



EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
INSTITUTE OF CARTOGRAPHY AND GEOINFORMATICS

FLOOD SUSCEPTIBILITY MAPPING BASED ON SATELLITE IMAGES AND
MACHINE LEARNING METHODS IN HUNGARY

BY
ÖMER FARUK ELMA

MSc THESIS
DEPARTMENT OF CARTOGRAPHY AND GEOINFORMATICS

LECT. VARGA ZSÓFIA
(Supervisor)

BUDAPEST – 2022

DECLARATION

I, undersigned Ömer Faruk Elma (neptun code: bz6t9z), declare that the present master's thesis is my original intellectual product in full and that I have not submitted any part or the whole of this work to any other institution. Permissions related to the use of copyrighted sources in this work are attached.

I AGREE to the publication of the accepted master's thesis in pdf form on the website of the Institute of Cartography and Geo-informatics.

Budapest, 15th May 2022



(signature of the student)

Table of contents

FLOOD SUSCEPTIBILITY MAPPING BASED ON SATELLITE IMAGES AND MACHINE LEARNING METHODS IN HUNGARY	1
DECLARATION	I
Table of contents	II
Abstract	IV
Keywords: Geographic information system (GIS); Machine Learning, Flood susceptibility map, Hungary, Support Vector Machine, Random Forest, Logistic Regression.....	IV
1. Introduction	1
2. Materials and method	4
2.1. Study area.....	4
2.2. Data used	5
2.2.1. Preparation of the flood sample points.....	5
2.2.2. Generation of the factors affecting the floods in the GIS-based software.....	6
2.2.2.1. Elevation.....	9
2.2.2.2. Aspect.....	9
2.2.2.3. Slope.....	9
2.2.2.4. Curvature	10
2.2.2.5. Plan curvature.....	10
2.2.2.6. Profile curvature	10
2.2.2.7. Distance to roads	10
2.2.2.9. NDVI.....	11
2.2.2.10. Soil	11
2.2.2.11. Rainfall	11
2.2.2.12. Land Use / Land Cover (LULC)	12
2.2.2.13. STI.....	12
2.2.2.14. SPI.....	12
2.2.2.15. TRI.....	12
2.2.2.16. TWI	13
2.3. Preparation of the dataset and establish of the model	13
2.3.1. Detecting the multicollinearity with VIF and Tolerance values.....	14
2.3.2. Discretization transforms of the continuous features for the machine learning	15
2.3.3. Training and test sets of the study	16
2.4. Machine learning models	16
2.4.1. Support vector machine (SVM).....	16

2.4.2. Logistic regression (LR).....	17
2.4.3. Random forest (RF).....	18
3. Results of the application and discussion	18
3.1. Flood susceptibility map produced using SVM	18
3.2. Flood susceptibility map produced using LR.....	20
3.3. Flood susceptibility map produced using RF	23
3.4. Validation of the flood susceptibility maps.....	25
4. Conclusion.....	27
5. Acknowledgement	28
6. References	28

Abstract

Flood is one of the most devastating natural catastrophes that occur and cause loss of life and property in Hungary. Therefore, the main goal of this research was to generate flood susceptibility maps and comparison of them in Hungary using three different models, namely support vector machine (SVM), logistic regression (LR), and random forest (RF). In the present study, 385 locations were assigned as flood sample points {0}, and 385 locations were assigned as non-flooding points {1} in the Google Earth program for modelling, and it separated into two parts, %70 and %30 as training and test sets, respectively. Furthermore, a total of 16 flood influencing factors were generated in the ArcGIS environment, including distance to roads, aspect, curvature, stream power index (SPI), elevation, distance to rivers, land-use/cover (LULC), normalized difference vegetation index (NDVI), rainfall, terrain roughness index (TRI), slope, profile curvature, soil type, topographic wetness index (TWI), plan curvature, and sediment transport index (STI) for use in the modelling.

The receiver operating characteristic (ROC) curve and area under curve (AUC) were utilized to assess the three flood susceptibility map's model performance. According to ROC curve validation, AUC rate values in the modelling, for SVM, RF, and LR, were 0.993 (99.3%), 0.999 (99.9%), and 0.993 (99.3%), respectively. Thus, based on these values, the RF approach was the most accurate way to produce flood susceptibility maps in Hungary, as it had the greatest area under the curve and the highest value of accuracy (0.999) in the modelling. Moreover, since SVM and LR methods show values very close to RF accuracy, they can also be accepted and used in the studies.

As a result, the flood susceptibility maps created with three different machine learning techniques used in this study can help the relevant institutions in their flood planning studies and flood control in order to reduce the flood effects that may occur in the future.

Keywords: Geographic information system (GIS), Machine Learning, Flood susceptibility map, Hungary, Support Vector Machine, Random Forest, Logistic Regression

1. Introduction

Flood disasters that have occurred from the past to the present have caused hundreds and thousands of losses of life and property and have affected Hungary and all Central European Countries on a large scale (Alfred & Uwe, 2003). Research and studies predict that the destructive effects of flood events in Europe may increase due to the effects of global warming and affect human life in a chaotic manner and fatality (Alfieri et al., 2015; Forzieri et al., 2016; Jochen et al., 2014; Jongman et al., 2014; Paprotny & Morales-Nápoles, 2017; Vousdoukas et al., 2017; Winsemius et al., 2016).

According to Statista Database Company (<https://www.statista.com/statistics/> accessed date: Thursday, 31 March 2022), floods account for the largest share in the types of natural disasters that occurred in Europe between 2001 and 2020, with 41 percent., which makes the flood disaster is the highest figure for any disaster type in Europe. In addition, flood disasters were the most common natural disaster worldwide (201 events), while hurricanes affect the most people (45.5 million) and cause the most economic losses (\$92.7 billion). Extreme temperatures were the most deadly type of disaster, accounting for 42% of all deaths, followed by floods, accounting for 1% of deaths in 2021 (CRED & UNDRR, 2021). As we can see from the map of river flood discharges in Europe between 1960-2010 that shown in (Blöschl et al., 2019), Hungary is one of the most flood-prone countries in Europe, and we can support this statement by looking at the highest ever observed Danube water levels and the flood defense that cost €58 million in 2013 (International Commission for the Protection of the Danube River(ICPDR), 2013). Therefore, flood risk assessment and prediction are major challenges in developing and developed cities. Rapid population growth, rapid urbanization, human factors, and climate change will increase the significance of this challenge (Huong & Pathirana, 2013). Climate changes combined with human factors, in the escalation of severe natural disasters, in terms of the effects of flood disasters on social life during and after the flood, financial damage, loss of life, and damage to the environment play an important role. For this reason, appropriate monitoring to identify areas prone to floods is considered crucial to mitigate risks and losses by preventing unexpected flooding (Arshad et al., 2019; Huong & Pathirana, 2013; Quesada-Román et al., 2020, 2022; Said et al., 2019; Yu et al., 2018).

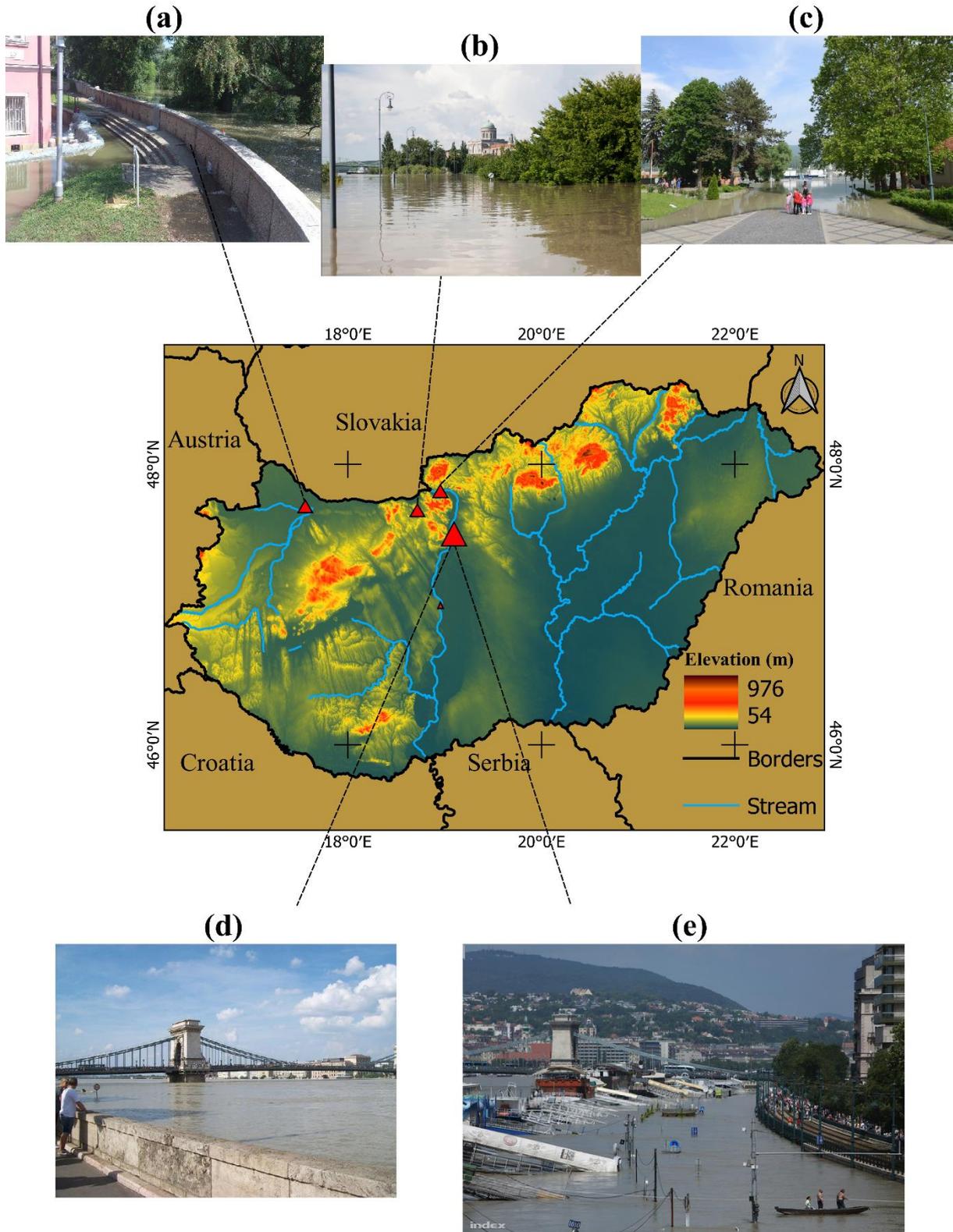


Fig.1. Selected location of inundated areas in the 2013 flood on the elevation map, (a) Győr, (b) Esztergom, (c) Nagymaros, (d) and (e) Budapest.

As seen in (Fig.1), extreme floods in the Hungarian portion of the Tisza River basin and Danube river basin have affected many towns, social and economic life along with the natural life (Albright, 2011). For example, the last flood happened in several locations in Hungary,

including the capital city Budapest. According to Directorate General for Disaster Management ([BM Országos Katasztrófavédelmi Főigazgatóság](#)), authorized teams performed more than 400 operations; however, a woman died in a flood disaster in Tótvázsony, Veszprém County after being dragged by the floodwaters. In addition to this fatal outcome, firefighters rescued about 14 people from flooded houses in the Fejér district and about 30 houses flooded in Szücsi, Heves District. Besides nature, it also economically damaged human-made structures such as buildings, roads, and power lines. (<https://floodlist.com/europe/hungary-czech-storms-floods-june-2020> accessed date: Thursday 31 March 2022). Moreover, we can understand how necessary the flood risk mapping is for the Hungarian sample study area, according to the lessons to be drawn from “The Great River Flood of Pest(1938)”, “The Great Flood of Szeged” that occurred in the past of Hungary and also by looking at major flooding events occurred in the Danube River Basin and Tisza River basin in the recent past in 2002, 2005, 2006, 2009, 2010, 2013 and 2014 years in Hungary. (<https://hungarytoday.hu/great-flood-pest-1838/> accessed date: Saturday 2 April 2022; <https://www.icpdr.org/main/danube-basin/hungary> accessed date: Wednesday 6 April 2022).

