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping. Catena 195, 104805.

Pham, B.T., Pradhan, B., Bui, D.T., Prakash, I., Dholakia, M., 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). Environ. Model. Softw. 84, 240–250.

Pham, B.T., Shirzadi, A., Bui, D.T., Prakash, I., Dholakia, M., 2018. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India. Int. J. Sediment Res. 33, 157–170.

Pourghasemi, H.R., Beheshtirad, M., Pradhan, B., 2016. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. Geomat. Nat. Hazards Risk 7, 861–885.

Pradhan, B., 2010. Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. Adv. Space Res. 45, 1244–1256.

Prăvălie, R., Costache, R., 2014. The analysis of the susceptibility of the flash-floods' genesis in the area of the hydrographical basin of Bāsca Chiojdului river/Analiza susceptibilitatii genezei viiturilor īn aria bazinului hidrografic al rāului Bāsca Chiojdului. Forum Geogr. 13, 39–49. http://dx.doi.org/10.5775/fg.2067-4635.2014.089.i

Prăvălie, R., Costache, R., 2013. The vulnerability of the territorial-administrative units to the hydrological phenomena of risk (flash-floods). Case study: the subcarpathian sector of Buzău catchment. Analele Univ. Din Oradea–Seria Geogr. 23, 91–98.

Qasem, S.N., Shamsuddin, S.M., 2011. Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. Appl. Soft Comput. 11, 1427–1438.

Saaty, T.L., 1980. The analytical hierarchy process, planning, priority. Resour. Alloc. RWS Publ. USA.

Sajedi-Hosseini, F., Choubin, B., Solaimani, K., Cerdà, A., Kavian, A., 2018. Spatial prediction of soil erosion susceptibility using a fuzzy analytical network process: application of the fuzzy decision making trial and evaluation laboratory approach. Land Degrad. Dev. 29, 3092–3103.

Sun, C.-C., 2010. A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods. Expert Syst. Appl. 37, 7745–7754.

Talukdar, S., Ghose, B., Salam, R., Mahato, S., Pham, Q.B., Linh, N.T.T., Costache, R., Avand, M., 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. Stoch. Environ. Res. Risk Assess. 1–24.

Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., Moore, J.H., 2018. Relief-based feature selection: introduction and review. J. Biomed. Inform.

Vakhshoori, V., Zare, M., 2018. Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? Geomat. Nat. Hazards Risk 9, 249–266.

Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., Shirzadi, A., Shahabi, H., Panahi, S., Costache, R., 2019. Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. J. Environ. Manage. 247, 712–729.

Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7, 161–187.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14, 1–37.

Yariyan, P., Janizadeh, S., Phong, T.V., Nguyen, H.D., Costache, R., Le, H.V., Pham, B.T., Pradhan, B., Tiefenbacher, J.P., 2020. Improvement of Best First Decision Trees Using Bagging and Dagging Ensembles for Flood-risk Mapping. Water Resour. Manag. https://doi.org/10.1007/s11269-020-02603-7

Yeon, Y.-K., Han, J.-G., Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol. 116, 274–283.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856.

Zaharia, L., Costache, R., Prăvălie, R., Ioana-Toroimac, G., 2017. Mapping flood and flooding potential indices: a methodological approach to identifying areas susceptible to flood and flooding risk. Case study: the Prahova catchment (Romania). Front. Earth Sci. 11, 229–247.

Zaharia, L., Costache, R., Prăvălie, R., Minea, G., 2015. Assessment and mapping of flood potential in the Slănic catchment in Romania. J. Earth Syst. Sci. 124, 1311–1324.

Zare, M., Pourghasemi, H.R., Vafakhah, M., Pradhan, B., 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab. J. Geosci. 6, 2873–2888.

Zhang, C.-X., Wang, G.-W., Zhang, J.-S., 2012. An empirical bias–variance analysis of DECORATE ensemble method at different training sample sizes. J. Appl. Stat. 39, 829–850.

Zhao, G., Pang, B., Xu, Z., Peng, D., Xu, L., 2019. Assessment of urban flood susceptibility using semi-supervised machine learning model. Sci. Total Environ. 659, 940–949.

Figure captions

Fig. 1 Study area location within Romania (Source: SRTM, 30 m and field survey database processing)

Fig. 2 Flash-Flood and Flood Predictors (a. Slope; b. Land use; c. Lithology; d. Hydrological Soil Group; e. Plan curvature; f. Profile curvature)

Fig. 3 Flash-Flood and Flood Predictors (a. Convergence Index; b. Modified Fournier Index; c. Aspect; d. TWI)

Fig. 4 Scheme of the workflow applied in the present research

Fig. 5 Sensitivity and Specificity values according to classification cutoff

Fig. 6 Classification plot of the observed groups and predicted probabilities

Fig. 7. Flash-Flood Potential Index (a. WOE; b. AHP-WOE; c. LR-WOE; d. CART-WOE; e. RBFNN-WOE)

Fig. 8 Optimal models architectures (a. CART-WOE; b. RBFNN-WOE)

Fig. 9 Performance indicator of RBFNN-WOE ensemble (a. ROC Curve; b. Pseudo-probability plot; c. Lift chart; d. Gain chart)

Fig. 10 Weights of FFPI classes

Fig. 11 ROC Curve (a. Success Rate; b. Prediction Rate)

Fig. 12 Flash-Flood Propagation Susceptibility Index (a. WOE; b. AHP-WOE; c. LR-WOE; d. CART-WOE; e. RBFNN-WOE)

Fig. 13 Weights of FFPSI classes

List of tables

Table 2 Multicollinearity assessment and feature selection

Table 3 Weights of Evidence values

Table 4 Pair-wise comparison matrix and normalized weights for each factor

Table 5 Confusion matrices computed for training phase of LR-WOE, CART-WOE and RBFNN-WOE models

Table 6 Importance of flash-flood predictors to FFPI models

Table 7 Statistical metrics used to validate the FFPI results

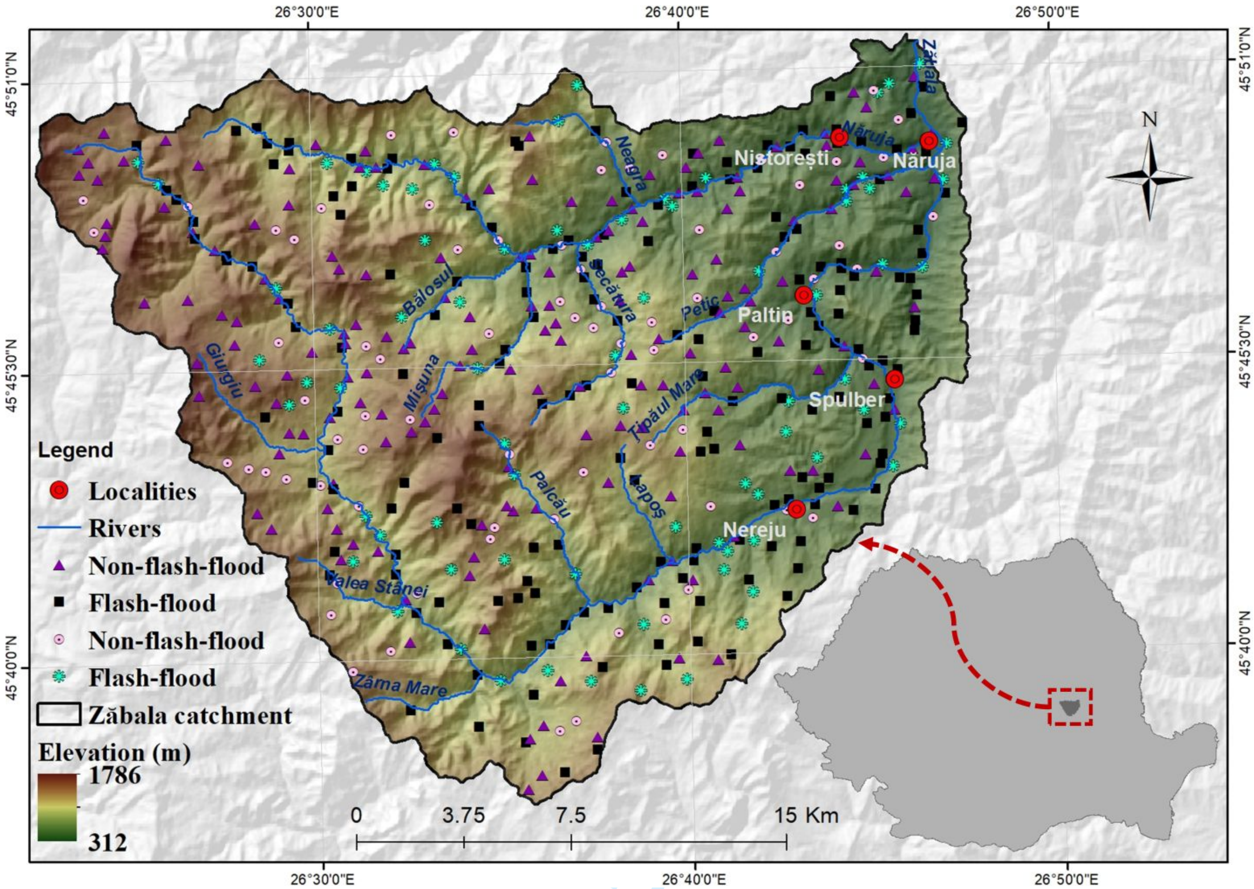


Fig. 1 Study area location within Romania

(Source: SRTM, 30 m and field survey database processing)

