Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model

Andaleeb Yaseen1 · Jianzhong Lu1 · Xiaoling Chen1

Accepted: 13 January 2022 / Published online: 11 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Floods are among the most destructive natural hazards. Therefore, their prediction is pivotal for flood management and public safety. Factors contributing to flood are different for every watershed as they depend upon the characteristics of each watershed. Therefore, this study evaluated the factors contributing to flood and the precise location of high and very high flood susceptibility regions in Karachi. A new ensemble model (LR-SVM-MLP) is introduced to develop the susceptibility map and evaluate influencing factors. This ensemble model was formed by employing a stacking ensemble on Logistic Regression (LR), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). A spatial database was generated for the Karachi watershed, which included; twelve conditioning factors as independent variables, 652 flood points and the same number of non-flood points as dependent variables. This data was then randomly divided into 70% and 30% to train and validate models, respectively. To analyse the collinearity among factors and to scrutinize each variable’s predictive power, multicollinearity test and Information Gain Ratio were applied, respectively. After training, the models were evaluated on various statistical measures and compared with benchmark models. Results revealed that the proposed ensemble model outperformed Logistic Regression (LR), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) and produced a precise and accurate map. Results of ensemble model showed 99% accuracy in training and 98% accuracy in testing datasets. This ensemble model can be used by flood management authorities and the government to contribute to future research studies.

Keywords Flash flood · Logistic regression · Support vector machine · Multi-layer perceptron · Ensemble classifier

1 Introduction
Floods are considered destructive natural hazards that causes millions of human deaths and result in billions of economic losses worldwide. The flood threat cannot be put aside because of future climate change Odoh and Chilaka (2012). It is predicted that the majority of the world will be threatened by the flood frequency and intensity (Jonkman 2005). Information about inundation can be obtained through different techniques of remote sensing, either air-borne or space-borne Schumann and Moller (2015).

The generation of susceptibility maps for flash floods is challenging, especially in large regions, as flash floods are quite complicated because they are area-dependent and arise nonlinearly to the variety of Spatio-temporal scales (Ahmadlou et al. 2019). Recently, promising results of machine learning models have been reported in literature worldwide for solving problems related to natural hazards.

For flood modelling, some of the qualitative multicriteria decision models include Analytical Hierarchy Process (AHP) (Kazakis et al. 2015; Rahmati et al. 2016), Fuzzy AHP (Ekmekcioğlu et al. 2021), SAW (Meshram et al. 2020), Interval Rough AHP (Sepehri et al. 2020), Frequency Ratio (FR) (Lee et al. 2012; Tehrany et al. 2015a), Weights of Evidence (WOE) (Rahmati et al. 2016; Tehrany et al. 2014a), and quantitative AI models for example; Logistic Regression (LR) (Fekete 2009; Tehrany et al. 2014a), Neuro-Fuzzy Logic (Tien Bui et al. 2016a; Mukerji et al. 2009), Decision Trees (DT) (Tehrany et al. 2013),
Support Vector Machine (SVM) (Tehrany et al. 2015b, 2015a). Among other ML models, MLP is one of the most used models for flood modelling as it has the capability of processing non-linear and multivariate data and has the potential of universal modelling (Youssef et al. 2011). Due to its prediction accuracy, SVM is becoming an emerging choice for hydrologists (Zhao et al. 2018; Tehrany et al. 2015a, b; Choubin et al. 2019). However, as different models act differently in given scenarios, no consensus has yet been reached on a single model for flood susceptibility modelling, and also each model carries some drawbacks with it. Therefore, scientists now address these problems by forming hybrids of different machine learning models together. Ensemble or hybrid machine learning models showed high accuracy and better performance than conventional methods in many previous studies (Pham et al. 2017; Saha et al. 2021).

Hybrid methods are extensively used in literature for flood modelling. Some ensemble methods, for example, adaptive neuro-fuzzy interference systems and their optimization algorithms (Razavi Termeh et al. 2018; Bui et al. 2016), have become famous for their effective prediction. Chapi et al. (2017) applied bagging ensemble on Logistic Model Tree, which performed best compared with other models. Similarly, Ngo et al. (2018) developed a new hybrid approach (FA-LM-ANN) by integrating Firefly Algorithm (FA), Levenberg–Marquardt (LM), and Artificial Neural Network (ANN) to study flood susceptibility, which proved to be the best model as compared to its benchmark models.

For this study, Logistic Regression (LR), Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) were chosen. Logistic Regression is simple but extremely effective in evaluating the relationship between dependent and independent variables for the classification problem has been used by several researchers (Chapi et al. 2017). For flood susceptibility, SVM uses the flood predictors’ non-linear transformations in higher dimensional feature space (Yilmaz 2010; Ghorbanzadeh et al. 2019; Nguyen et al. 2019). SVM reduces the test error by finding optimal hyper-plane that could separate flood and non-flood (support vectors) (Kalantar et al. 2018). MLP has been widely used for natural hazard prediction because it is highly capable of modelling the non-linear relationship between an explanatory variable and target variable (Kia et al. 2012).

The stacking classifier technique is used to make an ensemble of these models. The stacking classifier is reliable because it generates predictions based on two-level information. First, the base classifier predicts and then the meta classifier refines any biases occur in the prediction of base classifier (Hu et al. 2020). To make the prediction more efficient and reliable, choosing suitable models for base and meta classifier is essential.