Many different flood forecasting methods and techniques have been used in the literature. As researchers explain; (Pappenberger et al., 2006; Sarhadi et al., 2012), flood forecasting is complicated and complex that is difficult to do, and precise prediction of flood occurrence, both spatially and temporally, is a challenging task. In the previous studies, one of the most prevalent flood-prone areas and flood hazard areas analysis is the hydraulic models, which give information regarding the inundated area, water level, and water flow velocity based on one-dimensional or two-dimensional hydrologic models. (Gharbi et al., 2016; M. et al., 2014). Another common technique used in flood susceptibility maps is the multi-criteria decision analysis (MCDA) (Elsheikh et al., 2015; Gudiyangada Nachappa et al., 2020; Papaioannou et al., 2015). In the literature, along with the multi-criteria decision making, flood susceptibility maps created using the MHDM technique and the analytical hierarchy process(AHP) technique have been used as well in studies (Souissi et al., 2020).

Although there are many Geographic Information Systems based flood susceptibility maps in the literature, support vector machines (SVM) (Choubin et al., 2019; Tehrany et al., 2014; Tehrany, Pradhan, Mansor, et al., 2015), decision trees (Khosravi et al., 2018; Merz et al., 2013; Tehrany et al., 2013; Tingsanchali & Karim, 2010; Yariyan et al., 2020), frequency ratio (FR) (Lee et al., 2012; Samanta et al., 2018; Shafapour Tehrany et al., 2019), Logistic Regression (Al-Juaidi et al., 2018; Shafapour Tehrany et al., 2019; Tehrany et al., 2013) and

Artificial neural networks (Kia et al., 2012; Shafizadeh-Moghadam et al., 2018) are the most common and effective techniques used in flood mapping. In flood susceptibility mapping, the accuracy of the forecast, as well as the flood forecast, prediction, and model, is of great importance. Thus, in this study, three machine learning models were implemented to create a flood susceptibility map by using satellite images and GIS tools. These techniques are support vector machine(SVM), logistic regression(LR), and random forest(RF). Furthermore, the accuracy and contribution of machine learning methods to flood, prone areas to flood, and the results of flood susceptibility mapping were examined.

2. Materials and method

2.1. Study area

The main aim of this study is to obtain flood susceptibility maps in Hungary using three models: Support Vector Machines (SVM), logistic regression (LR), and Random Forest (FR). Hungary is a landlocked country located in the Carpathian Basin, lies between longitudes 22°55' E and 16°8' E and latitudes 48°35' N and 45°45' N in Central Europe with seven neighbors: Slovakia in the north, Ukraine in the northeast, Romania in the east and southeast, Serbia in the south, Croatia and Slovenia in the southwest and Austria in the West (Fig. 2). The country is mostly flat and in the north has mountlets, with the highest mountain of Hungary's Kékes at 1,014 meters above sea level. The country spans about 93,030 square kilometers of the Carpathian Basin, divided between land areas and water sources. The Land covers approximately 34,598 square miles, while water covers the remaining 1,320 square miles of Hungary's total size. There are two main and longest rivers which are called Danube and Tisza divide the country into the three parts; Transdanubia(the west side of the Danube River), the plain part of the country between the Danube River and Tisza river, and the Trans-Tisza region(the east side of the Tisza River) (Masek, 2018).

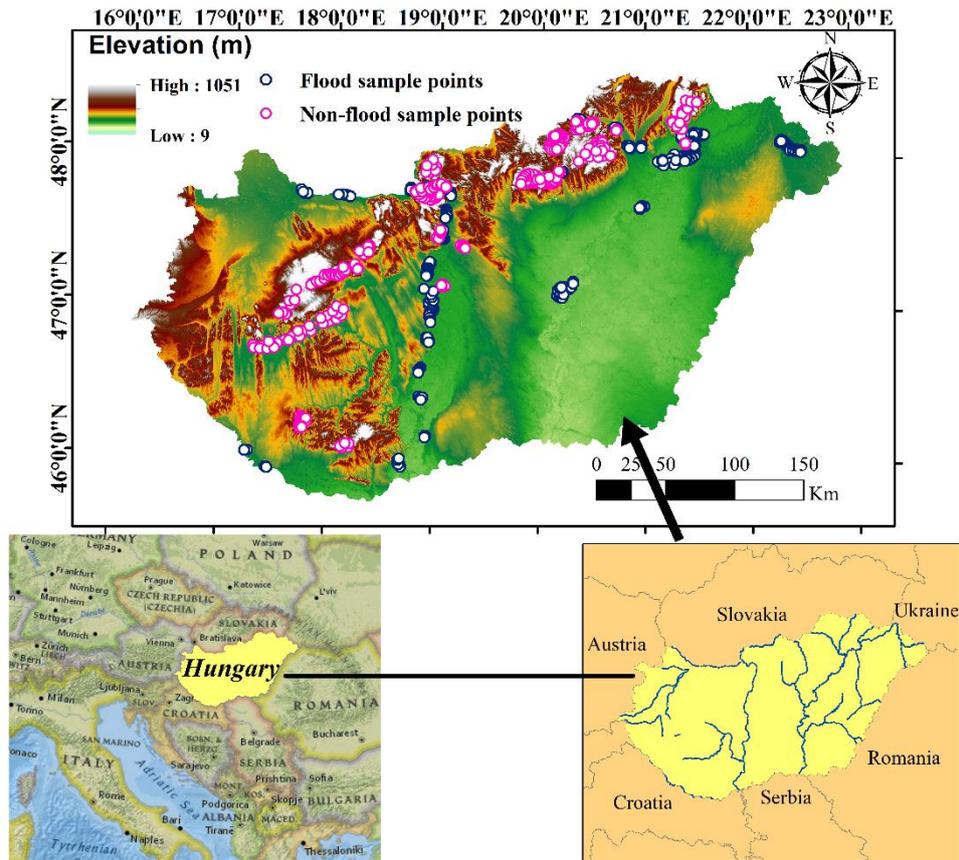


Fig. 2. The study area showing the distribution of flood and non-flood sample points locations

2.2. Data used

2.2.1. Preparation of the flood sample points

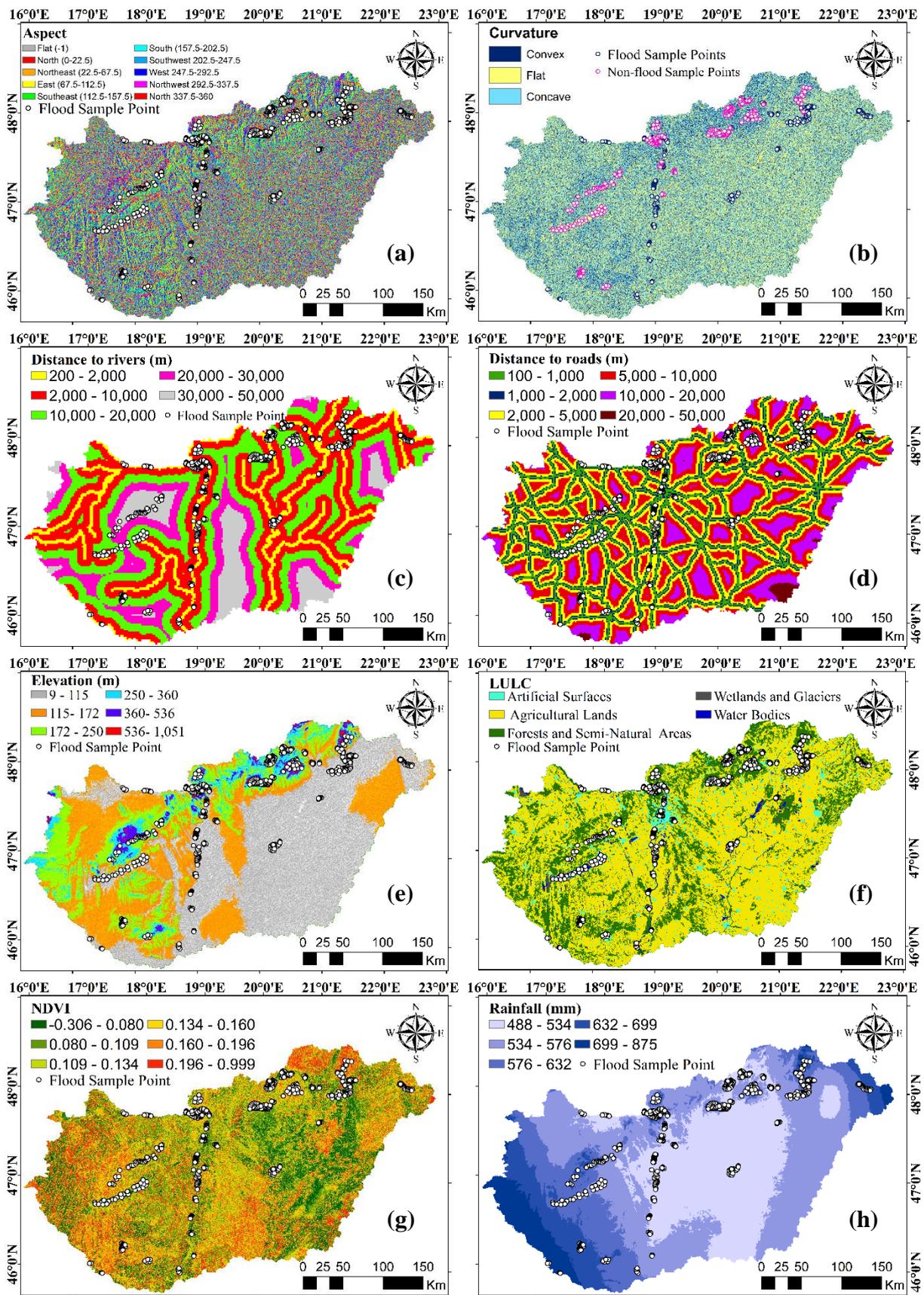
Generation of the flood inventory map is usually the initial and most crucial step to be applied in the flood susceptibility mapping (Masood & Takeuchi, 2012). In order to predict eventual flood events in any geographical area, it is one of the most critical points to analyze the flood records that have occurred in the past (Tehrany & Kumar, 2018). The flood inventory map, which shows the locations of previous floods, is required for the flood susceptibility modeling (Manandhar, 2010). It displays past flood records in a particular place. The flood sample points database, which is generated based on the past flood events, is both an important layer for the analysis of flood events and an essential factor in predicting the future flood events by associating them with the flood affecting factors (3.3.2). In the present study, a flood sample points map was generated by examining and using the existing historical flood data and flood reports of The International Commission for the Protection of the Danube River, The General Directorate of Water Management Hungary, local and global journal, and documentary sources.

Three hundred eighty-five(385) points were assigned as flood sample points and, on an equal number with 385 location points, were considered non-flood points.

The flood inventory map consists of 0 and 1 values, and values of 0 represent flood points, while values of 1 represent non-flood points. Non-flood points are randomly selected from mountainous areas with high points where there is no possibility of flooding. The flood inventory map is randomly divided into two parts, as %70 training dataset and %30 test dataset. In this case, 270 points were selected randomly and considered a dependent factor. Remained that %30 flood and non-flood locations, corresponding to 115 flood points, were used for testing.

2.2.2. Generation of the factors affecting the floods in the GIS-based software

According to the studies, there is no strictly conditioned and mandatory flood influencing factors list. Flood influencing factors were prepared elaborately by choosing among the factors selected and commonly used in the majority according to the studies in the literature (Dodangeh et al., 2020; Janizadeh et al., 2019; Khosravi et al., 2019; Shafizadeh-Moghadam et al., 2018). Therefore, 16 factors were used in this study to obtain the susceptibility maps. These flood influencing factors are distance to roads, aspect, curvature, stream power index (SPI), elevation, distance to rivers, land-use/cover (LULC), normalized difference vegetation index (NDVI), rainfall, terrain roughness index (TRI), slope, profile curvature, soil type, topographic wetness index (TWI), plan curvature, and sediment transport index (STI), respectively, as shown in (Fig. 3).