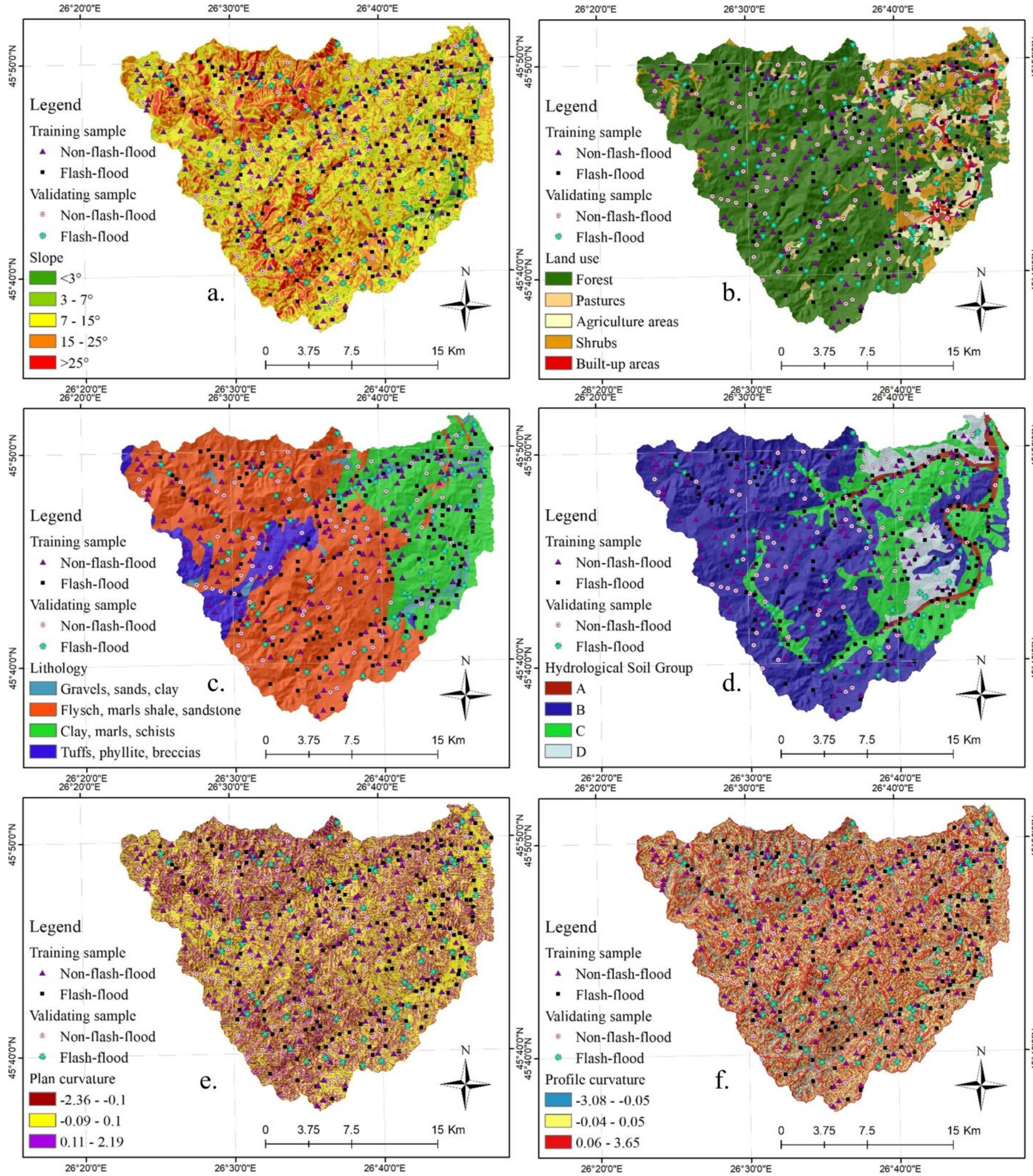


Fig. 2 Flash-Flood and Flood Predictors (a. Slope; b. Land use; c. Lithology; d. Hydrological Soil Group; e. Plan curvature; f. Profile curvature)

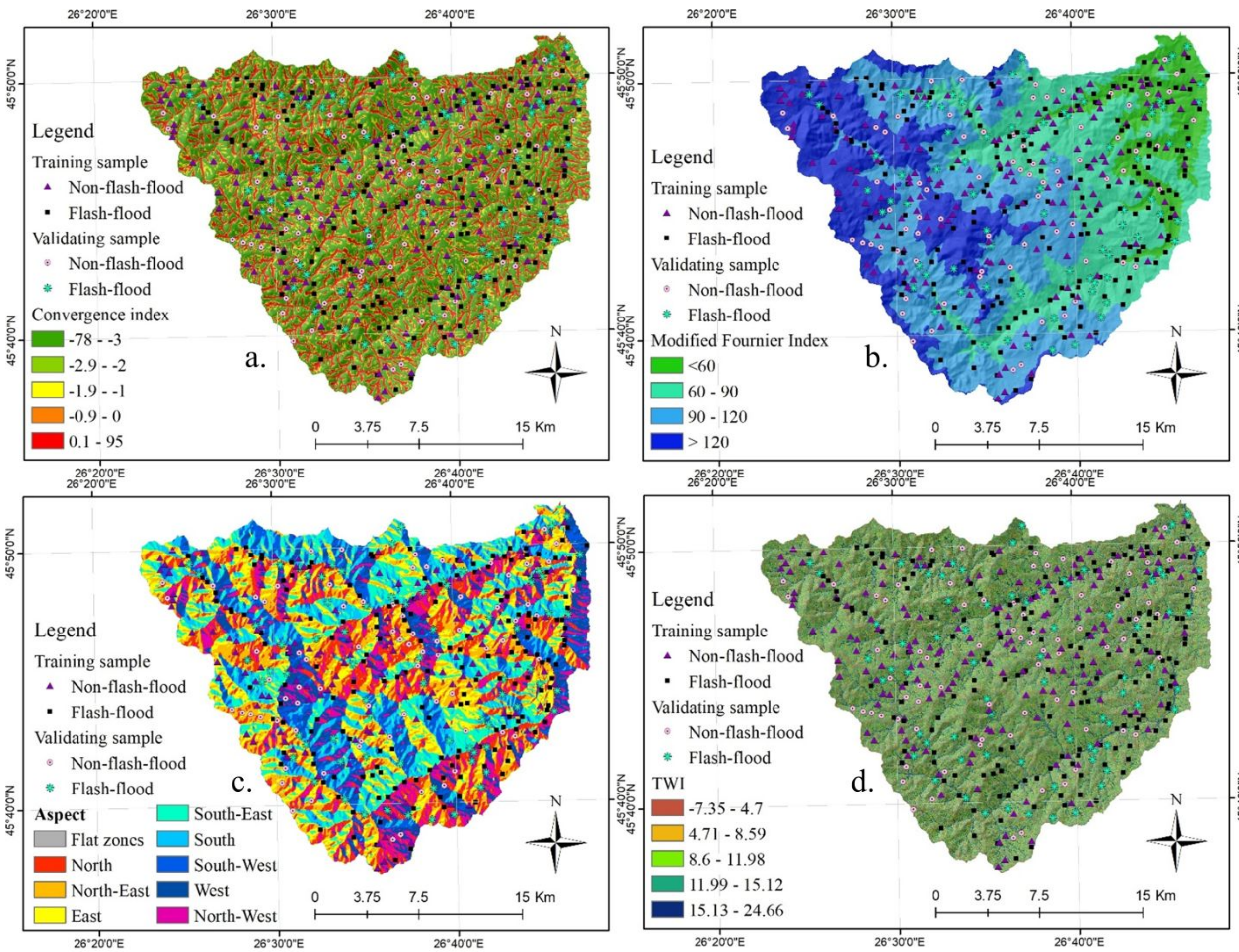


Fig. 3 Flash-Flood and Flood Predictors (a. Convergence Index; b. Modified Fournier Index; c. Aspect; d. TWI)