The Novelty of this research is (1) Ensemble of these three models using stacking classifier was never done before. (2) The effectiveness of this hybrid (LR_SVM-MLP) for flood susceptibility was never accessed. (3) To that of the authors best knowledge, no single study has been done for flood susceptibility analysis in Pakistan using machine learning models and contributing factors. The main objectives of this study were; (1) To compare the performance of the new ensemble model on flood susceptibility with commonly used models (2) To evaluate the importance of flood conditioning factors and their contribution to flood susceptibility in an arid region of Pakistan.

For the training process, 12 conditioning factors and 652 flood events were used. All models were trained individually, and their performances were compared with the ensemble model using several statistical measures. The fundamental purpose of this technique is not only to highlight flood-prone regions precisely and accurately but also to analyze the contributing factors so that authorities and government can plan flood management accordingly.

2 Methods

2.1 Study area

Located in the 24.86 N and 67.010 E in world coordinates, Karachi lies in the southern region of Pakistan. An increment from 466.5 Km² to 666.18 Km² was observed in the built-up area of Karachi from 1998 to 2018. It has a very humid to hot and worst climate in the summer season with very low annual precipitation (Raza et al. 2019). Two main rivers which pass from Karachi are Malir and Liyari. Malir flows from the east towards the south, while Liyari flows North to the Southwest Tariq et al.(2016). Karachi’s population jumped from 450,000 in 1947 (Hassan. 2017) to more than 16 million in 2017 (Shahbaz et al. 2017) and is expected to reach more than 20 million by the year 2025 (Mangi et al. 2020). Karachi has the major seaport of Pakistan and faces floods every 2–3 years. Since 2000, Karachi faced floods in 2000, 2006, 2007, 2011, 2013, 2019, 2020. Total 42 people were killed in the flood of 2019 and 41 in the flood of 2020 (Arif Hasan 2020) (Fig. 1).

2.2 Flood conditioning factors and database generation

Factors that affect flood probability vary as they mainly depend on the watershed characteristics (Tien Bui et al. 2016). Therefore it is vital to determine the influencing
factors for each watershed for the accurate flood susceptibility mapping of that area (Chapi et al. 2017).

This study is primarily based on geospatial data, extracted from Digital Elevation Model (DEM) and some other factors. A total of 12 factors were analyzed, which are elevation, Slope, Curvature, Stream Power Index (SPI), Topographic Wetness Index (TWI), Rainfall, Lithology, soil Type, Distance from Stream, Stream Density, Normalized Difference Vegetation Index (NDVI), and LandUse. The details of data and their sources are given in Table 1.

### Table 1  Sources of different data used in study

| Factors                  | Source                  | Links                                                                 |
|--------------------------|-------------------------|-----------------------------------------------------------------------|
| Elevation                | Digital Elevation Model  | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)  |
|                          | (DEM 30 m)              | [https://asterweb.jpl.nasa.gov/gdem.asp](https://asterweb.jpl.nasa.gov/gdem.asp) |
| Slope                    | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Curvature                | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Stream Power Index (SPI) | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Topographic Wetness Index (TWI) | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Distance to River        | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Stream Density           | Advance Spaceborne Thermal Emission and Reflection (ASTER) as Global Elevation Model (GDEM)                  |
| Lithology                | Geological Map          | [https://pubs.er.usgs.gov/publication/ofr97470C](https://pubs.er.usgs.gov/publication/ofr97470C) |
| Soil Type                | Soil Map                | Food and agricultural Organization (FAO) website [http://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/en/](http://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/en/) |
| Land Use                 | OLI of Landsat 8 image  | U.S. Geological Survey (USGS) [https://glovis.usgs.gov/app?fullscreen=0](https://glovis.usgs.gov/app?fullscreen=0) |
| NDVI                     | GIOVANNI                | [https://giovanni.gsfc.nasa.gov/giovanni/](https://giovanni.gsfc.nasa.gov/giovanni/) |
Fig. 2 Influencing factors for flood susceptibility mapping: a Elevation, b Slope, c Topographic Wetness Index, d curvature, e Stream Power Index, f Lithology, g Soil, h Land Use, i NDVI, j Rainfall, k Distance from River, l Stream Density
This data was processed in ArcGIS 10.8, and the database was constructed as the matrix of 3459 columns and 3260 rows and spatial database of cell size (30 m*30 m). It is essential to assemble data based on past flood events for future flood prediction (Manandhar, 2010; Bui et al. 2012). For this study, the flood of 2020 was used for data collection. As a large number of dependent variables gives more accurate results, total 652 flood points were collected.
from the areas that regularly faced floods in the past, and a similar number of non-flood points were collected to avoid biases of data. This data was then randomly classified as training data (70%) and validation data (30%); this method of random partition division repeatedly split the data without replacing the training and test dataset. The choice of contributing factors for this study was made using the information of previous literature (Tehrany et al., 2014a, b; Rahmati et al., 2016; Khoosravi et al. 2016).

The elevation is considered one of the most essential factors in flood analysis (Dodangeh et al. 2020). Previous research has shown that elevation and flood occurrence has an inverse relation with elevation; floods usually occur at relatively low elevation areas compared to the high elevation regions of the same area (Chen et al. 2020). The elevation of Karachi lies from −10 m–501 m. The elevation was constructed with five intervals from −10–59 m, 59.001–122 m, 122.01 m–198 m, 198.01 m–304 m, and 304.01 m–501 m (Fig. 2).