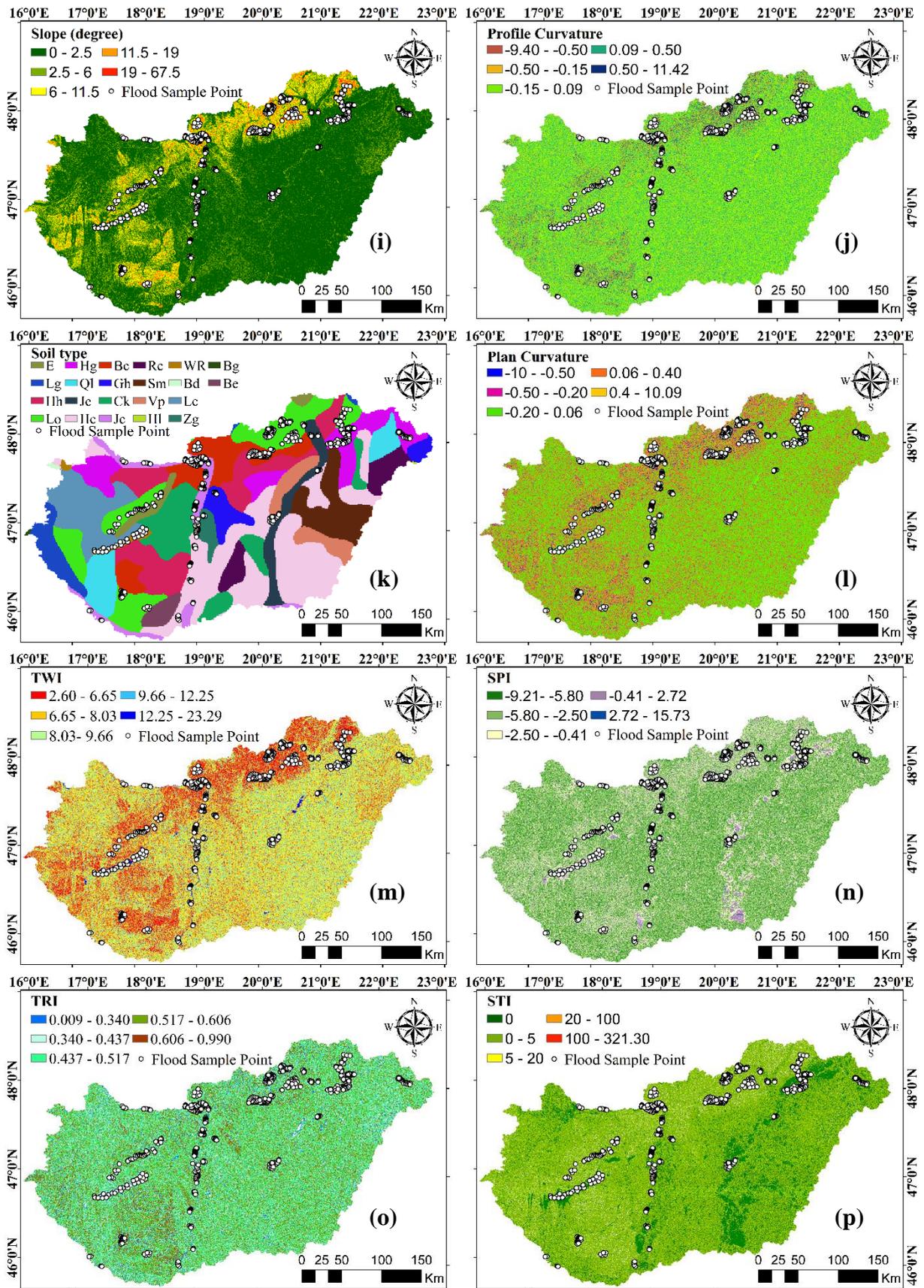


Fig. 3. Thematic representation of the flood influencing factors used in this study.

2.2.2.1. Elevation

The Elevation is also one of the significant and influential flood affecting factors involved in the flooding of an area. There is an opposite relationship between the altitude of a territory and flooding. The places with lower altitudes are more inclined to flooding (Choubin et al., 2019). The main reason for this is the rainwater flowing from the mountains at high levels, the snow, and glacial waters that melt and flow towards the spring and summer months. The elevation map was generated using the tools in the ArcGIS 10.3 interface of the arc-1 images downloaded from USGS. The digital elevation map (DEM) was generated using the shuttle radar topography mission (SRTM) 1 arc-second global images downloaded from [the United States Geological Survey \(USGS\)](#) web page using the tools in the ArcGIS 10.3 interface (Fig. 3e).

2.2.2.2. Aspect

The contribution of the aspect in previous works and studies on flood susceptibility maps demonstrates the importance of the aspect due to its widespread use as a factor influencing floods (Shafizadeh-Moghadam et al., 2018). Sunlit slopes are less humid and, therefore less prone to flooding. Slope aspect also has a significant influence on moisture content in the soil as well as meteorological conditions (Rahmati et al., 2016).). In the present study, the aspect map was generated from the digital elevation model (DEM) by using the tools in ArcGIS 10.3 and the aspect layer was divided into ten classes as represented in (Fig. 3a).

2.2.2.3. Slope

The wide distribution of low slope areas in the basin significantly affects the flood potential in the region (Khosravi, Pourghasemi, et al., 2016). There is an inverse proportionality between the flatness of the slope and the dewatering of the flooded area. Dewatering becomes more difficult as the slope decreases and the region flattens. In this case, considering the Great Hungarian Plain, which occupies the majority of the modern territory of Hungary, mainly in the eastern and southeastern areas of Hungary, it is clear that it will be difficult to clear this area from the water. The flowage volume and velocity are highly affected by the slope, which highlights the necessity of the slope. (Tien Bui et al., 2018) stated that the volume and velocity of runoff increase as the slope gradient increases. In the present study, the slope map was generated from the digital elevation model (DEM) by using the tools in ArcGIS 10.3, and the slope layer was classified into five categories as represented in (Fig. 3i). The slope angle in this study ranges from 0 to 67.5° degrees.

2.2.2.4. *Curvature*

The curvature represents the ground surface shape, such as flat, convex, and concave areas, which gives useful information that flat areas are more prone to flooding, as noted earlier on the slope (Khosravi et al., 2019; Tehrany, Pradhan, & Jebur, 2015). The curvature map is also generated from the digital elevation model (DEM) and consists of 3 classes as shown in (Fig. 3b). In these classes, negative(-) values represent convex, positive(+) values represent concave, and values corresponding to 0 represent flat (Youssef et al., 2015).

2.2.2.5. *Plan curvature*

The plan curvature (Fig. 3l) gives information about the convergence and divergence of the stream. Therefore, its used in the flood susceptibility mapping. Its derived from the digital elevation model (DEM).

2.2.2.6. *Profile curvature*

The profile curvature (Fig. 3j) raster were selected as on of the flood factor due to its effect on the flow rate and deceleration. Its derived from the digital elevation model (DEM).

2.2.2.7. *Distance to roads*

Since the distance from the road is one of the factors that can affect and prevent flood events, it is used as one of the factors affecting the flood in this study (Mukherjee & Singh, 2020; Shafapour Tehrany et al., 2017) . Wide and impermeable to water roads are the most suitable grounds for the transportation and accumulation of water in large-scale floods or continuous rainfall. For this reason, the distance from the road was chosen as a flood influencing factor. Distance to roads map was derived using the geographic information system software QGIS QuickOSM tool with the extracted Highway-Motorway and Highway-Primary vector data obtained from the [Open Street Map\(OSM\)](#). The distance to roads vector shapefile format was converted to raster format with appropriate resolution using ArcGIS 10.3 tool and divided into six classes as shown in (Fig. 3d).

2.2.2.8. *Distance to rivers*

Distance to rivers map was derived using geographic information system(ArcGIS) tools with the water course data(gdb) obtained from the [General Directorate of Water Management](#), technical spatial data service department.

2.2.2.9. NDVI

As its name signifies, Normalized Difference Vegetation Index (NDVI) is a factor that helps us understand the vitality of trees and vegetation cover on a particular region. The NDVI map (Fig. 3g) was prepared by using the Landsat 8 OLI/TIRS Collection 2 Level-2 satellite images (Obtained from the USGS Earth Explorer) and is calculated using the equation (1) below:

$$\text{NDVI} = (\text{NIR}-\text{RED}) / (\text{NIR}+\text{RED}) \quad (1)$$

The NDVI value range varies from -1 to +1. NDVI values close to +1 represent healthier and denser vegetation cover (Arabameri & Pourghasemi, 2019). As the researcher explained in the study conducted here (Kumar & Acharya, 2016), NDVI is very effective in flood events due to the inverse ratio between flooding and vegetation density. For this reason, NDVI was used in the present study as one of the factors applied in machine learning.

2.2.2.10. Soil

Soil type can be associated with the concept of water absorption, which the term we use when explaining in the distance to roads. For instance, the type of soil also may affects how much water is transported during the flood, just as concrete roads do not leak water and are a factor in flood formation (Rahmati et al., 2016). In the research paper (Huang et al., 2013) stated that, the type of soil surface has an essential part in the infiltration of the water. In addition to the water leakage of the soil type, the amount of water accumulated in it is also effective in the formation of floods. Therefore, soil type map (Fig. 3k) used as one of the factor affecting the flood in the susceptibility mapping. The soil type map was extracted from the vector data which downloaded from the [Food and Agriculture Organization of The United Nations](#) (FAO) for the Hungary study area.

2.2.2.11. Rainfall

Rainfall is one of the primary factor affecting flooding used in almost all flood susceptibility mapping studies (Tehrany et al., 2019). In the present study, rainfall map was derived using the data obtained (2011-2020) from the [Climatic Research Unit \(University of East Anglia\) and NCAS](#) and calculated using the inverse distance weighting (IDW) method in ArcGIS 10.3 . In general, while the annual precipitation is more in the southwestern and mountainous parts of the country, it decreases towards the middle flat parts and receives around 500 mm of annual precipitation.

Therefore, the annual precipitation amount is not proportionally distributed among the regions. While the spatial distribution of annual precipitation is concentrated in the southwest of the country, the great wide plain region received relatively less precipitation (Fig. 3h). The rainfall map is divided into 5 classes and the value range varies between 488 mm to 875 mm.

2.2.2.12. Land Use / Land Cover (LULC)

Land use is well up on the list as one of the factors affecting flood, due to widely used in previous comprehensive studies. The widespread use of land use not only in machine learning-based flood susceptibility mapping but also in other flood mapping methods and in another hazard assesment explains its magnitude (Khosravi, Nohani, et al., 2016; Roccati et al., 2021; Tehrani & Kumar, 2018). The land use map (Fig. 3f) was extracted from the data obtained via [The Copernicus Land Monitoring Service \(CLC 2018\)](#) and its classified into five categories.

2.2.2.13. STI

Sediment transportation index (STI) layer (Fig. 3p) were generated from digital elevation model (DEM) using the tools in the ArcGIS 10.3 interface and calculated and prepared using the equation (2) below:

$$STI = (A_s / 22.13)^m \times \sin(B / 0.0896)^n \quad (2)$$

where m, is the constant value used as 0.6, A_s is the catchment portion of the region, B is the slope layer (classified into quantile categories, used in the present study Fig. 3i) and n is the constant value used as 1.3.

2.2.2.14. SPI

Stream power index (SPI) is an essential factor in the flood susceptibility mapping as it is a measure of flow erosion power and runoff density (Florinsky, 2017).Stream power index (SPI) layer (Fig. 3n) were generated from digital elevation model (DEM) using the tools in the ArcGIS10.3 interface and calculated and prepared using the equation (3) below:

$$SPI = \ln(A_s * \tan \beta) \quad (3)$$

where A_s , is the catchment area (flow accumulation at gridcell) and β is the slope (Fig. 3 i) layer in radians that used in the present study.

2.2.2.15. TRI

Terrain roughness index (TRI), provides information on the roughness ratio of a particular land surface (Riley et al., 1999). Considering the influence of the ruggedness index to the flooding, terrain roughness index have been chosen for the present study . TRI layer (Fig. 3o)

were generated from digital elevation model (DEM) using the neighborhood tool in the ArcGIS 10.3 interface and calculated and prepared using the equation (4) below:

$$TRI = (FS_{mean} - FS_{min}) / (FS_{max} - FS_{min}) \quad (4)$$

Where FS_{mean} , is the focal statistical mean, FS_{min} is the focal statistical minimum and FS_{max} is the focal statistical maximum of the DEM raster.

2.2.2.16. TWI

Topographic wetness index (TWI) is a factor that provides information about topographic control and is frequently used on hydrological events and in flood susceptibility mapping (Sørensen et al., 2006). Due to there is a direct correlation between TWI values and prone to flooding, is used commonly in the flood susceptibility mapping. Areas with higher TWI values are more inundated areas. (Chen & Yu, 2011). TWI layer (Fig. 3m) were derived from digital elevation model (DEM) using the tool in the ArcGIS 10.3 interface and calculated using the equation (5) below:

$$TWI = \ln(A_s / \tan\beta) \quad (5)$$

2.3. Preparation of the dataset and establish of the model

The flood influencing factors that used in the present study, machine learning algorithms (RF, SVM, LR), separation of the test and training sets, validation technique and summary of the methodology were represented in flow chart (Fig. 4) below. These methods were comprehensive explained in the next steps.