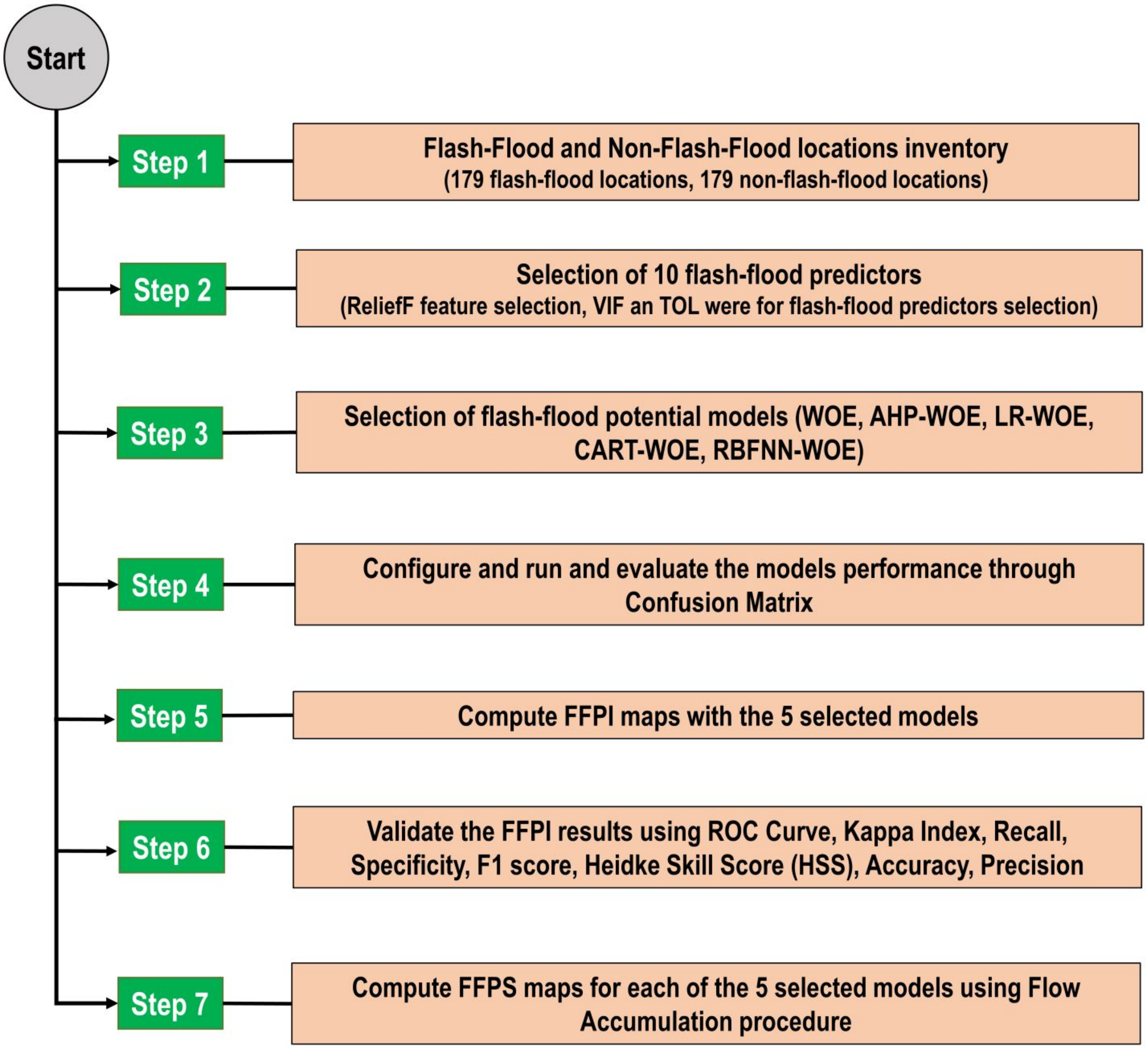


Fig. 4 Scheme of the workflow applied in the present research

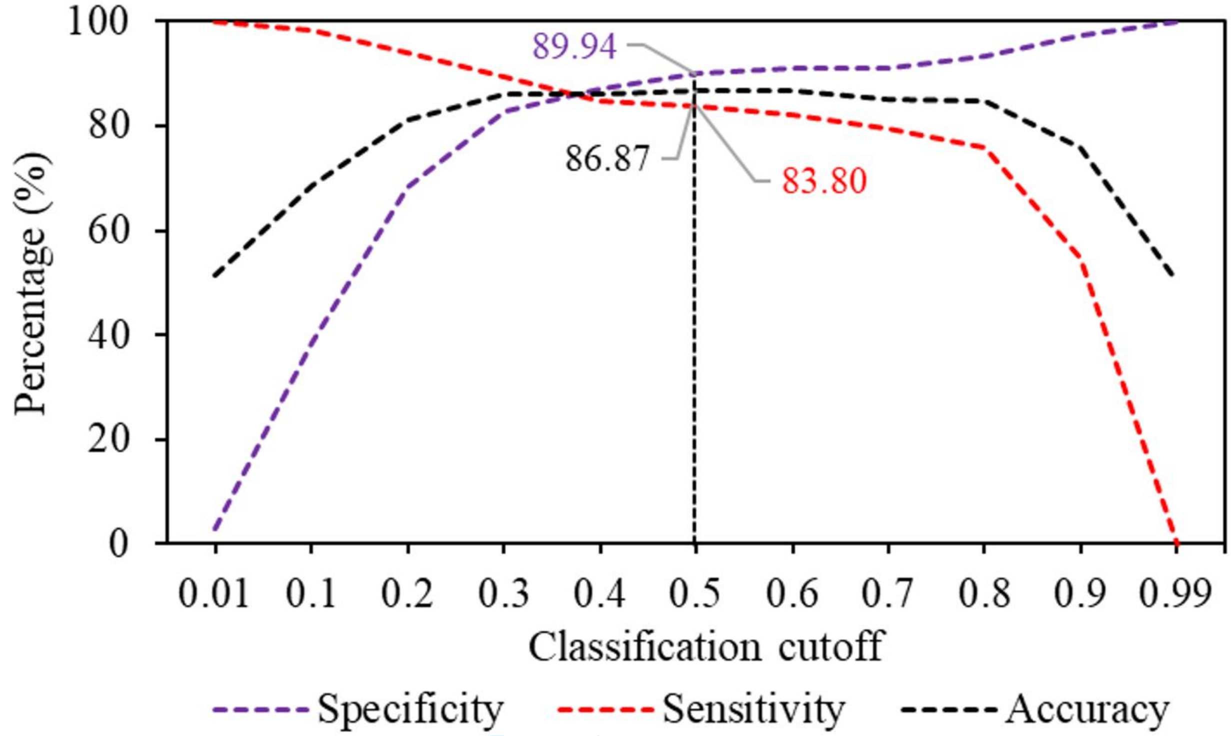


Fig. 5 Sensitivity and Specificity values according to classification cutoff

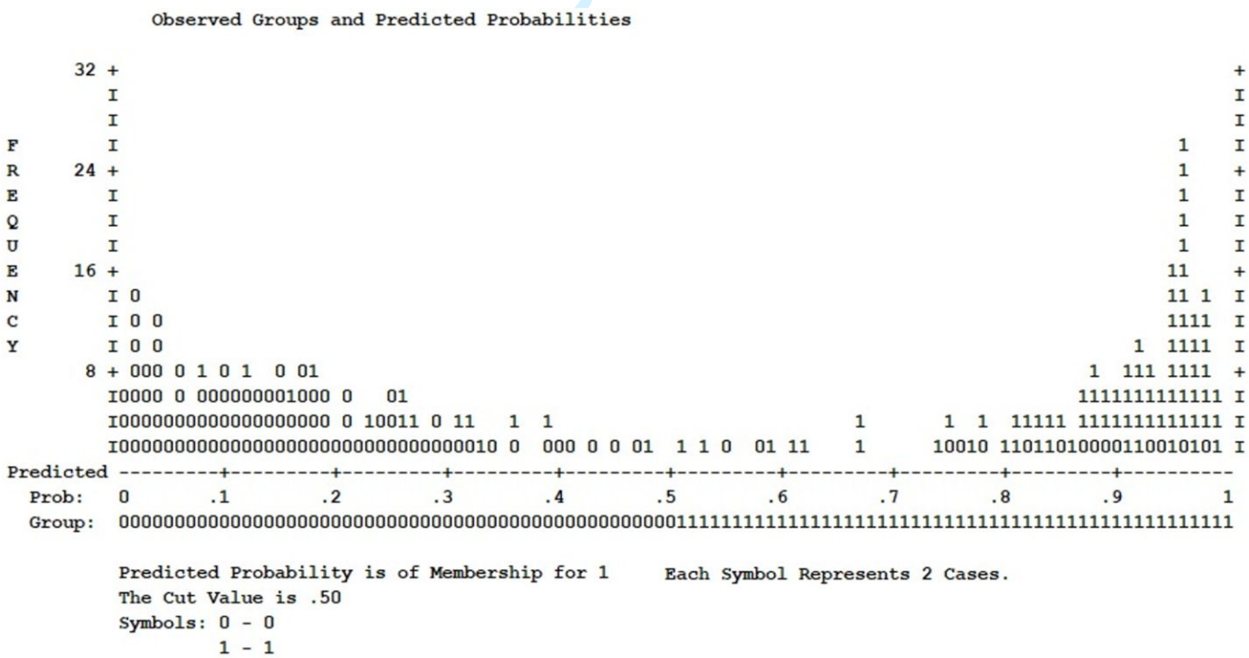