The surface of the slope area is highly pertinent to flood as floodwater flow with greater speed from the steep surface, and water tends to get absorbed in plain areas, led more damage (Steveaux et al. 2020). The slope of the study area was classified into five classes: 0–2.018, 2.018–5.55, 5.55–12.11, 12.11–23.46, 23.46–64.33. Curvature effect the flooding water budge; therefore, its study is necessary for flood modelling (Ahmadlou et al. 2019). The curvature value of study area lied in five intervals: 0.5485–20.159, 0.014878–0.54849, −0.25192–0.014877, −0.65213−0.25193, −13.859−0.65214. The topographic wetness index ascertained the spatial wetness status of the basin area, which affects the occurrence of floods in the region (Meles et al. 2020). TWI was calculated as follows (Beven and Kirkby 1979):

$$TWI = \ln\left(\frac{\varnothing}{\tan \beta}\right)$$

where, $\varnothing$ is the cumulative upslope area drainage through the point per unit contour length and $\tan \beta$ is the slope angle at the point. TWI was classified into the five classes: 2.8124–6.9503, 6.9504–8.541, 8.5418–10.451, 10.452–13.157, and 13.158–23.104. (Fig. 2). SPI determined the erosive power and discharge relative to a particular area (Poudyal et al. 2010).

$$SPI = As \tan \beta$$

where As is the area of specific basin, and $\beta$ is the local slope gradient (in degree).

Distance from the river is essential for identifying flood-prone areas in the watershed Tehrany et al. (2015a). Distance from the river was extracted in ArcGIS environment using Euclidean distance buffer and classified into eight classes: 0–0.5 km, 0.5–1 km, 1–1.5 km, 1.5–2 km, 2.5–3 km, 3–3.5 km, 3.5–4 km, >4 km. Stream density is another critical influencing factor while studying the flood susceptibility of an area. It is calculated by Elmore et al. (2013).

$$Stream\ density = \frac{stream\ length (m)}{Basin\ Area (km^2)}$$

Stream density was extracted from DEM in ArcGIS environment using line density. The five classes of stream density are: 0–20.603, 20.604–48.338, 48.339–77.657, 77.658–114.11, 114.12–202.07. The probability of flooding significantly increases with increasing rainfall events and the duration of rain (Lu et al. 2020). Rainfall with the monthly average of the last ten years was used in this study. Rainfall data was downloaded from GIOVANNI, and interpolation was calculated in the ArcGIS environment using Inverse Distance Weighted (IDW). This method has been proved to be effective for rainfall interpolation by Chen and Liu et al. (2012). The interpolation result was converted into five intervals which are, 15.083–15.509 mm, 15.51–15.962 mm, 15.963–16.424 mm, 16.425–16.796 mm, and 16.797–17.394 mm.

Lithology and soil type affects the hydrology of the basin. Areas with highly permeable subsoils and higher resistant rocks allow minor drainage (Çelik et al. 2012; Srivastava et al. 2014). The lithology and soil regulate flooding by regulating the erodibility and permeability in a watershed Stefanidis and Stathis (2013). Lithological data was obtained from the USGS site, while soil data was obtained from Food and Agriculture Organization (FAO). Both of these data were processed to extract data of the study area. The study area comprises four types of rocks and two types of soil (Fig. 2). Land Use types are significant indicators while assessing the probability of a flood in any area Rahmati et al. (2015). Land use types and NDVI greatly influence and control the infiltration; for example, areas with high forest land allow more infiltration and thus less runoff compared to the areas with the concrete surface (Tehrany et al. 2014a, 2014b). For land use analysis, the data was obtained from Raza et al. (2019); they classified Karachi into vegetation class, water class, settlement, and barren land. Similarly, NDVI is used to assess the relationship between vegetation and flooding. Tehrany et al. (2013). The index of NDVI ranges between −1 and 1. NDVI was calculated using Landsat 8 OLI imagery. After preprocessing of imagery, NDVI was calculated in ArcGIS using the formula Tucker and Sellers (1986)

$$NDVI = \frac{(near\ infrared\ _red)}{(near\ infrared\ +\ red)}$$
2.3 Multicollinearity Test

If the study uses multiple independent factors, the presence of collinearity can significantly influence the final results. Therefore, it is necessary to ascertain the collinearity among those factors (Arabameri et al. 2019; Wang et al. 2021). Variance Inflation Factor (VIF) helps multicollinearity detect the independent variables that can be further used in the model Arabameri et al. (2020a). Factor having VIF values greater than 4 pose severe concerns with multicollinearity. If the test reports any collinear variable, then this specific variable should be removed and not used for the prediction purpose in the model (Arabameri et al. 2019; Bui et al. 2019). In this study, twelve conditioning factors were considered for analysis; therefore, it is essential to check their collinearity. The VIF among variables will be determined by using the following equation. Where Tolerance is referred to as variability shown by an independent variable and not explained by other independent variables.

\[
\text{VIF} = \frac{1}{\text{Tolerance}}
\]  

(5)

2.4 Information Gain Ratio

As the present study consists of numerous factors, there is the possibility that some factors can reduce the model’s performance; therefore, to reduce uncertainty and noise in the result, it is necessary to the predictive power of each factor. If the value of the factor is 0, it will be removed from the study. Many previous researchers select this value to determine those factors that have zero influence on the study (Chapi et al. 2017; Khosravi et al. 2016). The predictive power of influencing factors will be evaluated using Information Gain Ratio (IGR); the higher the value of IGR, the higher the predictive power for the flood Chapi et al. (2017).