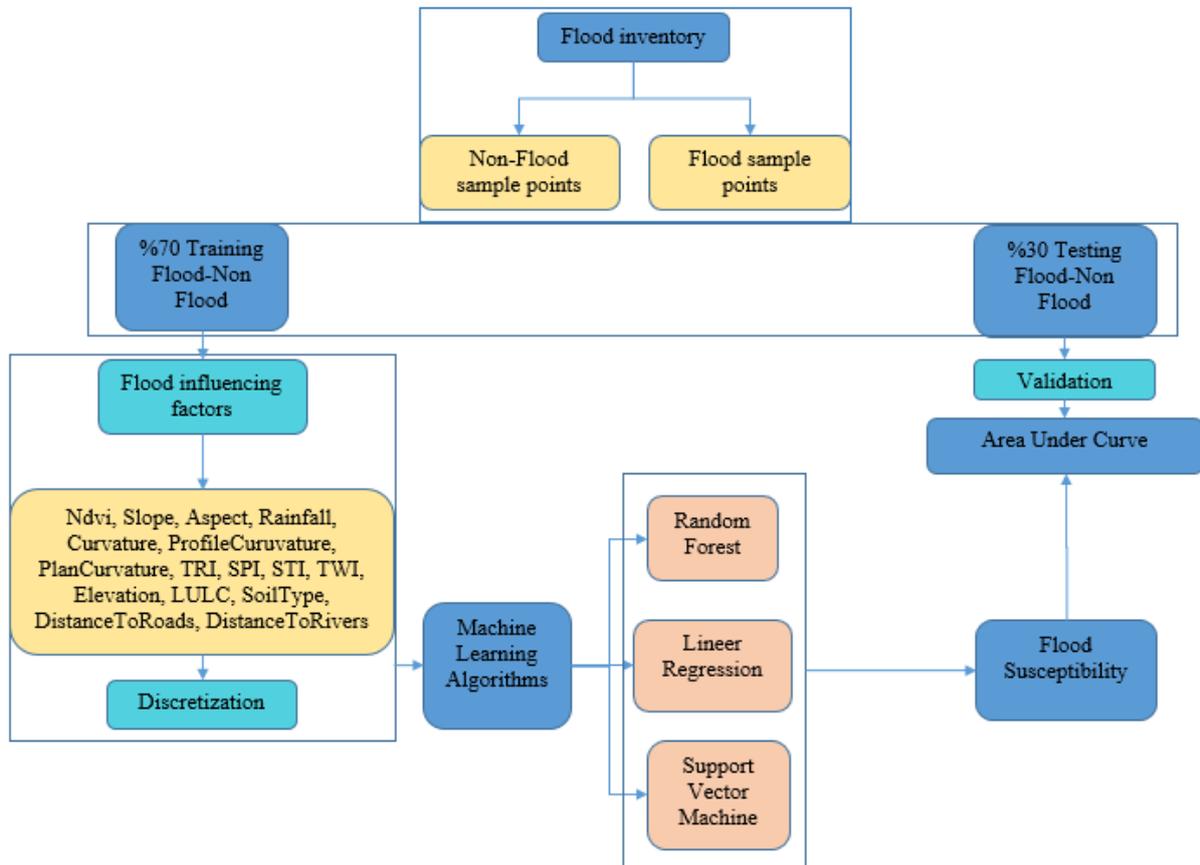


Fig. 4. Flowchart of the methodology

2.3.1. Detecting the multicollinearity with VIF and Tolerance values

In the machine learning, before starting the modeling, making the data to the convenient format ready for processing and organizing is one of the most important and first things to do. Therefore, the similarity and correlation values of the data used in this study were determined. Hence, a comprehensive multicollinearity test were applied to the data frame as a first step. The multicollinearity test is to detect if there is a high correlation among the two or more independent variables in a multiple regression model (Alin, 2010), which is among the data layers for our study. A linear relationship that may occur between the factors were used in the present study may lead to errors in the machine learning stages, to the confusion matrix, to the regression coefficients and to the all other parameters which used in the susceptibility mapping.

Thus, it will influence adversely the flood susceptibility map accuracy, which is the final product of the present study. However, multicollineratery can be eliminated by using the variance inflation factor (VIF) and tolerance values .The Tolerance values lower than 0.10 and VIF values above 10 indicate multicollinearity (Saha, 2017) and at this value range the similarity can be eliminate by removing the factors that indicate multicollinearity problems. By applying the multicollinearity test, VIF and tolerance values were calculated separately for each value in the python environment (Table 1).

(a)	VIF	Tolerance	(b)	VIF	Tolerance
Elevation	3.252073	0.307496	Elevation	3.251928	0.307510
Ndvi	2.150216	0.465070	Ndvi	2.149949	0.465127
Slope	2.257422	0.442983	Slope	2.256975	0.443071
Aspect	1.127767	0.886708	Aspect	1.124963	0.888918
Rainfall	1.521845	0.657097	Rainfall	1.521004	0.657461
Curvature	142.047526	0.007040	ProfileCurvature	1.406240	0.711116
ProfileCurvature	62.698990	0.015949	PlanCurvature	1.416149	0.706140
PlanCurvature	34.957595	0.028606	STI	1.081049	0.925028
STI	1.081449	0.924685	TWI	2.405980	0.415631
TWI	2.413781	0.414288	SPI	1.326078	0.754103
SPI	1.327497	0.753297	TRI	1.906638	0.524483
TRI	1.907059	0.524368	LULC	1.718042	0.582058
LULC	1.719023	0.581726	SoilType	1.596897	0.626215
SoilType	1.597311	0.626052	DistanceToRoads	1.445353	0.691873
DistanceToRoads	1.445509	0.691798	DistanceToRivers	1.777331	0.562641
DistanceToRivers	1.777551	0.562572			

Table 1. The multicollinearity analysis, VIF and Tolerance values results. The symbols (a) and (b) indicate that the multicollinearity test results before removing the curvature layer and after removing the curvature layer, respectively.

As (Table 1) indicates, the VIF value of each independent factor is less than 10 and the tolerance value is larger than 0.1 except the curvature factor (Gujarati et al., 2012; Kutner et al., 2004). The curvature layer shows linearity with the plan curvature and profile curvature factors. By removing the curvature layer from the test, the required linearity values were provided.

2.3.2. Discretization transforms of the continuous features for the machine learning

Before starting the model building process, we need to make the data set suitable and nominal for the analysis. Continuous data is converted into the discrete data in the python environment with the discretization method. After this stage is completed, the data set is divided

into two as testing and training. Since the statistically input variables are quite complex, discretization had been applied to the datasets. Flood factors that its used in the machine learning algorithms are discretized in certain standards, gives more effective performance results.

Discretization process were implemented by using the “cut” function which is in the pandas module. By using that function data layer values can be specify the according to the number of bins that we want to create based on the continious data and we can specify the labels for each bin according to the quantile classification.

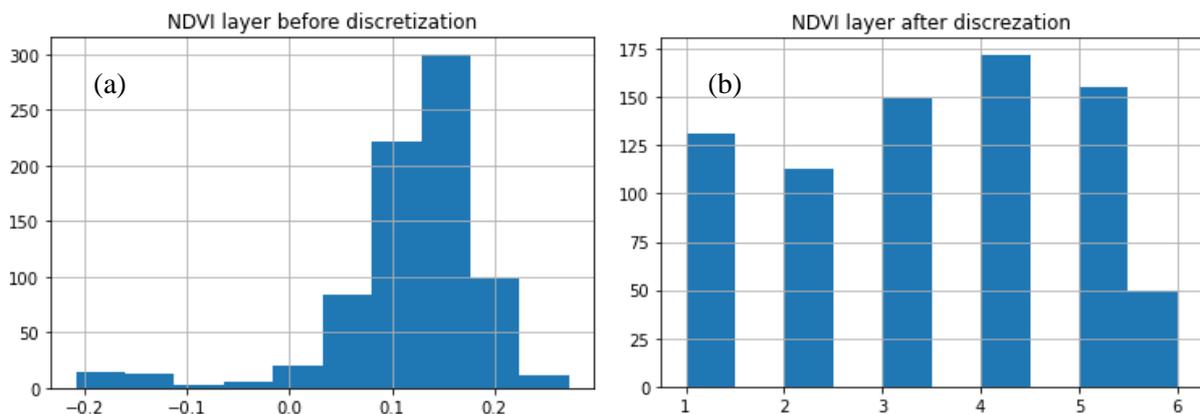


Fig. 5. Graphical representation of the selected NDVI factor; (a) before the discretization of the data, and (b) after the discretization of the data

The discretization process shown in (Fig. 5) were applied to all other flood factors separately. Thus, after the discretization process, machine learning applications became more adaptive and influential compared with the continuous data.

2.3.3. Training and test sets of the study

Separation of the model into training and test sets are the first priority in establishing the model and obtaining success and accuracy results. In the present study, model were separated into two parts, %70 and %30 (Gholamy et al., 2018) as training and test sets, respectively.

2.4. Machine learning models

As seen in the flowchart (Fig. 4), flood susceptibility maps studies were carried out with three machine learning methods in the present research; support vector machine (SVM), logistic regression (LR) and random forest (RF), respectively.

2.4.1. Support vector machine (SVM)

One of the most popular models used in flood susceptibility mapping is the support vector machine (SVM) (Gudiyangada Nachappa et al., 2020; Tehrany et al., 2019; Tehrany, Pradhan, Mansor, et al., 2015). Due to widespread and shared use in the susceptibility mappings

many times, support vector machine algorithms were selected in the present study to generate the flood susceptibility maps as one of the models. Support vector machines are supervised learning techniques that are used in classification (Jebur et al., 2014), regression and finding outliers, such as evaluated in the present study to estimate the {0} and {1} values for flood prediction. The method commonly used in the SVM is the kernel linear mathematical systems function together with it were employed for the data transformation and determination in the SVM model. The SVM, together with a collection of linear indicator functions, is a commonly used machine learning model that has been used to solve function determination problems (Vapnik, 1999). SVMs are based on the concept of determining the optimum hyperplane for dividing a dataset into two classes (Rustam et al., 2020). The kernel type and related parameters that are chosen have a direct influence on SVM productivity and precision.

2.4.2. Logistic regression (LR)

Logistic regression has been used to forecast the occurring probability of the natural disasters such as landslides and floods (Hosmer Jr et al., 2013). Logistic regression is comparable to multiple linear regression and enables one to make a connection and correlation for both flood event (dependent variable) and flood influencing factors (independent variable). Logistic regression is one of the most used statistical models in the literature to analyse and predict natural disasters such as landslides and floods (Yesilnacar & Topal, 2005). The goal of logistic regression in flood susceptibility study, which is stated by (Pradhan, 2010), is to find the best proper model to reflect the relationship between the dependent (flood inventory map) and independent (flood influencing factors) variables.

In the present modeling, the flood occurrence (dependent variable) consists of values {0} and {1}; where zero values represents the flood points, and one values represents non-flood points. The relationship between the dependent variable and other variables is expressed as follows (6):

$$p = 1 / (1 + e^{-z}) \quad (6)$$

Where p is the the flood probability index (Bai et al., 2012), explained as flood occurrence values varying between 0 and 1. In the equation (7) shown below, z is the linear regression. The logistic regression method tries adapting the data to an equation with the following explanation:

$$z = b_0 + b_1x_1 + b_2x_2+\dots \dots + b_nx_n \quad (7)$$

Where b_0 is a fixed value representing the method's intercept and the logistic regression coefficients are denoted by b_n , while the flood affecting factors (slope, rainfall, STI, TRI, etc.) are denoted by x_n .

2.4.3. Random forest (RF)

The Random Forest (RF) technique is an ensemble of the decision trees used to predict categorization or regression. RF has attracted increased interest in recent studies because of its potential to generate great classification and categorization results as well as its processing speed (Du et al., 2015). When predicting the product, a random set of characteristics is chosen at each stage, and each production is weighted by the value generated from the votes received. The majority vote, depending on the results of evaluated decision trees, leads to a single final categorization decision tree. In the flood susceptibility mapping, RF is one of the most important non-parametric ensemble learning approaches (Criminisi & Shotton, 2013; Ghorbanzadeh et al., 2019).