Fig. 6 Classification plot of the observed groups and predicted probabilities

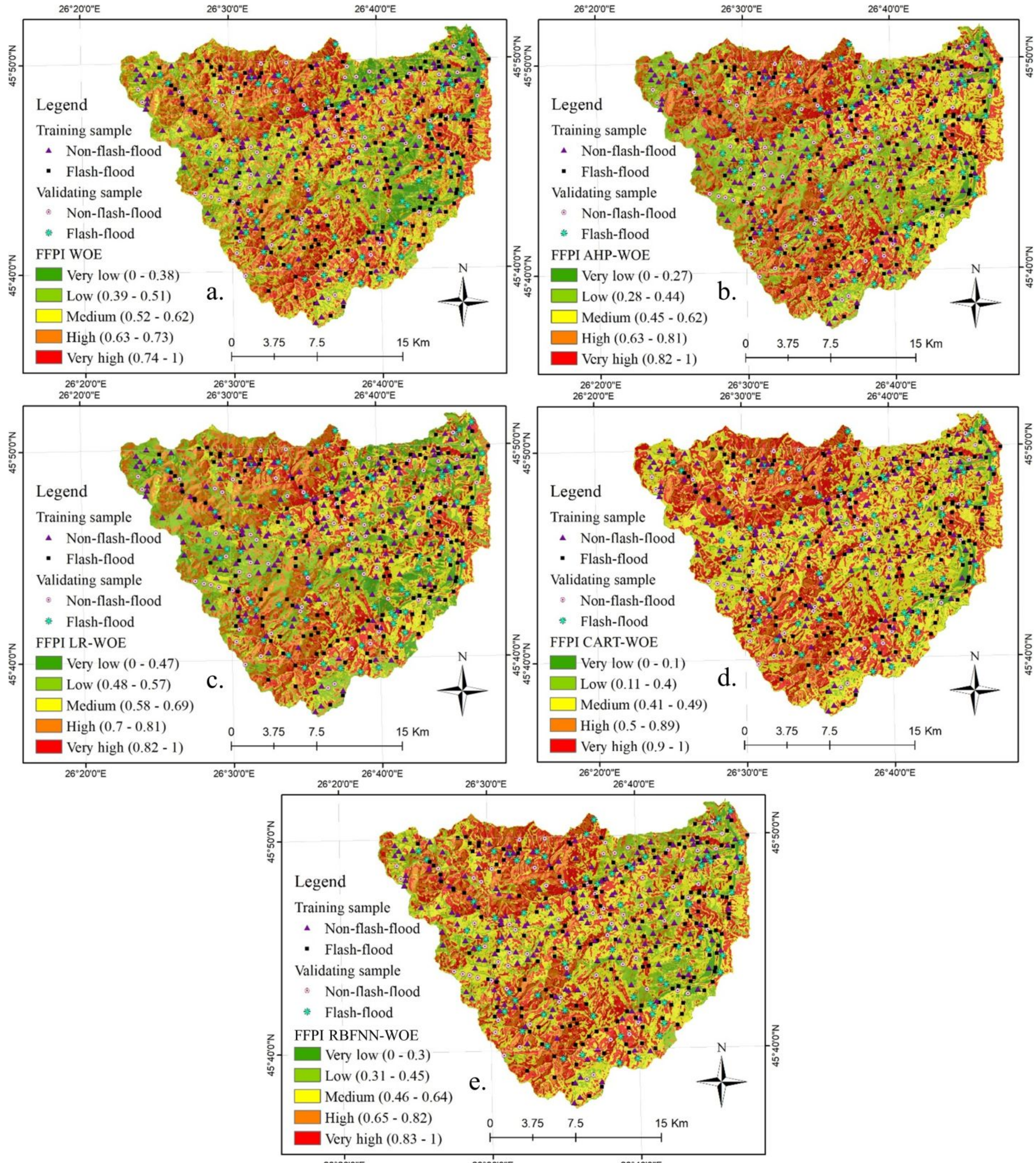


Fig. 7. Flash-Flood Potential Index (a. WOE; b. AHP-WOE; c. LR-WOE; d. CART-WOE; e. RBFNN-WOE)

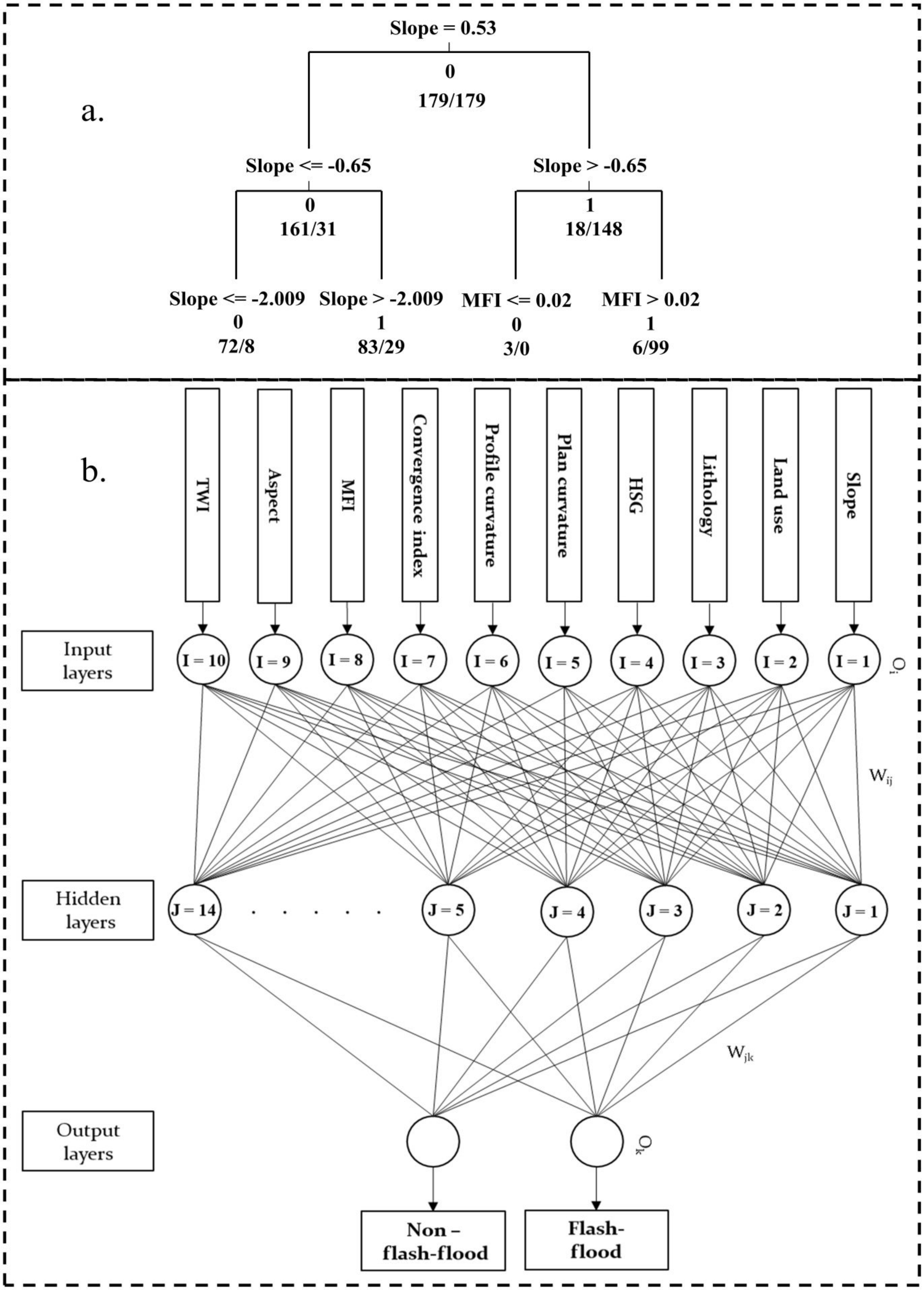


Fig. 8 Optimal models architectures (a. CART-WOE; b. RBFNN-WOE)