Let D be the training dataset composed of n number of samples, and n (Yi, D) is the number of samples present in the training data D, belongs to the label of class Yi (flood or non-flood). The IGR for each conditioning factor, for example, the slope, will obtain as follows:

\[
\text{IGR}(D, \text{Slope}) = \frac{\text{Entropy}(D) - \text{Entropy}(D, \text{Slope})}{\text{SplitEntropy}(D, \text{Slope})}
\]  

(6)

\[
\text{Entropy}(D) = -\sum_{i=1}^{2} \frac{n(Yi, \text{Slope})}{|D|} \log_2 \frac{n(Yi, \text{Slope})}{|D|}
\]  

(7)

\[
\text{Entropy}(D, \text{Slope}) = \sum_{j=1}^{m} \frac{|D_j|}{|D|} \text{Entropy}(D)
\]  

(8)

\[
\text{SplitEntropy}(D, \text{Slope}) = -\sum_{j=1}^{m} \frac{|D_j|}{|D|} \log_2 \frac{|D_j|}{|D|}
\]  

(9)

2.5 Frequency Ratio

To analyse the relative importance of each factor and the contribution of each class of factors towards flood, flood frequency ratios were analysed Gayen et al. (2019). The value of the frequency ratio can be expressed as the percentage of flood pixels shared in the domain of pixels of the whole area. FR can be calculated as follows:

\[
\text{FR} = \frac{Li/Ci}{L/C}
\]  

(10)

where Li represents the flood cells in the ith category, whereas Ci represents total cells in the ith category, L is the total flood cells, and C is the total cells. The FR values greater than 1 represent the large concentration of that class in the flood area, while if the value of the class is less than 1 it suggests that this class has a very little contribution of flood cells in the data layer.

3 Theoretical background of the models

3.1 Logistic Regression (LR) Model

Logistic regression is extensively used to analyze binary variables (Das et al. 2010; Dai et al. 2001; Chen et al. 2017), and LR is used to evaluate the relationship between dependent and independent variables. In this study, a standard LR model will be applied to evaluate the relationship between the conditioning factors, an independent variable (predictor values) of this study, and flood occurrence, a dependent variable.

Based on these conditioning factors (predictor values), the absence or presence of characteristics will be predicted by the maximum likelihood method Xu et al. (2013). Results of LR will be evaluated based on the probability of the dependent variable as they are constrained to fall more in the category of 0 or 1 Shahabi et al. (2015). The higher the coefficient, the more impact it has on the flood occurrence. The following equations will be used to derive the probability for flood from the LR coefficient.

\[
P = e^{\beta} / (1 + e^{\beta})
\]  

(11)

\[
Z = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n
\]  

(12)
where $P$ represents the probability of flood, $Z$ is the linear combination or linear logistic model, $b_0$ represents the intercept of the model, $n$ shows the number of flood conditioning factors, $b$ tells the weights of each condition factor, and $X$ represents the flood influencing factors such as elevation.

### 3.2 Support vector machine (SVM)

Support Vector Machine is a widely popular model of machine learning. A hyper-plane has been generated using the training datasets when converting from the actual support vector machine datasets to high dimensional feature space has occurred Choi et al. (2020). The functional performance of the model depends mainly on the usage of a suitable kernel. It is a common observation that, like other neural networks, SVM also faces the problem of over-fitting and under-fitting. The theory of SVM has been documented by several researchers Kecman (2001) in the high-dimensional feature space, the hypothetical space of support vector machine is limited to the linear function. These hypotheses are then trained with a learning algorithm based on optimisation theory. These algorithms apply statistically extract learning theory. In this way, optimizing the machine to generalize is achieved by the fine-tuning of the learning machine. An appropriate choice of kernel allows the non-separable data in the original input space to become separable in feature space. The kernel can be defined as the function that directly calculates the inner product from the input points.

### 3.3 Multi-layer perceptron (MLP)

Artificial Neural Networks act as the black box to impersonate the human brain’s structure and function Kia et al. (2012). MLP has high stability while having a smaller structure than most other neural networks Wang et al. (2021), thus selected for the present study. The structure of an MLP consists of an input layer, hidden layer, and output layer. Flood conditioning factors are associated with the hidden layer, the output layer is the flood or non-flood, and the main aim of the hidden layer is to convert input into output layer (Fig. 3). The weight adjustments among neurons respond to errors between target output values and actual output values. The training of the neural nets in MLP consists of basically two major steps; 1. Use the forward propagation for the input data (conditioning factors) through the hidden layer to get the output, and then to estimate the difference, this output is compared with preset values 2. Adjust the weights of the connection so that the best results can be obtained with the minimum difference. The classification function of MLP for the present study can be written as Pham et al. (2017).

$$t_i = f(x_i)$$  \hspace{1cm} (13)

$x_i$ is the ith the vector of flood conditioning factor, whereas $t_i$, $i = 0$ is for non-flood pixels and 1 for flood pixels, and $f(x_i)$ is a hidden function optimized by adjustable network weights for the given architecture during the training process.