3. Results of the application and discussion

As a result, in the present study, 16 factors were analyzed, and their effects on flooding were tested. As mentioned earlier, these factors are namely elevation, aspect, slope, curvature, plan curvature, profile curvature, soil type, distance to roads, distance to rivers, NDVI, LULC, rainfall, STI, SPI, TRI and TWI. Among these factors, the curvature layer, which shows similarity according to the results of the multicollinearity test (Table 1), was excluded from the modelling. Hereupon, the flood influencing layers were classified with quantile classification in order to better analyze the flood susceptibility maps (Tehrany et al., 2019; Umar et al., 2014). The influence of these remaining 15 factors on the probability of flooding was tested using three different machine learning algorithms. These factors are support vector machines (SVM), logistic regression (LR), and random forest (RF).

3.1. Flood susceptibility map produced using SVM

Support vector machine (SVM) was implemented using the linear kernel parameter among four kernel types as LN, PL, RBF and SIG, and flood susceptibility maps were produced in the GIS environment by transferring the weight coefficients obtained from jupyter notebook to ArcGIS 10.3 interface. As Fig. 6 indicate, feature importance is ordered from largest to smallest according to weighting coefficient. Thus, according to SVM, the least influential parameters in Table 2 were profile curvature, plan curvature, and distance to roads, whereas the most influential ones were elevation, slope, and rainfall.

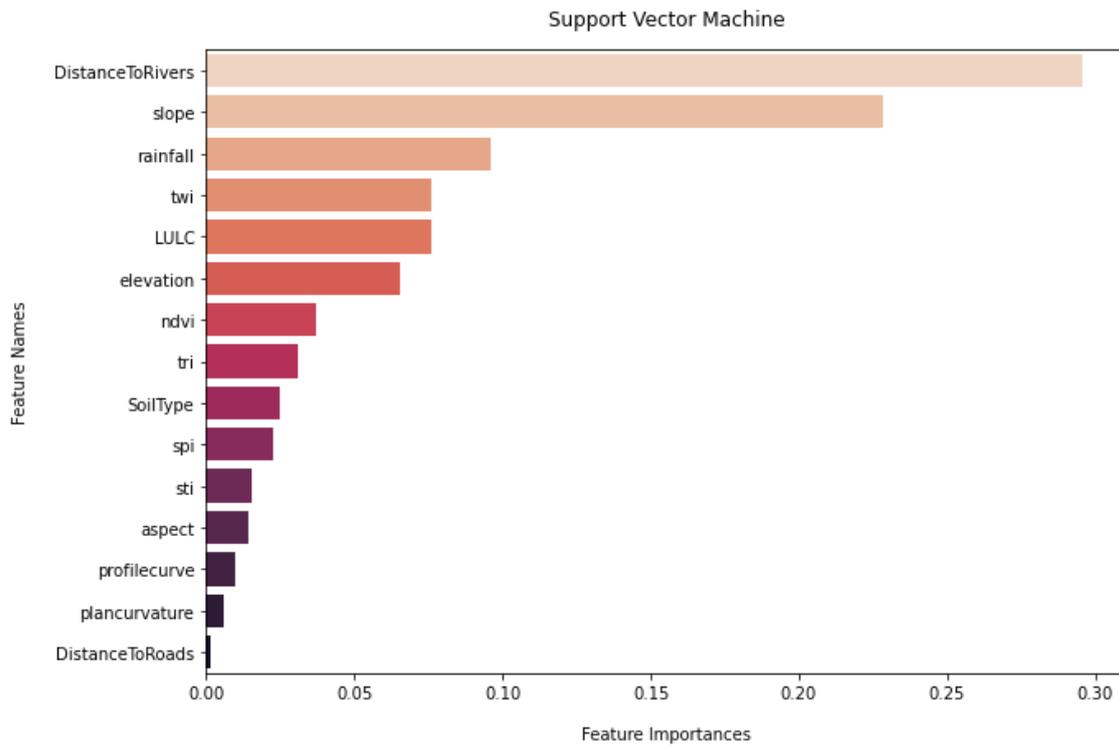


Fig. 6. Bar graph representation of the factors affecting flood based on SVM

	Flood influencing factors	Feature importances
14	DistanceToRivers	0.295599
2	slope	0.228614
4	rainfall	0.095927
8	twi	0.076212
11	LULC	0.076025
0	elevation	0.065463
1	ndvi	0.036954
10	tri	0.030885
12	SoilType	0.024768
9	spi	0.022766
7	sti	0.015598
3	aspect	0.014308
5	profilecurve	0.009617
6	plancurvature	0.005947
13	DistanceToRoads	0.001316

Table 2. Calculated weight coefficients based on SVM

The mathematical multiplication of the rescaled layers affecting the flood, and the weight coefficients calculated using the support vector machine, were made in the ArcGIS environment and shown in the following equation:

$$SVM = (0.295599 * DistanceToRivers) + (0.228614 * Slope) + (0.095927 * Rainfall) + (0.076212 * TWI) + (0.076025 * LULC) + (0.065463 *$$

$$\begin{aligned}
 & \text{Elevation}) + (0.036954 * NDVI) + (0.030885 * TRI) + (0.024768 * \\
 & \text{SoilType}) + (0.022766 * SPI) + (0.015598 * STI) + (0.014308 * \\
 & \text{Aspect}) + (0.009617 * ProfileCurvature) + (0.005947 * \\
 & \text{PlanCurvature}) + (0.001316 * DistanceToRoads)
 \end{aligned}
 \tag{8}$$

The final result SVM raster was derived from the equation above (8) by multiplying the each layer's raster images with the weight coefficients. After that, the generated raw flood susceptibility raster image was categorized into five vulnerable regions with the following area percentages for each class as shown in the Fig. 7; very high (%21.79), high (%28.92), moderate (%26.52), low (%16.60), and very low (%6.17).

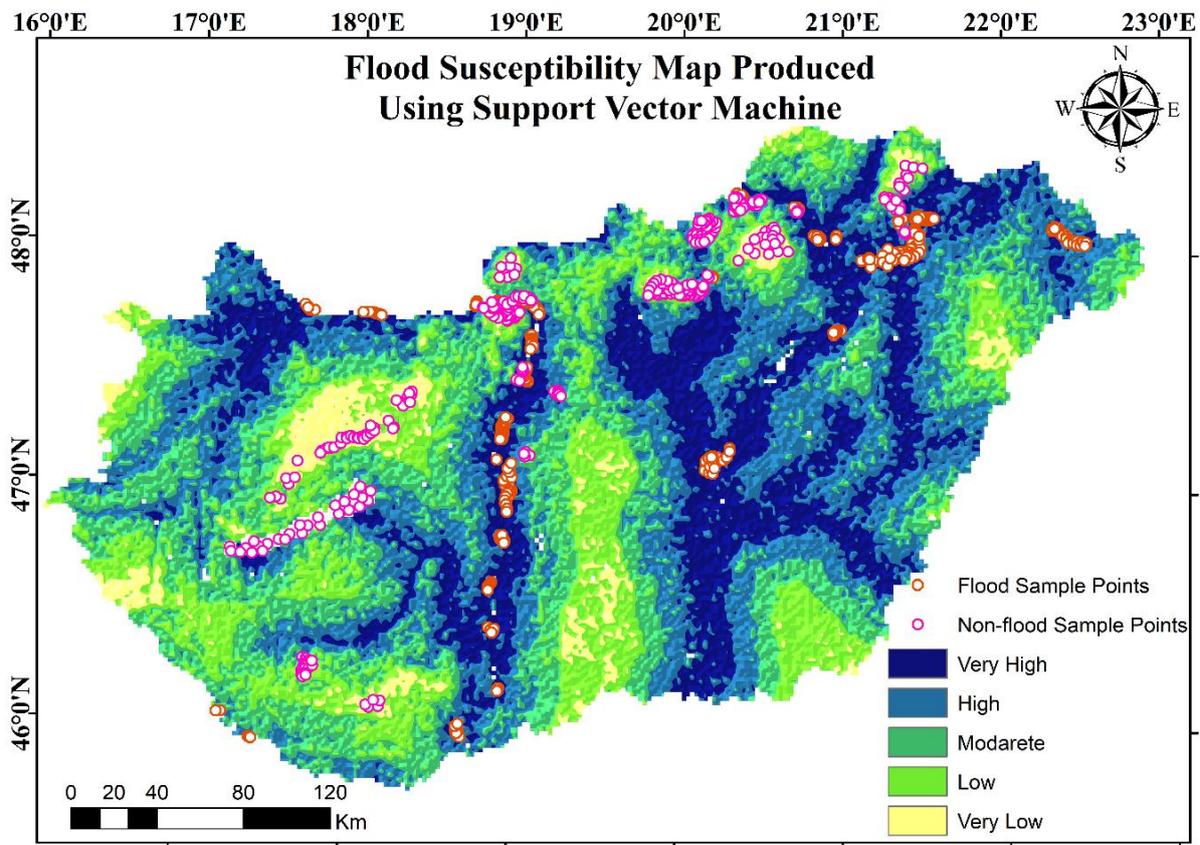


Fig. 7. Flood susceptibility map produced using SVM

3.2. Flood susceptibility map produced using LR

LR modelling was prepared using the Jupyter Notebook platform and LR coefficients (Table 3) were calculated separately for each flood influencing factor in the Jupyter environment.

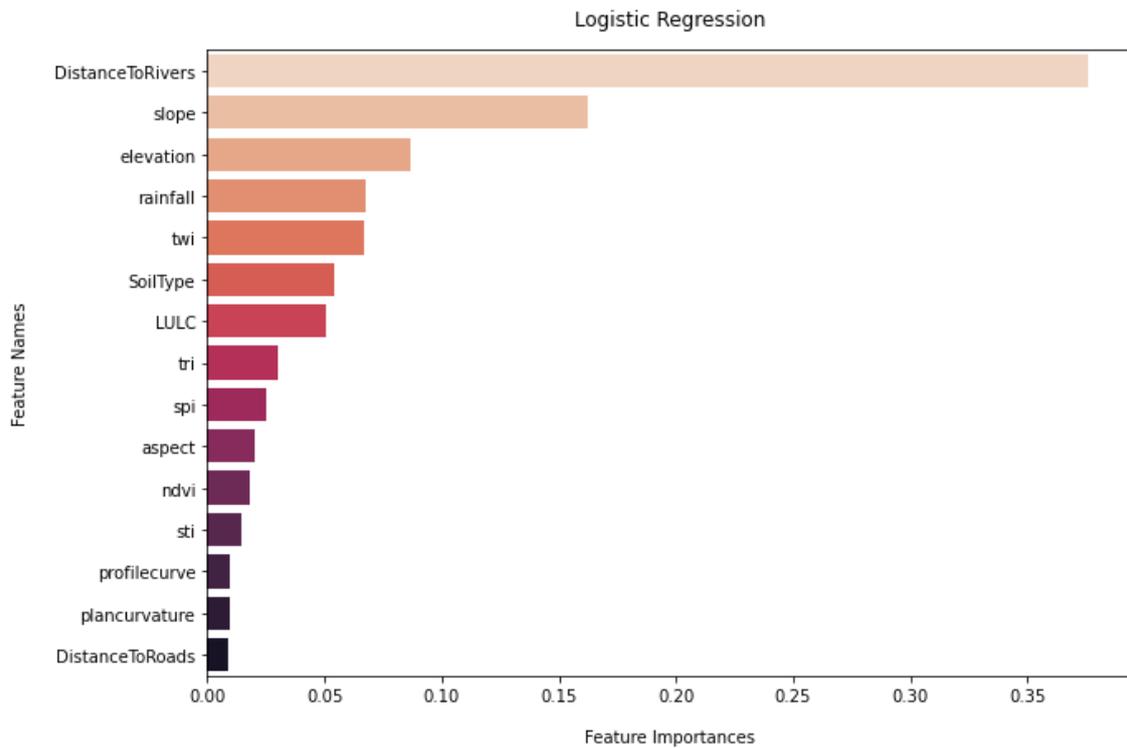


Fig. 8. Bar graph representation of the factors affecting flood based on LR

As Fig. 8 and Table 3 indicate, low LR weight values represent the factors that have a more negligible effect on flood occurrence, while the high-value weight coefficients represent the factors that have more influence on flood occurrence. Thus, in the present study, distance to roads, plan curvature, profile curvature and STI have the lowest weights, while distances to rivers, slope and elevation factors have the highest weights.