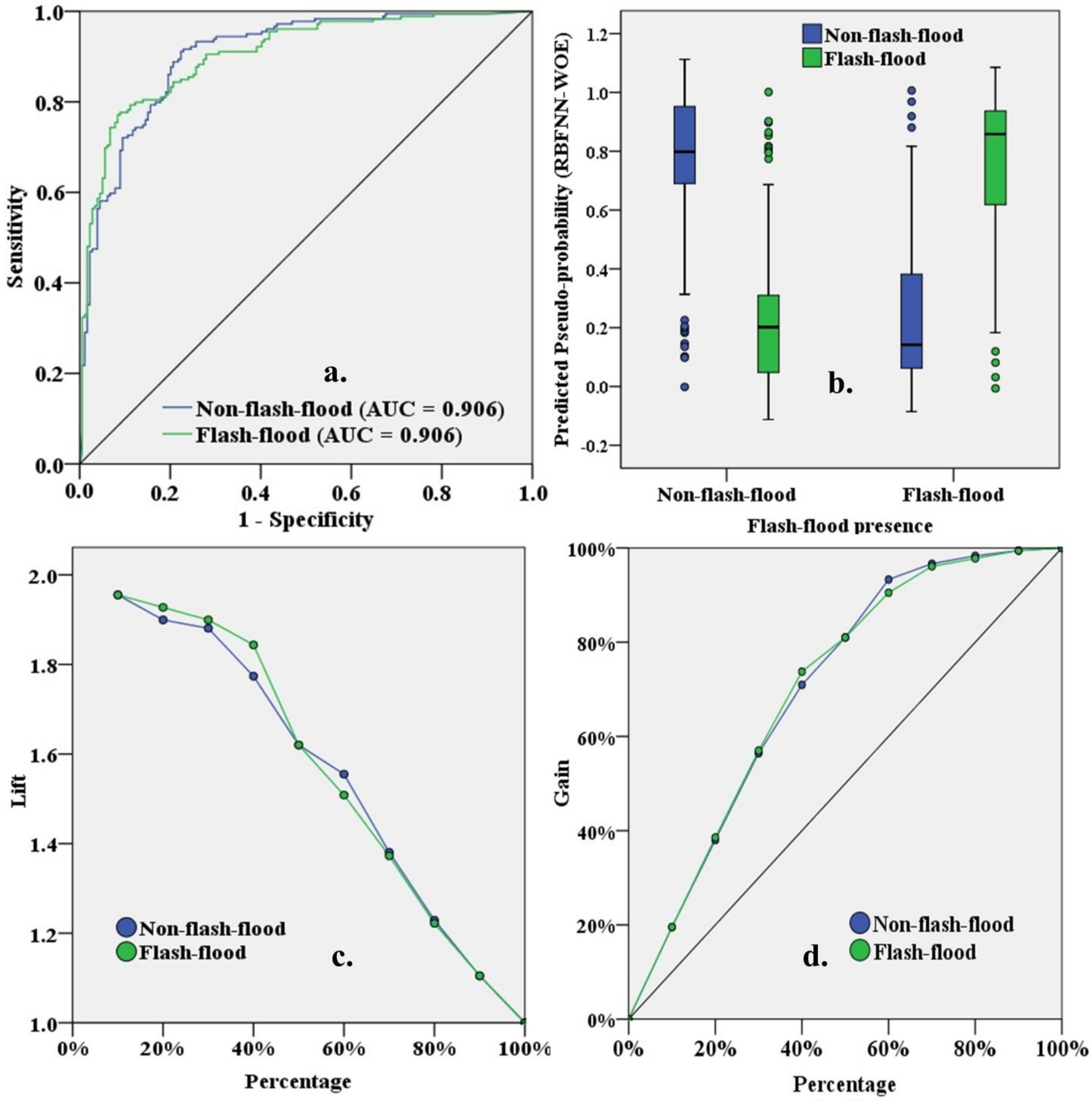


Fig. 9 Performance indicator of RBFNN-WOE ensemble (a. ROC Curve; b. Pseudo-probability plot; c. Lift chart; d. Gain chart)

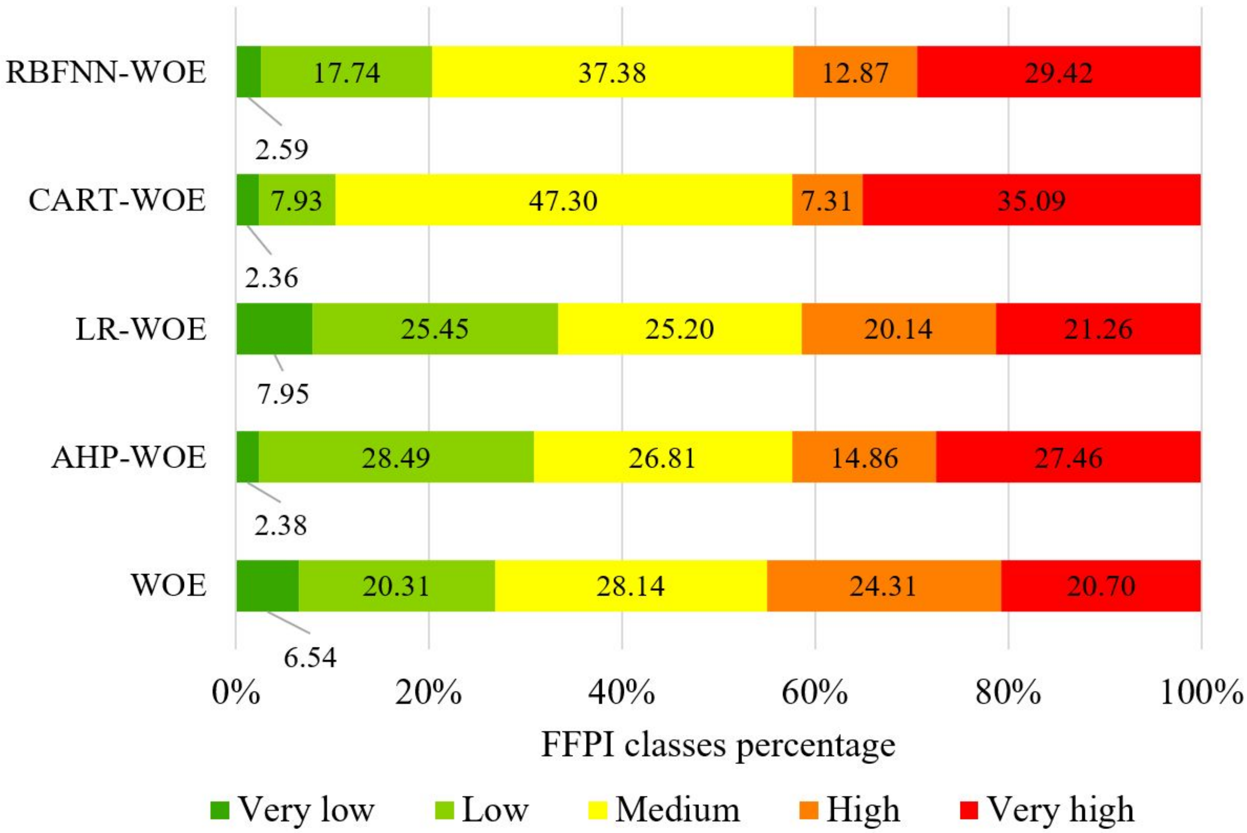
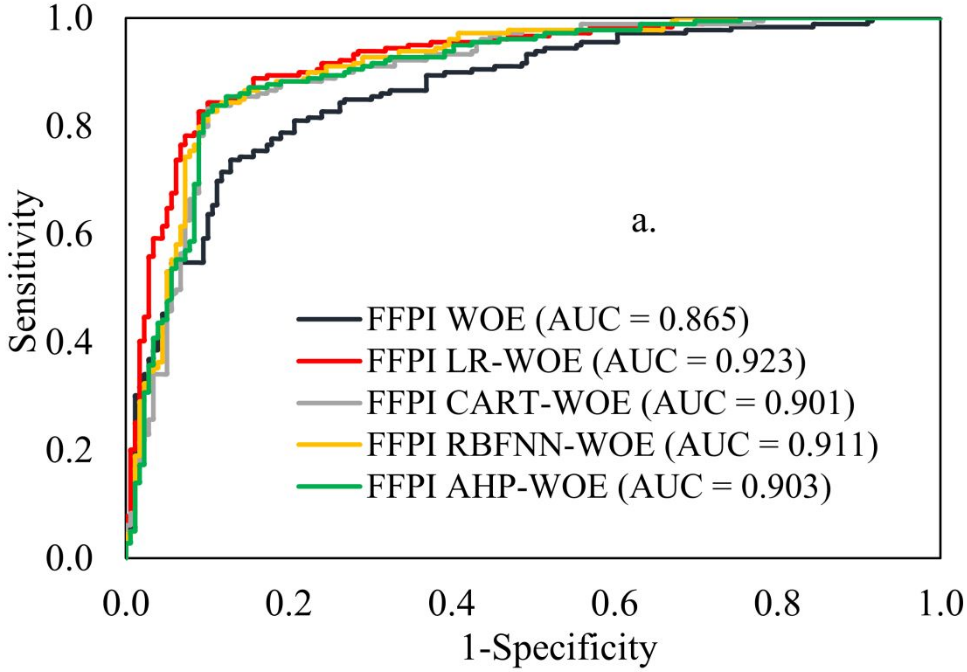