The error which occurs during the training input pattern is equal to the difference between the network output $o_k$ and target out $d_k$ and can be expressed by the following equation.

$$e_k = (d_k - o_k)$$  \hspace{1cm} (14)

Weights adjustments among layers can reduce the error propagating from the output back through a neural network Lee et al. (2003). Adjustments of weights will be done by the following equation

$$\Delta w_{ij}(n + 1) = \eta (\delta j * o_i) + \alpha \Delta w_{ij}(n)$$  \hspace{1cm} (15)

where $\Delta w_{ij}(n + 1)$ and $\Delta w_{ij}(n)$ are the weight changes of epochs $(n + 1)$ and $(n)$, respectively, $\eta$ is the learning rate, $\delta$ is the rate of change in the error, and $\alpha$ is the momentum coefficient.

![Fig. 3 The general structure of an MLP model used for flash flood modelling](image)
3.4 Stacking classifier for flood susceptibility mapping

Stacking classifier is the global classifier that combines the output of different classifiers. In the stacking classifier, the final prediction is made using two steps. In the first step, base learning is used to predict the values from the dataset to feed as the input for the second-stage learner. The meta-model is often simple, so it can provide the smooth interpretation of predictions made by the base classifier. Therefore, it is recommended that linear models should be used as meta classifier for instant linear regression, for regression task, which predicts numeric value and logistic regression for classification tasks that predicts class labels. The stacking classifier’s overall performance depends on the selection of models Dou (2020). The powerful and complex models are usually used as base classifiers Pourghasemi (2017); therefore, the SVM and MLP are used as base classifiers in this study. SVM and MLP proved to be an excellent combination as ensemble models in previous research as well (Chen 2017; Pourghasemi 2017; Hu 2020).

In this study, Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) will be used as base classifiers while Logistic Regression (LR) will use as a meta-classifier Fig. 4.

3.5 Comparison and evaluation of the models

All models will be compared based on their accuracy, confusion matrix (cross table), sensitivity which is true positive (TP) or recall, specificity (True Negative), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Area Under Receiver Operating Curve (AUROC) for both training and validation dataset. The confusion matrix shows the four types of results; True Positive, False Positive, True Negative, and True Positive. Based on these possible four results, the precision and recall are formulated as (Onan 2015).

\[
\text{Precision} = \frac{TP}{TP + FP}
\]  
(16)

\[
sensitivity = \frac{TP}{TP + FN}
\]  
(17)

\[
\text{specificity} = \frac{TN}{TN + FP}
\]  
(18)

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  
(19)

Moreover, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) will be used for flood mapping (Tien Bui et al. 2016b; Kia et al. 2012). MAE will be used for validation purposes, while RMSE is sensitive to the large dataset (Chai and Draxler (2014).

\[
\text{RMSE} = \sqrt{1/n \sum_{i=1}^{n} (X_{\text{predicted}} - X_{\text{actual}})^2}
\]  
(20)

\[
\text{MAE} = \sqrt{1/n \sum_{i=1}^{n} |X_{\text{predicted}} - X_{\text{actual}}|}
\]  
(21)

While \( n \) represents the total number of samples in training or validation datasets, \( X_{\text{predicted}} \) is the predicted values of training or validation datasets, and \( X_{\text{actual}} \) is the actual values we got as output values from the model.

Receiver Operating Characteristic (ROC) curve was first used by DeLeo (1993) as another way of finding the quality and predictive power of probabilistic models (Pham et al. 2017; Shahabi et al. 2015). Graphically, the sensitivity is plotted on the x-axis while the specificity rate is plotted on the y-axis Gorsevski et al. (2006). The quantitative index, AUROC, measures the general performance of the model Pham et al. (2017). The more the AUROC, the better performance of the model is it ranges from 0.5, which is an inaccurate model, to 1, which is a perfect model (Tien Bui et al. 2016). AUROC can be calculated as:

\[
\text{AUROC} = \sum TP + \sum TN/P + N
\]  
(22)

Figure 5 represents basic steps used for this study.

3.6 Sensitivity analysis

In this study, the sensitivity of flood conditioning factors was analysed by the jackknife test (Park, 2015). It is believed to have a high capability to deal with a wide range...
of practical problems (Bandos et al. 2017). This test is done using the percentage of relative decrease of the AUC to determine the contribution of factors following equation (Park 2015):

\[
PRDi = \frac{100}{C2} = \frac{[\text{AUCall} - \text{AUCi}]}{\text{AUCall}} \quad (23)
\]

In this equation AUCall is the value of AUC calculated from the prediction by all factors. AUCi and PRDi are the AUC values and percentage of relative decrease of AUC respectively, where the ith factor removed from the process of prediction.

4 Results

4.1 Data preparation

In this study, a total of 12 influencing factors and 652 flood points were selected to generate the flood susceptibility map of Karachi, Pakistan. As evaluating flood-prone areas is based on the application of binary models, 652 points of non-flood points were used for the analysis. These non-flood locations were selected randomly at relatively high elevation areas where chances for flood were meager. The data was divided into 70% for training and 30% for testing models. This division is essential for the construction of a proper database.

4.2 Selection of conditioning factor

Based on VIF values, it is found that no factor had a VIF value greater than 4. Therefore no factor had the problem of multicollinearity. The highest VIF is shown by rainfall (3.9), while the lowest is shown by NDVI (1.14), but overall all the values showed negligible collinearity (Table 2). This result shows that all factors are independent of each other.