	Flood influencing factors	Feature importances
14	DistanceToRivers	0.375889
2	slope	0.162576
0	elevation	0.086731
4	rainfall	0.067276
8	twi	0.067229
12	SoilType	0.053899
11	LULC	0.050703
10	tri	0.030027
9	spi	0.025333
3	aspect	0.020439
1	ndvi	0.017969
7	sti	0.014564
5	profilecurve	0.009413
6	plancurvature	0.009384
13	DistanceToRoads	0.008568

Table 3. Weight coefficients based on LR

The mathematical multiplication of the rescaled layers affecting the flood, and the weight coefficients calculated using logistic regression model, were made in the ArcGIS environment and shown in the following equation:

$$\begin{aligned}
 LR = & (0.375889 * DistanceToRivers) + (0.162576 * Slope) + \\
 & (0.086731 * Elevation) + (0.067276 * Rainfall) + (0.067229 * TWI) + \\
 & (0.053899 * SoilType) + (0.050703 * LULC) + (0.030027 * TRI) + \\
 & (0.025333 * SPI) + (0.020439 * Aspect) + (0.017969 * NDVI) + \\
 & (0.014564 * STI) + (0.009413 * ProfileCurvature) + \\
 & (0.009384 * PlanCurvature) + (0.00856 * DistanceToRoads)
 \end{aligned}
 \tag{9}$$

The final result SVM raster was derived from the equation above (9) by multiplying the each layer's raster images with the weight coefficients. After that, the generated raw flood susceptibility raster image was categorized into five vulnerable regions with the following area percentages for each class as shown in the Fig. 9; very high (%21.16), high (%24.91), moderate (%25.62), low (%16.27), and very low (%12.04).

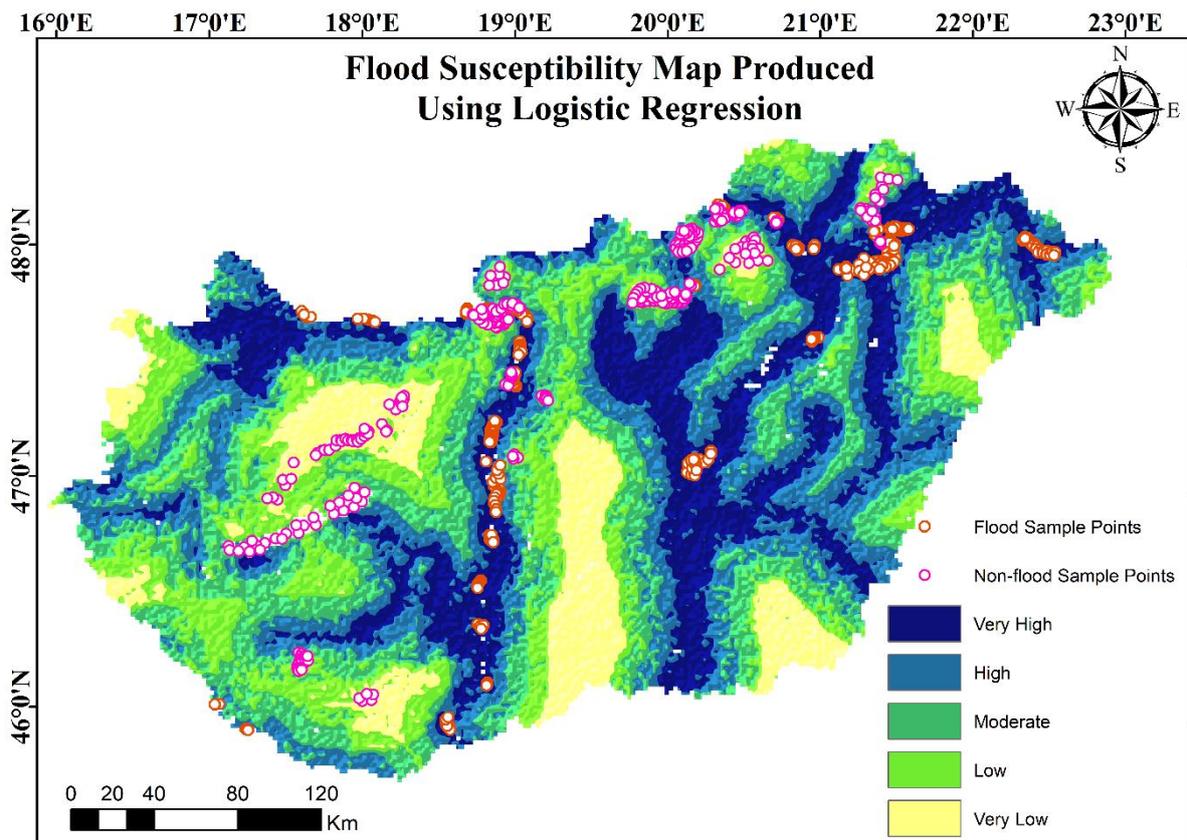


Fig. 9. Flood susceptibility map produced using LR

3.3. Flood susceptibility map produced using RF

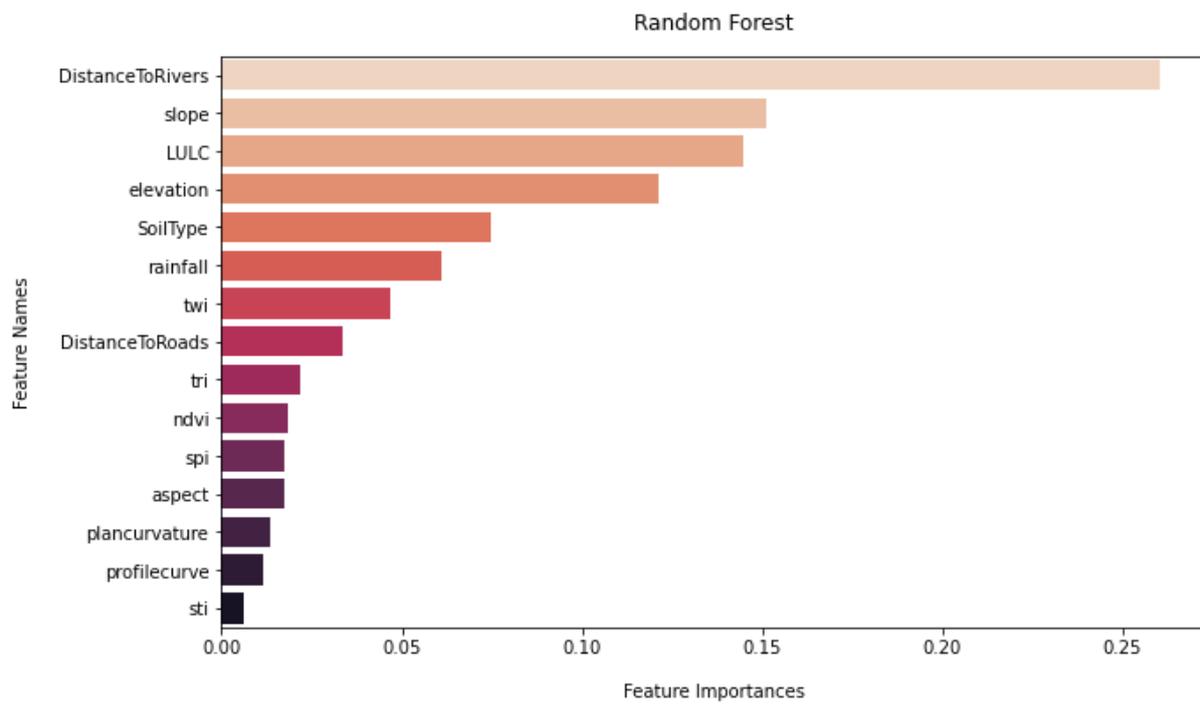


Fig. 10. Bar graph representation of the factors affecting flood based on RF

	Flood influencing factors	Feature importances
14	DistanceToRivers	0.260197
2	slope	0.150996
11	LULC	0.144734
0	elevation	0.121298
12	SoilType	0.074753
4	rainfall	0.060781
8	twi	0.046948
13	DistanceToRoads	0.033589
10	tri	0.021825
1	ndvi	0.018312
9	spi	0.017510
3	aspect	0.017504
6	plancurvature	0.013574
5	profilecurve	0.011742
7	sti	0.006238

Table 4. Weight coefficients based on RF

Based on the [Table 4](#) results, with a value of 0.260197, the distance to rivers is the highest and most essential factor affecting the flood, as in the other two models. The mathematical multiplication of the rescaled layers affecting the flood, and the weight coefficients calculated using random forest (RF), were made in the ArcGIS environment and shown in the following equation:

$$\begin{aligned}
RF = & (0.260197 * DistanceToRivers) + (0.150996 * Slope) + \\
& (0.144734 * LULC) + (0.121298 * Elevation) + (0.074753 * Soil) + \\
& (0.060781 * Rainfall) + (0.046948 * TWI) + \\
& (0.033589 * DistanceToRoads) + (0.021825 * TRI) + (0.018312 * NDVI) \quad (10) \\
& +(0.017510 * SPI) + (0.017504 * Aspect) + (0.013574 * PlanCurvature) \\
& +(0.011742 * ProfileCurvature) + (0.006238 * STI)
\end{aligned}$$

The final result RF raster was derived from the equation above (10) by multiplying the each layer's raster images with the weight coefficients. After that, the generated raw flood susceptibility raster image was categorized into five vulnerable regions with the following area percentages for each class as shown in the Fig. 11; very high (%16.85), high (%27.88), moderate (%26.46), low (%19.80), and very low (%9.01).

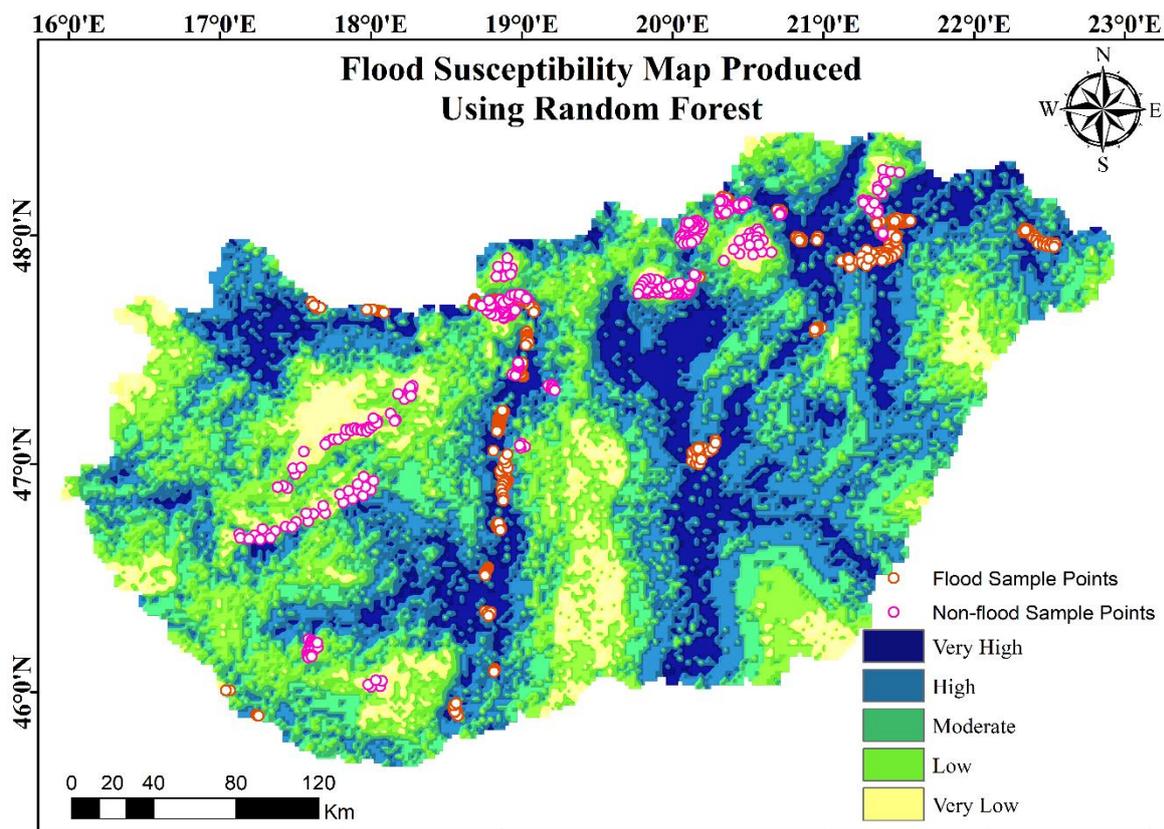


Fig. 11. Flood susceptibility map produced using RF

3.4. Validation of the flood susceptibility maps

In the machine learning, performance measurement is a critical issue. In the present study, a flood prediction ratio of performance appraisal values was produced using the training dataset to evaluate the performance and the testing dataset for validation for each of the three machine models. AUC is one of the most essential evaluation criteria for classification model performance used in the literature (Tien Bui et al., 2018). Therefore, as a result, the receiver operating characteristic (ROC) curve and area under curve (AUC), was utilized to assess the flood susceptibility map's performance.