Fig. 10 Weights of FFPI classes



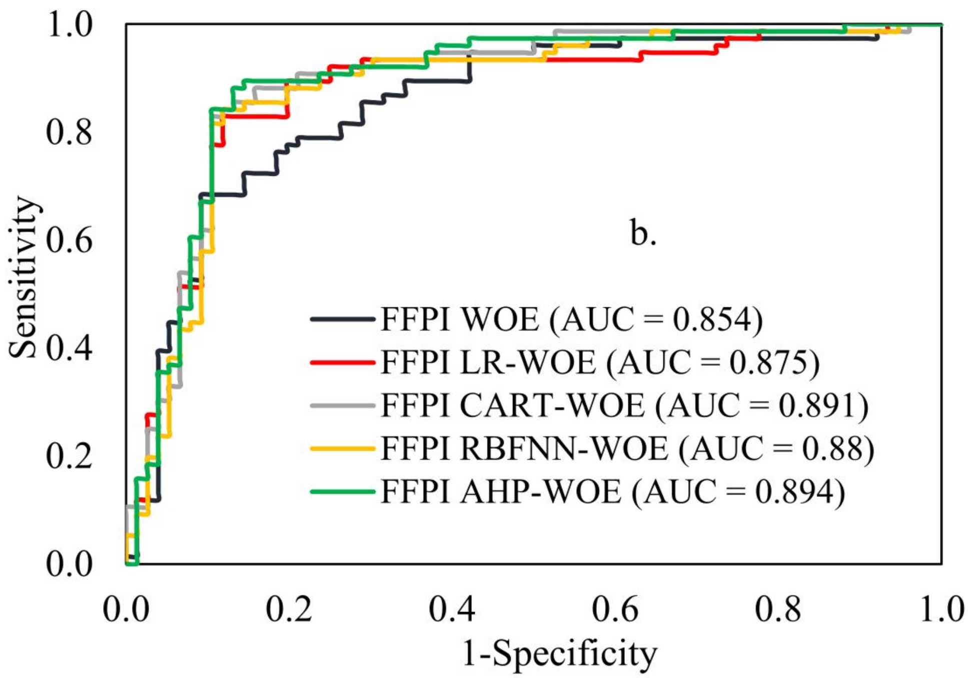


Fig. 11 ROC Curve (a. Success Rate; b. Prediction Rate)

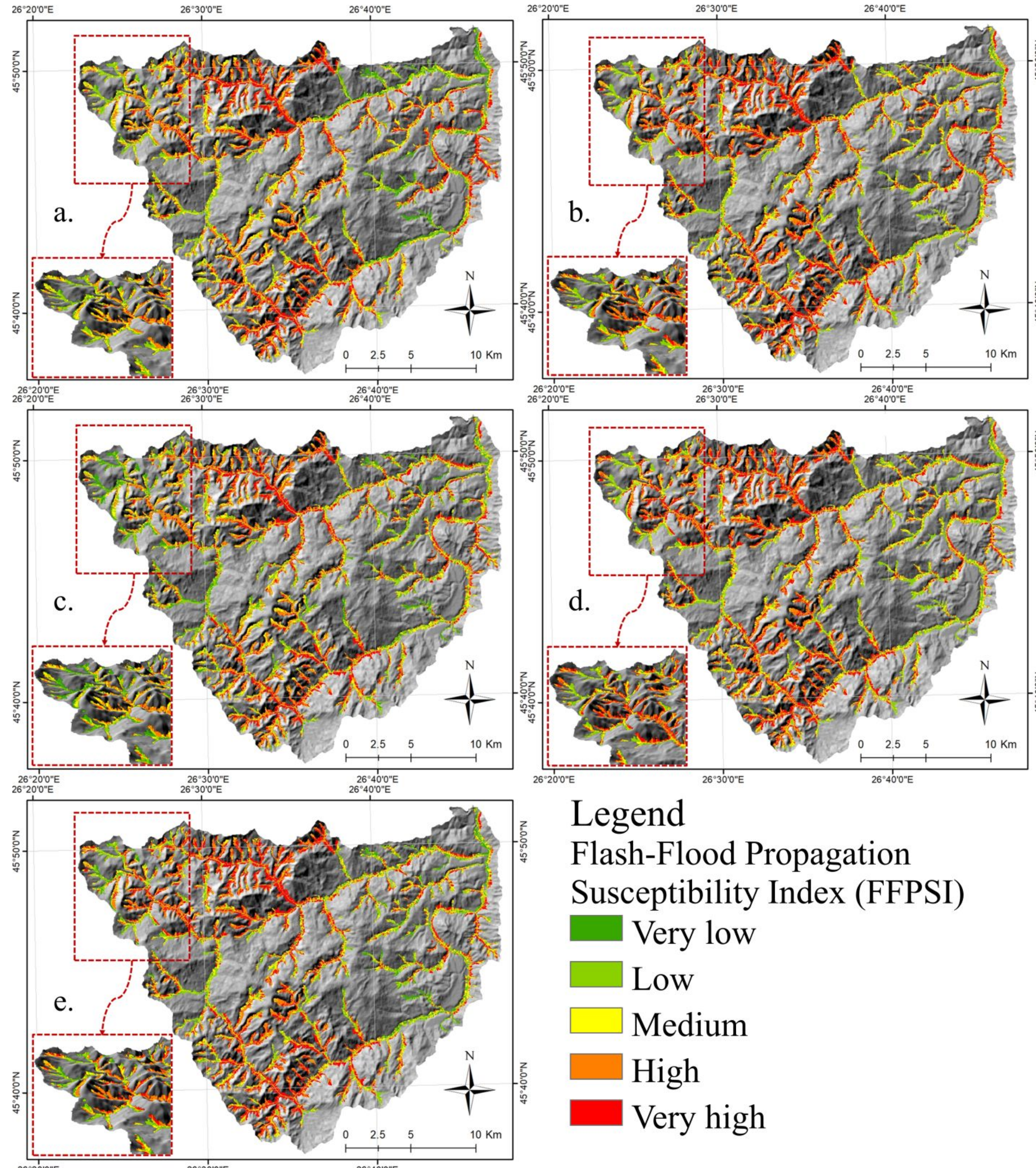


Fig. 12 Flash-Flood Propagation Susceptibility Index (a. WOE; b. AHP-WOE; c. LR-WOE; d. CART-WOE; e. RBFNN-WOE)