4.3 Information Gain Ratio of effective Factors

To access the effect of conditioning factors on the flood. The predictive power of all factors was analyzed using Information Gain Ratio (IGR) method on the training dataset. It helps in eliminating the unwanted factors or the factors whose presence create noise or are unproductive for the data Pham et al. (2016). Results showed that some factors had a strong influence on the study, while some showed zero contribution (Fig. 6).

LULC showed the highest value of IGR (0.55), which means LULC has the highest impact on flood occurrence, followed by elevation and rainfall (0.5). While ‘distance from the river’ is removed because it showed zero contribution in flood, as shown in Fig. 6, similar results were obtained by Bui et al. (2015).

4.4 Spatial Relationship between flood probability and FCFs

FR of each class shows its contribution in the probability of flood probability. The highest FR values among slopes (100%) lie in the class 0–2. Similarly, all the flood pixels lie in the class range – 10 m–56 m in case of elevation, which showed that low lying areas of the study area are more prone to flood, and in the case of lithology, all flood points fell in the class of Paleogene sedimentary rocks; all
other classes of these factors showed negligible contribution towards flood (Table 3). Other factors have the distribution of flood pixels in more than two classes. This assessment gave us an absolute understanding of the contribution of each class of all the influencing factors in the flood occurrence in the study area Table 3.

4.5 Stacking ensemble of the models

Although single machine learning models gives decisive predictions, integrating these machine learning algorithms or with some other statistical techniques gives better performance (Hong et al. 2018; Arabameri et al. 2019). Ensembles methods gave much better and precise outcomes for the flood susceptibility analysis in the past Pham et al. (2017). Therefore, an ensemble model was constructed for flood susceptibility of the study area using the stacking ensemble technique.

For the stacking of models, Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) were trained as a base classifier on the training dataset, whereas Logistic Regression was trained on the outcome of the base classifier. Different optimizations of models were applied to get the best result and avoid overfitting.

To obtain the best performance from SVM, SVM was trained by using four types of kernels, namely polynomial kernel (PL), Radial based function (RBF), Linear Kernel (LN), and sigmoid kernel (SIG). Except for RBF, all other kernels showed the problem of overfitting or under-fitting, ‘RBF’ kernel with a C value of 1.0 showed perfect performance with a root mean square value of 0.18. Some previous research also tried different kernels (Tien Bui et al. 2012; Tehrany et al. 2015a, b), and the results showed that the radial base kernel gave the best performance. Numerous studies confirm that RBF outweighs other kernel functions in the case of flood susceptibility (Chen et al. 2020; Yang and Cervone 2019). Similarly, MLP was executed several times with the different number of neurons and hidden layers, and the final choice was based on the highest accuracy and lowest RMSE. After many trials and errors, we found that MLP in our case gave the best performance with 20 hidden layers and 30 output neurons. (Table 4).

4.6 Model validation and comparison

After training the models on their best optimization and hyper-parameter tuning, their performances were compared by the difference in their accuracy, precision, recall, Etc. The evaluation was performed both on the training and validation datasets because the training dataset represents the model’s fitting skill and the validation dataset indicates the model’s generalizations ability. In the case of the validation dataset highest accuracy was 98 obtained by ensemble model, SVM and MLP showed an accuracy of 96%, while LR showed 93.9% accuracy. For the training

---

**Table 2** Multicollinearity of 12 conditioning factors

| Variables     | VIF  |
|---------------|------|
| Curvature     | 1.360|
| Elevation     | 3.265|
| Rainfall      | 3.905|
| Slope         | 3.265|
| SPI           | 3.206|
| Stream Density| 1.922|
| Twi           | 3.046|
| LULC          | 2.195|
| Soil          | 3.458|
| NDVI          | 1.149|
| TPI           | 1.70 |
| Distance from river | 1.562 |