	precision	recall	f1-score	support
0	0.97	0.95	0.96	97
1	0.96	0.98	0.97	134
Accuracy			0.97	231
macro avg	0.97	0.96	0.96	231
weighted avg	0.97	0.97	0.97	231

accuracy SVC: 0.9653679653679653
 precision SVC: 0.9632352941176471
 recall SVC : 0.9776119402985075
 f1 score SVC : 0.9703703703703703

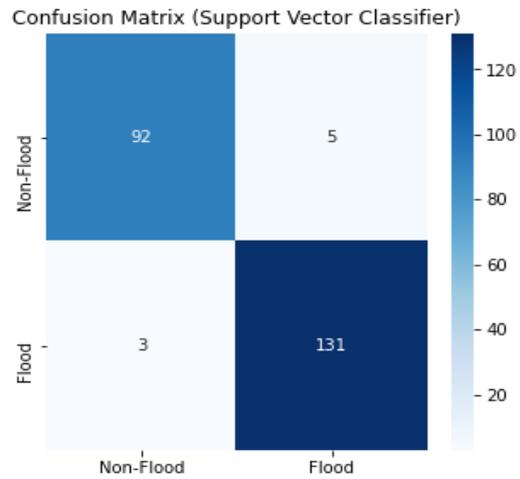


Fig. 12. SVM Confusion Matrix and other performance metrics

	precision	recall	f1-score	support
0	0.98	0.95	0.96	97
1	0.96	0.99	0.97	134
Accuracy			0.97	231
macro avg	0.97	0.97	0.97	231
weighted avg	0.97	0.97	0.97	231

accuracy LR: 0.9696969696969697
 precision LR: 0.9635036496350365
 recall LR : 0.9850746268656716
 f1 score LR : 0.9741697416974171

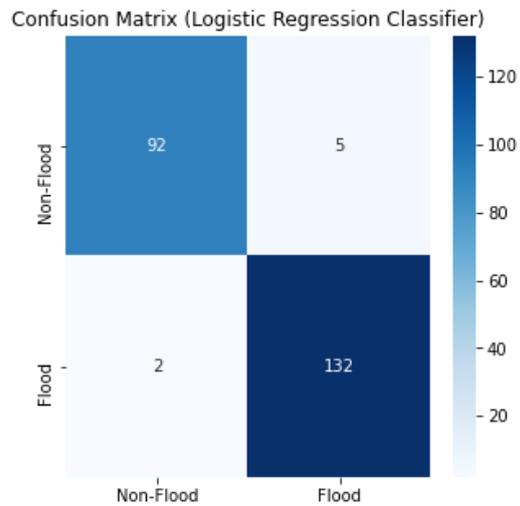


Fig. 13. LR Confusion Matrix and other performance metrics

	precision	recall	f1-score	support
0	0.98	1.00	0.99	97
1	1.00	0.99	0.99	134
Accuracy			0.99	231
macro avg	0.99	0.99	0.99	231
weighted avg	0.99	0.99	0.99	231

accuracy RFC: 0.9913419913419913
precision RFC: 1.0
recall RFC : 0.9850746268656716
f1 score RFC : 0.9924812030075187

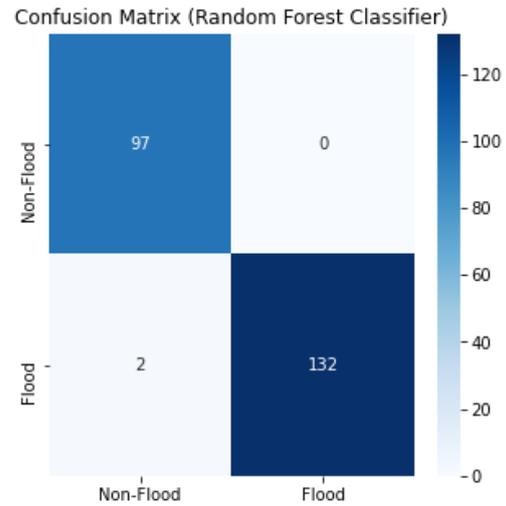


Fig. 14. RF Confusion Matrix and other performance metrics

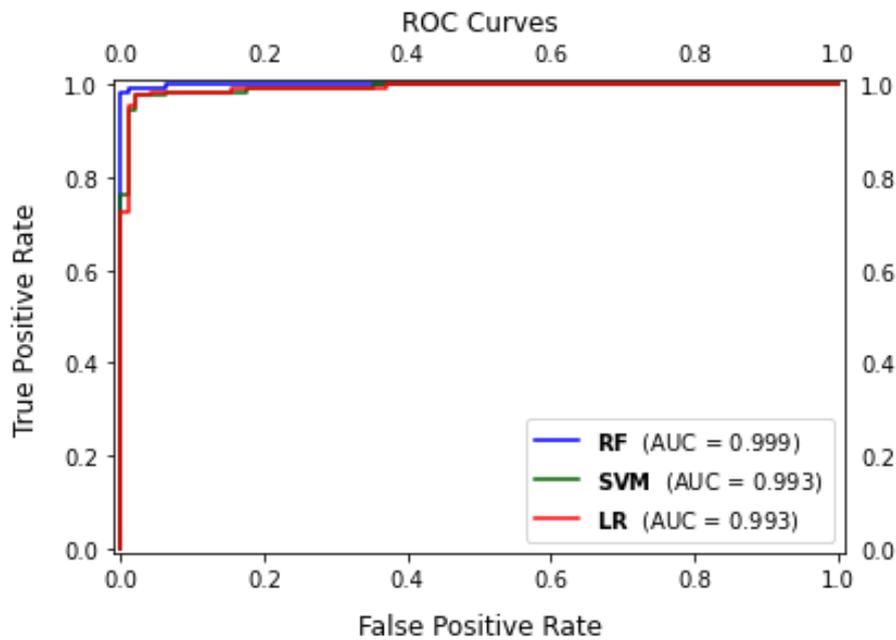


Fig. 15. ROC curve plotting of the SVM, LR and RF models

ROC is a probability curve, and AUC represents the degree or measure of separability. AUC represents the degree or measure of separability, whereas ROC is a probability curve (Davis & Goadrich, 2006). As indicated in Fig. 15, the ROC curve is depicted in two dimensions, with the false positive rates (specificity) on the x-axis and true positive rates (sensitivity) on the y-axis. ROC curves of the created integrative models during the learning and validation phases, with the red curve indicating the logistic regression (0.993) AUC value, the green curve representing the support vector machine(0.993) AUC value, and the blue curve representing the random forest (0.999) AUC value. Based on these findings, it can be stated that the RF approach was the most accurate way for producing flood susceptibility maps in Hungary, as it had the

greatest area under the curve and the highest value of accuracy (0.999) among the other two methods.

4. Conclusion

Flood is one of the most dangerous and destructive natural disasters globally. Therefore, the production of flood susceptibility maps plays an essential role in making provisions against flood hazards. 16 flood influencing factors, namely distance to roads, aspect, curvature, stream power index (SPI), elevation, distance to rivers, land-use/cover (LULC), normalized difference vegetation index (NDVI), rainfall, terrain roughness index (TRI), slope, profile curvature, soil type, topographic wetness index (TWI), plan curvature, and sediment transport index (STI), were used in the present study, along with three machine learning models, support vector machine (SVM), logistic regression (LR), and random forest (RF) for modelling and produce flood susceptibility maps in Hungary. Multicollinearity tests (VIF) and Tolerance methods were used to measure the relevance of these 16 Flood influence factors. However, the curvature layer was excluded from the study because it showed multicollinearity and the remaining 15 factors were used. The accuracy of the models and flood susceptibility maps based on the training and validation datasets was determined using the ROC curve and AUC. The application of the ROC Curve in the validation stage demonstrates that, in comparison to other techniques, the Random Forest model results have the best AUC performance with the 0.999. The other two models, SVM and LR, showed the same performance with an AUC value of 0.993.

In conclusion, the approach of all flood susceptibility mapping models was satisfactory and dependable. According to the SVM model, an overall area of 21.79 percent was considered severely prone to flooding. In addition, 28.92% of the areas were classified as high-risk, 26.52% of them as moderate-grade, 16.60% of them as low-grade, and 6.17% of them as very low flood susceptibility based on the SVM model results. For RF and LR, these numbers show the proportion of places prone to flooding, with a ratio of 16.85% extremely high, 27.88% high, 26.46% moderate, 19.80% low, 9.01% very low, and 21.16% very high, 24.91%, 25.62% moderate, 16.27% low, 12.04% very low, respectively.

Although, in the literature, there are previous studies for the flood mappings, such as a neural network approach in Hungary (Skakun, 2010), the research with complex and effective machine learning methods was implemented for the first time in the present study. In conclusion, strategies and flood susceptibility maps proposed here might be useful in disaster management, guiding emergency plans and to the relevant institutions and organizations.

5. Acknowledgement

I would first like to thank my supervisor, Lect. Varga Zsófia for the helpful comments on the dissertation work. And I also thank the Department of Cartography and Geoinformatics for the resources and knowledge they provided me. I want to thank the USGS, other web-enabling institutes and the General Directorate of Water Management for providing the water-course data I used in my study. Finally, I would like to thank my family and my girlfriend, who were always there for me during this process.

6. References

- Al-Juaidi, A. E. M., Nassar, A. M. & Al-Juaidi, O. E. M. (2018). Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences*, 11(24), 765. <https://doi.org/10.1007/s12517-018-4095-0>
- Alfieri, L., Burek, P., Feyen, L. & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, 19(5), 2247–2260. <https://doi.org/10.5194/hess-19-2247-2015>
- Alfred, B. & Uwe, G. (2003). Flood Risk in Central Europe. *Science*, 300(5622), 1099. <https://doi.org/10.1126/science.1083624>
- Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370–374.
- Arabameri, A. & Pourghasemi, H. R. (2019). 13 - Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R (H. R. Pourghasemi & C. B. T.-S. M. in G. I. S. and R. for E. and E. S. Gokceoglu (eds.); pp. 299–321). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-815226-3.00013-2>
- Arshad, B., Ogie, R., Barthelemy, J., Pradhan, B., Verstaevel, N. & Perez, P. (2019). Computer Vision and IoT-Based Sensors in Flood Monitoring and Mapping: A Systematic Review. In *Sensors* (Vol. 19, Issue 22). <https://doi.org/10.3390/s19225012>
- Bai, S., Wang, J., Zhang, Z. & Cheng, C. (2012). Combined landslide susceptibility mapping after Wenchuan earthquake at the Zhouqu segment in the Bailongjiang Basin, China. *Catena*, 99, 18–25.
- Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., ... Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, 573(7772), 108–111. <https://doi.org/10.1038/s41586-019-1495-6>
- Chen, C.-Y. & Yu, F.-C. (2011). Morphometric analysis of debris flows and their source areas using GIS. *Geomorphology*, 129(3), 387–397. <https://doi.org/https://doi.org/10.1016/j.geomorph.2011.03.002>
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. & Mosavi, A. (2019). An ensemble

- prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment*, 651, 2087–2096.
<https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.10.064>
- CRED & UNDRR. (2021). The Non-COVID Year in Disasters. *CRED*.
- Criminisi, A. & Shotton, J. (2013). *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media.
- Davis, J. & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240.
- 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. *Science of The Total Environment*, 705, 135983.
<https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.135983>
- Du, P., Samat, A., Waske, B., Liu, S. & Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 38–53.
- Elsheikh, R. F. A., Ouerghi, S. & Elhag, A. R. (2015). Flood Risk Map Based on GIS, and Multi Criteria Techniques (Case Study Terengganu Malaysia). In *Journal of Geographic Information System* (Vol. 07, Issue 04, pp. 348–357). <https://doi.org/10.4236/jgis.2015.74027>
- Florinsky, I. V. (2017). An illustrated introduction to general geomorphometry. *Progress in Physical Geography: Earth and Environment*, 41(6), 723–752. <https://doi.org/10.1177/0309133317733667>
- Forzieri, G., Feyen, L., Russo, S., Vousdoukas, M., Alfieri, L., Outten, S., Migliavacca, M., Bianchi, A., Rojas, R. & Cid, A. (2016). Multi-hazard assessment in Europe under climate change. *Climatic Change*, 137(1), 105–119. <https://doi.org/10.1007/s10584-016-1661-x>
- Gharbi, M., Soualmia, A., Dartus, D. & Masbernat, L. (2016). Comparison of 1D and 2D hydraulic models for floods simulation on the Medjerda Riverin Tunisia. *J. Mater. Environ. Sci*, 7(8), 3017–3026.
- Gholamy, A., Kreinovich, V. & Kosheleva, O. (2018). *Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation*.
- Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D. & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, 11(2), 196.
- Gudiyangada Nachappa, T., Tavakkoli Piralilou, S., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O. & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590, 125275.
<https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125275>

- Gujarati, D. N., Porter, D. C. & Gunasekar, S. (2012). *Basic econometrics*. Tata mcgraw-hill education.
- Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Huang, J., Wu, P. & Zhao, X. (2013). Effects of rainfall intensity, underlying surface and slope gradient on soil infiltration under simulated rainfall experiments. *CATENA*, *104*, 93–102.
<https://doi.org/https://doi.org/10.1016/j.catena.2012.10.013>
- Huong, H. T. L. & Pathirana, A. (2013). Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam. *Hydrology and Earth System Sciences*, *17*(1), 379–394.
<https://doi.org/10.5194/hess-17-379-2013>
- International Commission for the Protection of the Danube River(ICPDR). (2013). *Floods in June 2013 in the Danube River basin*. <https://www.icpdr.org/main/resources/floods-june-2013-drb-0>
- Janizadeh, S., Avand, M., Jaafari, A., Phong, T. Van, 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*(19). <https://doi.org/10.3390/su11195426>
- Jebur, M. N., Mohd Shafri, H. Z., Pradhan, B. & Tehrany, M. S. (2014). Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery. *Geocarto International*, *29*(7), 792–806.
- Jochen, H., Daniel, L., T., V. A., Mahé, P., James, N. R., J., T. R. S., Ben, M., Xavier, F., Cezar, I. & Anders, L. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences*, *111*(9), 3292–3297. <https://doi.org/10.1073/pnas.1222469111>
- Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., Bouwer, L. M., Pflug, G., Rojas, R. & Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, *4*(4), 264–268. <https://doi.org/10.1038/nclimate2124>
- 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. *Natural Hazards*, *83*(2), 947–987. <https://doi.org/10.1007/s11069-016-2357-2>
- Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I. & Tien Bui, D. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of The Total Environment*, *627*, 744–755.
<https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.01.266>
- Khosravi, K., Pourghasemi, H. R., Chapi, K. & Bahri, M. (2016). Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon’s entropy, statistical index, and weighting factor models. *Environmental Monitoring and Assessment*, *188*(12), 656.
<https://doi.org/10.1007/s10661-016-5665-9>

- Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H. -B., Gróf, G., Ho, H. L., Hong, H., Chapi, K. & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, 573, 311–323. <https://doi.org/10.1016/j.jhydrol.2019.03.073>
- Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A. & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, 67(1), 251–264. <https://doi.org/10.1007/s12665-011-1504-z>
- Kumar, R. & Acharya, P. (2016). Flood hazard and risk assessment of 2014 floods in Kashmir Valley: a space-based multisensor approach. *Natural Hazards*, 84(1), 437–464. <https://doi.org/10.1007/s11069-016-2428-4>
- Kutner, M. H., Nachtsheim, C. J., Neter, J. & Wasserman, W. (2004). *Applied linear regression models* (Vol. 4). McGraw-Hill/Irwin New York.
- Lee, M.-J., Kang, J. & Jeon, S. (2012). Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. *2012 IEEE International Geoscience and Remote Sensing Symposium*, 895–898. <https://doi.org/10.1109/IGARSS.2012.6351414>
- M., M., B., B., S., B., G., D. B., M., P. & R., R. (2014). Flooding Hazard Mapping in Floodplain Areas Affected by Piping Breaches in the Po River, Italy. *Journal of Hydrologic Engineering*, 19(4), 717–731. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000840](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000840)
- Manandhar, B. (2010). Flood plain analysis and risk assessment of Lothar Khola. *Master of Science Thesis in Watershed Management. Tribhuvan University Institute of Forestry Pokhara, Nepal.*
- Masek, Z. (2018). Settlement History of the Middle Tisza Region in the 4th–6th centuries AD. According to the Evaluation of the Material from Rákóczi-falva-Bagi-földek 5–8–8A sites. *DISSERTATIONES ARCHAEOLOGICAE EX INSTITUTO ARCHAEOLOGICO UNIVERSITATIS DE ROLANDO EÖTVÖS NOMINATAE*, 597–619.
- Masood, M. & Takeuchi, K. (2012). Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. *Natural Hazards*, 61(2), 757–770. <https://doi.org/10.1007/s11069-011-0060-x>
- Merz, B., Kreibich, H. & Lall, U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, 13(1), 53–64. <https://doi.org/10.5194/nhess-13-53-2013>
- Mukherjee, F. & Singh, D. (2020). Detecting flood prone areas in Harris County: a GIS based analysis. *GeoJournal*, 85(3), 647–663. <https://doi.org/10.1007/s10708-019-09984-2>
- Papaioannou, G., Vasiliades, L. & Loukas, A. (2015). Multi-Criteria Analysis Framework for Potential Flood Prone Areas Mapping. *Water Resources Management*, 29(2), 399–418. <https://doi.org/10.1007/s11269-014-0817-6>
- Pappenberger, F., Matgen, P., Beven, K. J., Henry, J.-B., Pfister, L. & Fraipont, P. (2006). Influence of uncertain

- boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources*, 29(10), 1430–1449. <https://doi.org/https://doi.org/10.1016/j.advwatres.2005.11.012>
- Paprotny, D. & Morales-Nápoles, O. (2017). Estimating extreme river discharges in Europe through a Bayesian network. *Hydrology and Earth System Sciences*, 21(6), 2615–2636. <https://doi.org/10.5194/hess-21-2615-2017>
- Pradhan, B. (2010). Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology*, 9(2).
- Quesada-Román, A., Ballesteros-Cánovas, J. A., Granados-Bolaños, S., Birkel, C. & Stoffel, M. (2020). Dendrogeomorphic reconstruction of floods in a dynamic tropical river. *Geomorphology*, 359, 107133. <https://doi.org/https://doi.org/10.1016/j.geomorph.2020.107133>
- Quesada-Román, A., Ballesteros-Cánovas, J. A., Granados-Bolaños, S., Birkel, C. & Stoffel, M. (2022). Improving regional flood risk assessment using flood frequency and dendrogeomorphic analyses in mountain catchments impacted by tropical cyclones. *Geomorphology*, 396, 108000. <https://doi.org/https://doi.org/10.1016/j.geomorph.2021.108000>
- Rahmati, O., Pourghasemi, H. R. & Zeinivand, H. (2016). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, 31(1), 42–70. <https://doi.org/10.1080/10106049.2015.1041559>
- Riley, S. J., DeGloria, S. D. & Elliot, R. (1999). Index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5(1–4), 23–27.
- Roccati, A., Paliaga, G., Luino, F., Faccini, F. & Turconi, L. (2021). GIS-Based Landslide Susceptibility Mapping for Land Use Planning and Risk Assessment. In *Land* (Vol. 10, Issue 2). <https://doi.org/10.3390/land10020162>
- Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W. & Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8, 101489–101499.
- Saha, S. (2017). Groundwater potential mapping using analytical hierarchical process: a study on Md. Bazar Block of Birbhum District, West Bengal. *Spatial Information Research*, 25(4), 615–626. <https://doi.org/10.1007/s41324-017-0127-1>
- Said, N., Ahmad, K., Riegler, M., Pogorelov, K., Hassan, L., Ahmad, N. & Conci, N. (2019). Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications*, 78(22), 31267–31302. <https://doi.org/10.1007/s11042-019-07942-1>
- Samanta, S., Pal, D. K. & Palsamanta, B. (2018). Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Applied Water Science*, 8(2), 66. <https://doi.org/10.1007/s13201-018-0710-1>
- Sarhadi, A., Soltani, S. & Modarres, R. (2012). Probabilistic flood inundation mapping of ungauged rivers: Linking GIS techniques and frequency analysis. *Journal of Hydrology*, 458–459, 68–86. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.06.039>

- Shafapour Tehrany, M., Kumar, L., Neamah Jebur, M. & Shabani, F. (2019). Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk*, 10(1), 79–101. <https://doi.org/10.1080/19475705.2018.1506509>
- Shafapour Tehrany, M., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W. & Xie, X. (2017). GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. *Geomatics, Natural Hazards and Risk*, 8(2), 1538–1561. <https://doi.org/10.1080/19475705.2017.1362038>
- Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K. & Shirzadi, A. (2018). Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *Journal of Environmental Management*, 217, 1–11. <https://doi.org/https://doi.org/10.1016/j.jenvman.2018.03.089>
- Skakun, S. (2010). A neural network approach to flood mapping using satellite imagery. *Computing and Informatics*, 29(6), 1013–1024.
- Sørensen, R., Zinko, U. & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, 10(1), 101–112. <https://doi.org/10.5194/hess-10-101-2006>
- Souissi, D., Zouhri, L., Hammami, S., Msaddek, M. H., Zghibi, A. & Dlala, M. (2020). GIS-based MCDM – AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia. *Geocarto International*, 35(9), 991–1017. <https://doi.org/10.1080/10106049.2019.1566405>
- Tehrany, M. S., Jones, S. & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA*, 175, 174–192. <https://doi.org/https://doi.org/10.1016/j.catena.2018.12.011>
- Tehrany, M. S. & Kumar, L. (2018). The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. *Environmental Earth Sciences*, 77(13), 490. <https://doi.org/10.1007/s12665-018-7667-0>
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, 504, 69–79. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2013.09.034>
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332–343. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.03.008>
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. (2015). Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment*, 29(4), 1149–1165. <https://doi.org/10.1007/s00477-015-1021-9>

- Tehrany, M. S., Pradhan, B., Mansor, S. & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA*, 125, 91–101. <https://doi.org/https://doi.org/10.1016/j.catena.2014.10.017>
- Tien Bui, D., Khosravi, K., Li, S., Shahabi, H., Panahi, M., Singh, V. P., Chapi, K., Shirzadi, A., Panahi, S., Chen, W. & Bin Ahmad, B. (2018). New Hybrids of ANFIS with Several Optimization Algorithms for Flood Susceptibility Modeling. In *Water* (Vol. 10, Issue 9). <https://doi.org/10.3390/w10091210>
- Tingsanchali, T. & Karim, F. (2010). Flood-hazard assessment and risk-based zoning of a tropical flood plain: case study of the Yom River, Thailand. *Hydrological Sciences Journal*, 55(2), 145–161. <https://doi.org/10.1080/02626660903545987>
- Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N. & Tehrany, M. S. (2014). Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena*, 118, 124–135.
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M. & Feyen, L. (2017). Extreme sea levels on the rise along Europe's coasts. *Earth's Future*, 5(3), 304–323. <https://doi.org/https://doi.org/10.1002/2016EF000505>
- Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtoet, W., Lucas, P. L., van Vuuren, D. P. & Ward, P. J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, 6(4), 381–385. <https://doi.org/10.1038/nclimate2893>
- Yariyan, P., Janizadeh, S., Van Phong, T., Nguyen, H. D., Costache, R., Van Le, H., Pham, B. T., Pradhan, B. & Tiefenbacher, J. P. (2020). Improvement of Best First Decision Trees Using Bagging and Dagging Ensembles for Flood Probability Mapping. *Water Resources Management*, 34(9), 3037–3053. <https://doi.org/10.1007/s11269-020-02603-7>
- Yesilnacar, E. & Topal, T. (2005). Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, 79(3–4), 251–266.
- Youssef, A. M., Pradhan, B. & Sefry, S. A. (2015). Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences*, 75(1), 12. <https://doi.org/10.1007/s12665-015-4830-8>
- Yu, M., Yang, C. & Li, Y. (2018). Big Data in Natural Disaster Management: A Review. In *Geosciences* (Vol. 8, Issue 5). <https://doi.org/10.3390/geosciences8050165>