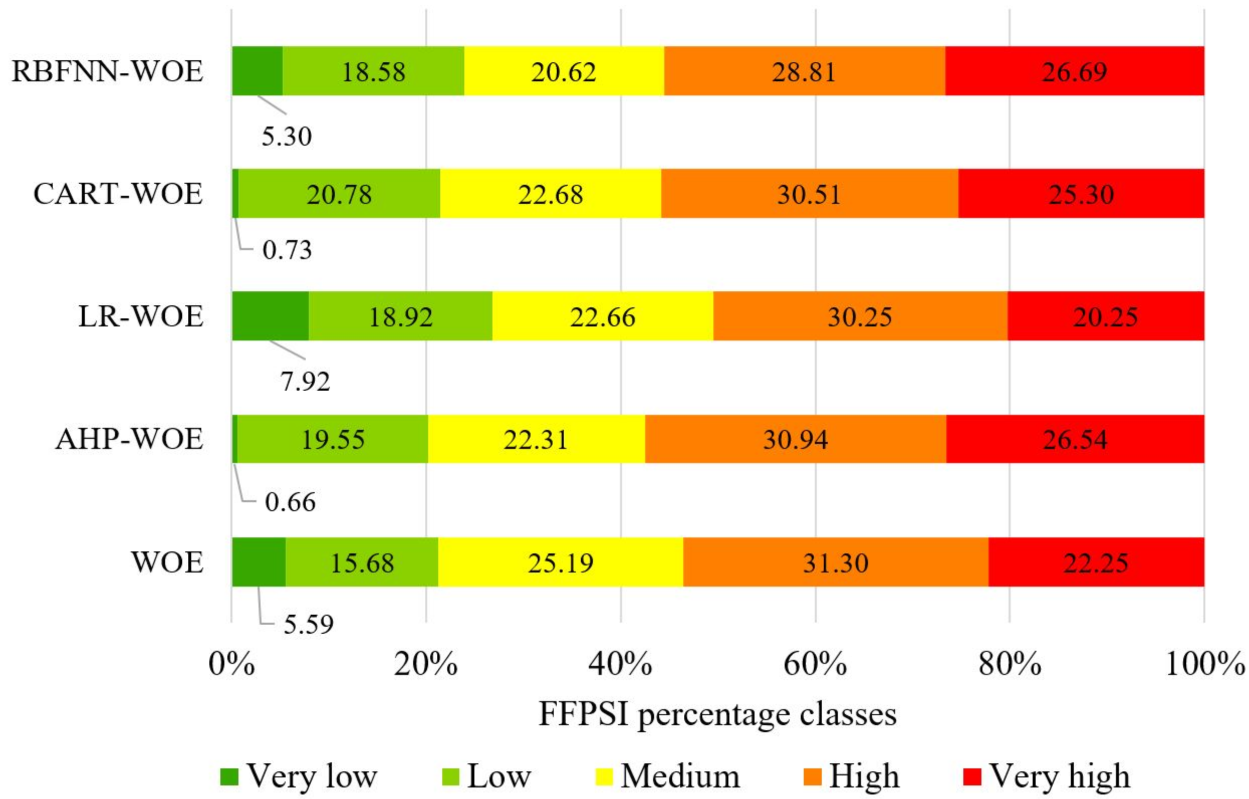


Fig. 13 Weights of FFPSI classes

Table 1 Data used, source, resolution, scale and type

Data	Source	Resolution	Scale	Type
Digital Elevation Model (DEM)	Shuttle Radar Topography Mission (SRTM)	30 m	-	Spatial
Flash-Flood points	General Inspectorate for Emergency Situation (GIES) of Romania; mass-media	-	-	Spatial
Non-Flash-Flood points	Aerial imagery; field survey	-	-	Spatial
Rainfall (mm/year)	Worldclim v2	-	-	Spatial
Land use/cover	Corine Land Cover, 2018	1 km	-	Spatial
Hydrological Soil Groups	Digital Soil Map of Romania	-	1:200000	Spatial
Lithology	Digital Geological Map of Romania	-	1:200000	Spatial

Table 2 Multicollinearity assessment and feature selection

Flash-Flood Predictor	TOL	VIF	ReliefF Attribute
HSG	0.486	2.059	0.014
Profile curvature	0.917	1.091	0.016
Slope	0.785	1.273	0.215
Plan curvature	0.874	1.144	0.039
Lithology	0.411	2.432	0.003
MFI	0.417	2.398	0.036
Aspect	0.976	1.025	0.006
Convergence Index	0.672	1.488	0.026
Land use	0.943	1.061	0.015
TWI	0.795	1.258	0.019

Table 3 Weights of Evidence values

Factor	Class	*Class pixels*	*Flash-Flood pixels*	*WOE coefficients*
Slope	< 3°	14392	0	-4.01

	3 – 7°	48164	2	-2.28
	7 – 15°	288025	30	-1.50
	15 – 25°	213187	123	-1.50
	> 25°	43974	24	-1.50
Land use	Forest	476027	145	0.06
	Pastures	17399	8	0.36
	Agriculture areas	38914	13	-1.50
	Shrubs	64551	7	-1.50
	Built-up areas	10845	6	-1.50
Lithology	Gravels, sands, clay	22179	11	0.63
	Flysch, marls shale, sandstone	393103	102	-2.50
	Clay, marls, schists	146757	60	-2.50
	Tuffs, phyllite, breccias	45692	6	-0.77
HSG	A	383646	25	-2.16
	B	20939	70	3.09
	C	48205	65	2.09
	D	154946	19	-2.50
Plan curvature	-2.36 – -0.1	144551	56	0.40
	-0.09 – 0.1	329253	91	-0.12
	0.11 – 2.19	133938	32	-0.24
Profile curvature	-3.08 – -0.05	72259	33	0.56
	-0.04 – 0.05	240659	75	0.14
	0.04 – 3.65	294824	71	-0.32
Convergence index	-78 - -3	312086	107	0.25
	-2.9 - -2	65245	24	0.16
	-1.9 - -1	55305	19	0.08
	-0.9 – 0	42598	6	-0.87
	0.1 - 95	132508	23	-0.73
MFI	< 60	70506	44	0.89
	60 – 90	178411	71	0.44
	90 – 120	216598	49	-0.40
	> 120	142227	15	-2.50
Aspect	Flat surfaces	1994	0	-1.85
	North	81189	28	0.11
	North-East	98234	39	0.30
	East	88164	31	0.14
	South-East	86794	17	-0.54
	South	66559	17	-0.23
	South-West	57424	16	-0.13
	West	59678	13	-0.40
	North-East	67706	18	-0.19
TWI	-7.35 - 4.7	69875	17	-0.26
	4.71 – 8.59	143756	42	-0.06
	8.6 – 11.98	79221	22	-0.12
	11.99 – 15.12	261587	89	0.22
	15.13 – 24.66	53278	9	-0.65

Table 4 Pair-wise comparison matrix and normalized weights for each factor

Factor and classes/categories	***Pair-wise comparison matrix***										***Normalized weights***
Factors	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	

[1] Slope angle	1										0.260
[2] Land use	1/2	1									0.179
[3] Lithology	1/3	1/2	1								0.111
[4] HSG	1/5	1/4	1/2	1							0.045
[5] Plan curvature	1/3	1/2	1	3	1						0.085
[6] Profile curvature	1/4	1/3	1/2	2	2	1					0.110
[7] Convergence index	1/4	1/3	1/2	2	1	1/2	1				0.050
[8] MFI	1/4	1/3	1/2	2	1	1/4	2	1			0.070
[9] Aspect	1/7	1/6	1/5	1/3	1/4	1/4	4	1/3	1		0.040
[10] TWI	1/5	1/4	1/3	1	1/2	1/3	2	1/2	3	1	0.050

Table 5 Confusion matrices computed for training phase of LR-WOE, CART-WOE and RBFNN-WOE models

	Observed	**Predicted**		**Percent Correct**
		0	1	
LR-WOE	0	161	18	89.94%
	1	29	150	83.79%
	Overall Percentage	51.76%	48.24%	86.87%
CART-WOE	0	161	18	89.94%
	1	31	148	82.68%
	Overall Percentage	53.63%	46.37%	86.31%
RBFNN-WOE	0	156	23	87.15%
	1	31	148	82.68%
	Overall Percentage	51.32%	48.68%	84.91%

Table 6 Importance of flash-flood predictors to FFPI models

Predictors	LR-WOE (β_i)	CART-WOE	RBFNN-WOE
Aspect	1.253	0.003	0.073
Convergence Index	0.102	0.081	0.074
HGS	0.098	0.029	0.062
Land use	1.603	0.001	0.122
Lithology	0.099	0.015	0.078
MFI	0.835	0.059	0.102
Plan curvature	0.377	0.034	0.078
Profile curvature	-0.215	0.002	0.069
Slope	1.274	0.559	0.251
TWI	0.431	0.037	0.091

Table 7 Statistical metrics used to validate the FFPI results

Metrics	Training dataset					Validating dataset				
	FFPI WOE	FFPI LR-WOE	FFPI CART-WOE	FFPI RBFNN-WOE	FFPI AHP-WOE	FFPI WOE	FFPI LR-WOE	FFPI CART -WOE	FFPI RBFNN-WOE	FFPI AHP-WOE

TP	137	151	144	149	147	55	65	63	67	69
TN	147	163	159	161	161	63	67	67	64	65
FP	42	28	35	30	32	16	11	13	9	7
FN	32	16	20	18	18	13	9	9	12	11
Sensitivity	0.811	0.904	0.878	0.892	0.891	0.809	0.878	0.875	0.848	0.863
Specificity	0.778	0.853	0.820	0.843	0.834	0.797	0.859	0.838	0.877	0.903
Accuracy	0.793	0.877	0.846	0.866	0.860	0.803	0.868	0.855	0.862	0.882
K index	0.587	0.754	0.693	0.732	0.721	0.604	0.737	0.711	0.724	0.763
Precision	0.765	0.844	0.804	0.832	0.821	0.775	0.855	0.829	0.882	0.908
F1 score	0.787	0.873	0.840	0.861	0.855	0.791	0.867	0.851	0.865	0.885

Change of authorship request form (pre-acceptance)

Please read the important information on page 4 before you begin

This form should be used by authors to request any change in authorship including changes in corresponding authors. Please fully complete all sections. Use black ink and block capitals and provide each author's full name with the given name first followed by the family name.