---

![Information Gain Ratio](image-url)
| Factor                | Class                  | Pixel in domain | Flood pixel | Frequency Ratio |
|----------------------|------------------------|-----------------|-------------|-----------------|
|                      | No | %     | No  | %     |                |
| Slope                | 0–2.0183  | 4,013,495 | 97.44 | 288,900 | 100  | 1.0  |
|                      | 2.0184–5.5504 | 88,322  | 2.144 | 0  | 0  | 0  |
|                      | 5.5505–12.11 | 14,741   | 0.35 | 0  | 0  | 0  |
|                      | 12.111–23.463 | 2014    | 0.04 | 0  | 0  | 0  |
|                      | 23.464–64.334 | 210     | 0.00 | 0  | 0  | 0  |
| Curvature            | 0.5485–20.159 | 41       | 0.00 | 33,300 | 11.52 | 0.99 |
|                      | 0.014878–0.54849 | 541,051 | 13.13 | 255,600 | 88.47 | 0  |
|                      | 0.014878–0.54849 | 3,577,625 | 86.86 | 0  | 0  | 0  |
|                      | 0.65213–0.25193 | 63      | 0.00 | 0  | 0  | 0  |
|                      | 13.859–0.65214 | 2       | 4.85 | 0  | 0  | 0  |
| Soil                 | Calcaric regosol | 1,332,732 | 32.35 | 280,000 | 89.14 | 0.98 |
|                      | Lithosols | 2,786,049 | 67.64 | 8900  | 10.85 | 0.01 |
| Lithology            | Paleogene sedimentary rocks | 3,688,904 | 89.56 | 243,000 | 84.11 | 0.84 |
|                      | Neogene sedimentary rock | 94,038   | 2.27 | 45,900 | 15.88 | 0.15 |
|                      | Quaternary sediments | 93,538   | 2.27 | 0  | 0  | 0  |
|                      | Water     | 148,498  | 3.60 | 0  | 0  | 0  |
| Land use             | Water     | 69,758   | 1.72 | 900  | 0.31 | 0.03 |
|                      | Settlement | 722,266 | 17.86 | 258,300 | 89.40 | 0.87 |
|                      | Barren land | 3,001,990 | 74.27 | 22,500 | 7.78 | 0.01 |
|                      | Vegetation | 247,940  | 6.13 | 7200  | 2.49 | 0.07 |
| Stream Density       | 0–20.603 | 1,410,838 | 34.92 | 44,000 | 15.23 | 0.06 |
|                      | 20.604–48.338 | 1,497,870 | 37.08 | 184,500 | 62.69 | 0.64 |
|                      | 48.339–77.657 | 894,851  | 22.15 | 81,000 | 28.21 | 0.28 |
|                      | 77.658–114.11 | 205,764  | 5.09 | 17,100 | 5.95 | 0.05 |
|                      | 114.12–202.07 | 29,863   | 0.73 | 4500  | 1.56 | 0.01 |
| TWI                  | 2.8124–6.9503 | 1,317,735 | 31.99 | 76,500 | 26.47 | 0.26 |
|                      | 6.9504–8.5417 | 2,402,148 | 58.32 | 192,600 | 66.66 | 0.37 |
|                      | 8.5418–10.451 | 360,033  | 8.74 | 18,900 | 6.54 | 0.24 |
|                      | 10.452–13.157 | 36,486   | 0.88 | 900   | 0.31 | 0.11 |
|                      | 13.158–23.104 | 2380     | 0.05 | 0  | 0  | 0  |
| SPI                  | 9.2792–19.246 | 1,552,721 | 37.69 | 147,600 | 51.09 | 0.40 |
|                      | 6.843–9.2791 | 2,338,970 | 56.78 | 133,200 | 46.10 | 0.24 |
|                      | 5.2188–6.8429 | 208,299  | 5.09 | 7200  | 2.49 | 0.14 |
|                      | 3.8899–5.2187 | 18,072   | 0.43 | 900   | 0.31 | 0.21 |
|                      | 0.41993–3.8898 | 720     | 0.01 | 0  | 0  | 0  |
| Distance from River (Km) | 0–0.5 | 803,109 | 19.49 | 44,000 | 15.23 | 0.06 |
|                      | 0.5–1 | 671,489 | 16.30 | 57,500 | 19.90 | 0.10 |
|                      | 1–1.5 | 535,889 | 13   | 40,300 | 13.94 | 0.09 |
|                      | 1.5–2 | 400,699 | 9.7  | 41,500 | 14.36 | 0.13 |
|                      | 2.5–3 | 288,449 | 7    | 30,500 | 10.55 | 0.13 |
|                      | 3–3.5 | 205,989 | 5    | 75,100 | 25.99 | 0.46 |
|                      | 3.5–4 | 132,179 | 3.20 | 0  | 0  | 0  |
|                      | > 4   | 1,080,979 | 26.24 | 0  | 0  | 0  |
samples, the ensemble model ranked first with an accuracy of 99%, followed by MLP (98%), SVM (97%), and LR (94.8%). In the case of precision, the ensemble model showed 98% precision for the validation dataset and 99% for the training dataset, which showed that this model gave exceptionally precise locations of flood susceptibility.

Receiver operating characteristics curve was calculated on both training and validation datasets. In both training and validation datasets, it is observed that the ensemble model obtained the highest performance. The results of sensitivity and specificity of the ensemble model showed that the number of correctly classified flood pixels was 99% for the training dataset and 99.8% for validation datasets. Similarly, 97.6% and 99% were correctly classified as non-flood pixels for training and validation datasets, respectively, by ensemble model (Tables 5, 6, and Fig. 7).

It is necessary to categorize the final prediction map of flood susceptibility into different classes to easily visualize flood probability in the study area (Pham et al. 2017; Tehrany et al. 2014b).

| Table 3 (continued) |
|---------------------|
| **Factor** | **Class Pixel in domain** | **Flood pixel** | **Frequency Ratio** |
| **No** | **%** | **No** | **%** |
| **Rainfall (mm/)** | | | |
| 15.083–15.509 | 363,464 | 8.82 | 63,225 | 21.88 | 0.32 |
| 15.51–15.962 | 320,593 | 7.78 | 73,225 | 25.34 | 0.42 |
| 15.963–16.424 | 690,938 | 16.77 | 65,225 | 22.57 | 0.17 |
| 16.425–16.796 | 1,975,558 | 47.96 | 87,225 | 30.19 | 0.08 |
| 16.797–17.394 | 768,228 | 18.65 | 0 | 0 | 0 |
| **Elevation** | | | |
| 10–59 | 2,210,702 | 53.67 | 288,900 | 100 | 1 |
| 59.001–122 | 1,257,046 | 30.51 | 0 | 0 | 0 |
| 122.01–198 | 389,341 | 9.45 | 0 | 0 | 0 |
| 198.01–304 | 196,127 | 4.76 | 0 | 0 | 0 |
| 304.01–501 | 65,566 | 1.59 | 0 | 0 | 0 |
| **NDVI** | | | |
| 0.5777–0.2742 | 2463 | 0.06 | 1800 | 0.63 | 0.12 |
| 0.2741–0.197 | 31,549 | 0.78 | 163,800 | 57.41 | 0.87 |
| 0.1969–0.1651 | 1,680,915 | 41.64 | 118,800 | 41.64 | 0.01 |
| 0.165–0.1305 | 2,313,539 | 57.32 | 900 | 0.31 | 0.01 |
| 0.1304–0.1012 | 7618 | 0.18 | 0 | 0 | 0 |