Please note: In author collaborations where there is formal agreement for representing the collaboration, it is sufficient for the representative or legal guarantor (usually the corresponding author) to complete and sign the Authorship Change Form on behalf of all authors.

Section 1: Please provide the current title of manuscript
(For journals: Please provide the manuscript ID, title and/or DOI if available.)
(For books: Please provide the title, ISBN and/or DOI if available.)

Manuscript ID no. in case of unpublished manuscript: TGEI-2021-0559
DOI in case of published manuscript:
ISBN (for books):

Title: **Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning**

Change of authorship request form (pre-acceptance)

Section 2: Please provide the previous authorship, in the order shown on the manuscript before the changes were introduced. Please indicate the corresponding author by adding (CA) behind the name.

	First name(s)	Family name	ORCID or SCOPUS id, if available
1st author	Romulus	Costache	https://orcid.org/0000-0002-6876-8572
2nd author	Quoc Bao	Pham	https://orcid.org/0000-0002-0468-5962
3rd author	Alireza	Arabameri	https://orcid.org/0000-0002-1142-1666
4th author	Daniel Constantin	Diaconu	https://orcid.org/0000-0002-0468-6703
5th author	Iulia	Costache	https://orcid.org/0000-0002-1779-8173
6th author	Anca	Crăciun	https://orcid.org/0000-0003-1113-906X
7th author	Nicu	Ciobotaru	https://orcid.org/0000-0002-6052-985X
8th author	Manish	Pandey	https://orcid.org/0000-0001-8291-2043
9th author	Aman	Arora	https://orcid.org/0000-0001-9396-8720
10th author	Sk Ajim	Ali	https://orcid.org/0000-0001-7488-5591
11th author	Binh Thai	Pham	https://orcid.org/0000-0001-9707-840X
12th author	Hoang	Nguyen	https://orcid.org/0000-0001-6122-8314
13th author	Hoang Anh	Tuan	

Please use an additional sheet if there are more than 7 authors.

Section 3: Please provide a justification for change. Please use this section to explain your reasons for changing the authorship of your manuscript, e.g. what necessitated the change in authorship? Please refer to the (journal) policy pages for more information about authorship. Please explain why omitted authors were not originally included and/or why authors were removed on the submitted manuscript.

The modification in the authorship position was based on the contribution. The new author was added because of his valuable contribution in the revised version of the manuscript. All authors agreed in this changes.

Change of authorship request form (pre-acceptance)

Section 4: Proposed new authorship. Please provide your new authorship list in the order you would like it to appear on the manuscript. Please indicate the corresponding author by adding (CA) behind the name. <u>If the corresponding author has changed, please indicate the reason under section 3.</u>

	First name(s)	Family name (this name will appear in full on the final publication and will be searchable in various abstract and indexing databases)
1st author	Romulus	Costache
2nd author	Quoc Bao	Pham
3rd author	Alireza	Arabameri
4th author	Daniel Constantin	Diaconu
5th author	Iulia	Costache
6th author	Anca	Crăciun
7th author	Nicu	Ciobotaru
8 th author	Manish	Pandey
9th author	Aman	Arora
10th author	Sk Ajim	Ali
11th author	Binh Thai	Pham
12th author	Hoang	Nguyen
13th author	Hoang Anh	Tuan
14th author	Mohammadtaghi	Avand

Please use an additional sheet if there are more than 7 authors.

Change of authorship request form (pre-acceptance)

Section 5: Author contribution, Acknowledgement and Disclosures. Please use this section to provide a new disclosure statement and, if appropriate, acknowledge any contributors who have been removed as authors and ensure you state what contribution any new authors made (if applicable per the journal or book (series) policy). Please ensure these are updated in your manuscript - after approval of the change(s) - as our production department will not transfer the information in this form to your manuscript.

New acknowledgements:

New Disclosures (financial and non-financial interests, funding):

New Author Contributions statement (if applicable per the journal policy): The new author was added because of his valuable contribution in the revised version of the manuscript.

State 'Not applicable' if there are no new authors.

Change of authorship request form (pre-acceptance)

Section 6: Declaration of agreement. *All* authors, unchanged, new and removed *must* sign this declaration.
(NB: Please print the form, sign and return a scanned copy. Please note that signatures that have been inserted as an image file are acceptable as long as it is handwritten. Typed names in the signature box are unacceptable.) *** Please delete as appropriate. Delete all of the bold if you were on the original authorship list and are remaining as an author.**

	First name	Family name		Signature	Affiliated institute	Date
1st author	Romulus	Costache			National Institute of Hydrology and Water Management, București-Ploiești Road, 97E, 1st District, 013686, Bucharest, Romania	17.09.2021
2nd author	Quoc Bao	Pham			Institute of Applied Technology, Thu Dau Mot University, Binh Duong province, Vietnam	17.09.2021
3rd author	Alireza	Arabameri			Department of Geomorphology, Tarbiat Modares University, Tehran 36581-17994, Iran	17.09.2021
4th authors	Daniel Constantin	Diaconu			Centre for Integrated Analysis and Territorial Management, University of Bucharest, Bucharest 010041, Romania	17.09.2021
5th author	Iulia	Costache			Department of Meteorology and Hydrology, Faculty of Geography, University of Bucharest, Bucharest 010041, Romania	17.09.2021
6th author	Anca	Crăciun			Danube Delta National Institute for Research and Development,165 Babadag Street, 820112, Tulcea, Romania	17.09.2021
7th author	Nicu	Ciobotaru			National Institute of Hydrology and Water Management, București-Ploiești Road, 97E, 1st District, 013686, Bucharest, Romania	17.09.2021

Change of authorship request form (pre-acceptance)

8 th author	Manish	Pandey			University Center for Research & Development (UCRD), Chandigarh University, Mohali-140413, Punjab, India	17.09.2021
9 th author	Aman	Arora			Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi- 10025, Delhi, India	17.09.2021
10 th author	Sk Ajim	Ali			Department of Geography, Faculty of Science, Aligarh Muslim University (AMU), Aligarh, UP 202002, India	17.09.2021
11 th author	Binh Thai	Pham			University of Transport Technology, Hanoi 100000, Vietnam	17.09.2021
12 th author	Hoang	Nguyen			Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Vien st., Duc Thang ward, Bac Tu Liem dist.,	17.09.2021
13 th author	Hoang Anh	Tuan			Faculty of Geodetic map and Land Management, Hanoi University of Mining and Geology, 18 Vien Street, Duc Thang Ward, Bac Tu Liem	17.09.2021
14th author	Mohammadtaghi	Avand			Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Tehran, P.O. Box 14115-	17.09.2021

Please use an additional sheet if there are more than 7 authors.

Change of authorship request form (pre-acceptance)

Important information. Please read.

- Please return this form, fully completed, to Springer Nature. We will consider the information you have provided to decide whether to approve the proposed change in authorship. We may choose to contact your institution for more information or undertake a further investigation, if appropriate, before making a final decision.
- By signing this declaration, all authors guarantee that the order of the authors are in accordance with their scientific contribution, if applicable as different conventions apply per discipline, and that only authors have been added who made a meaningful contribution to the work.
- Please note, we cannot investigate or mediate any authorship disputes. If you are unable to obtained agreement from all authors (including those who you wish to be removed) you must refer the matter to your institution(s) for investigation. Please inform us if you need to do this.
- If you are not able to return a fully completed form within **30 days** of the date that it was sent to the author requesting the change, we may have to withdraw your manuscript. We cannot publish manuscripts where authorship has not been agreed by all authors (including those who have been removed).
- Incomplete forms will be rejected.