| Table 4 Results of hyper parameter tuning of models |
|---------------------|
| **Algorithm** | **Parameters** |
| **LR** | C = 1.0, class_weight = None, dual = False, fit_intercept = True, intercept_scaling = 1, l1_ratio = None, max_iter = 100, multi_class = 'auto', n_jobs = None, penalty = 'l2', random_state = None, solver = 'lbfgs', tol = 0.0001, verbose = 0, warm_start = False |
| **SVM** | C = 1.0, break_ties = False, cache_size = 200, class_weight = None, coef0 = 0.0, decision_function_shape = 'ovr', degree = 3, gamma = 'scale', kernel = 'rbf', max_iter = -1, probability = False, random_state = None, shrinking = True, tol = 0.001, verbose = False |
| **MLP** | activation = 'relu', alpha = 0.0001, batch_size = 'auto', beta_1 = 0.9, beta_2 = 0.999, early_stopping = False, epsilon = 1e-08, hidden_layer_sizes = (11, 20, 30), learning_rate = 'constant', learning_rate_init = 0.001, max_fun = 15,000, max_iter = 100, momentum = 0.9, n_iter_no_change = 10, nesterovs_momentum = True, power_t = 0.5, random_state = None, shuffle = True, solver = 'adam', tol = 0.0001, validation_fraction = 0.1, verbose = False, warm_start = False |
Table 5  Model performance for training dataset for new ensemble model and other benchmark models

|                     | LR    | SVM   | MLP   | LR-SVM-MLP |
|---------------------|-------|-------|-------|------------|
| True Negative (TN)  | 446   | 469   | 468   | 457        |
| True Positive (TP)  | 444   | 446   | 445   | 470        |
| False Positive (FP) | 25    | 2     | 3     | 1          |
| False Negative (FN)| 32    | 30    | 31    | 19         |
| Sensitivity         | 96.01 | 96    | 93.4  | 96.1       |
| Specificity         | 93.68 | 99    | 99.3  | 99.7       |
| Precision           | 0.95  | 0.94  | 0.98  | 0.98       |
| Recall              | 0.96  | 0.96  | 0.97  | 0.07       |
| Accuracy            | 94.8  | 97    | 98    | 99         |
| Root Mean Square Error (RMSE) | 0.24  | 0.183 | 0.189 | 0.14       |
| Mean Absolute Error (MAE) | 0.060 | 0.033 | 0.035 | 0.021      |

Table 6  Model performance for validation dataset for new ensemble model and other benchmark models

|                     | LR    | SVM   | MLP   | LR-SVM-MLP |
|---------------------|-------|-------|-------|------------|
| True Negative (TN)  | 193   | 205   | 204   | 219        |
| True Positive (TP)  | 193   | 193   | 198   | 171        |
| False Positive (FP) | 13    | 1     | 2     | 1          |
| False Negative (FN)| 8     | 8     | 3     | 16         |
| Sensitivity         | 93.2  | 96    | 98.5  | 91.4       |
| Specificity         | 94.6  | 99.5  | 99    | 99.8       |
| Precision           | 0.94  | 0.92  | 0.96  | 0.99       |
| Recall              | 0.93  | 0.93  | 0.98  | 0.95       |
| Accuracy            | 93.8  | 96.6  | 96    | 98         |
| Root Mean Square Error (RMSE) | 0.22  | 0.148 | 0.110 | 0.198      |
| Mean Absolute Error (MAE) | 0.051 | 0.022 | 0.012 | 0.039      |

Fig. 7  Receiver operating characteristic (ROC) curve a: Training dataset, b: Validation dataset
Final flood susceptibility maps were constructed in five classes based on quantile classification; for example, classes of ensemble model were categorized in very low or non-susceptibility category (0.0–0.001), low (0.001–0.221), moderate (0.221–0.862), high (0.862–0.992), and very high (0.992–0.999). Table 7 represents the percentage of the area that lies in each class. The ensemble model showed that almost 16% of the region comes under the category of very high and 23% under the class of high susceptibility; both of these regions collectively made 39% of the high susceptibility zone (Fig. 8).

| Class   | LR (Km²) | %  | SVM (Km²) | %  | MLP (Km²) | %  | Ensemble (Km²) | %  |
|---------|----------|----|-----------|----|-----------|----|----------------|----|
| Very low| 630,258.68 | 17 | 742,548.2 | 20.03 | 674,299.39 | 18.19 | 722,960.96 | 19.50 |
| Low     | 852,427.64 | 22.9 | 744,375.7 | 20.08 | 799,383.23 | 21.56 | 767,380.096 | 20.70 |
| Medium  | 802,612.35 | 21.65 | 747,083.5 | 20.15 | 760,295.55 | 20.51 | 745,017.92 | 20.09 |
| High    | 716,998.33 | 19.3 | 788,454.3 | 21.27 | 742,257.6 | 20.02 | 850,915.008 | 22.95 |
| Very High| 704,484.73 | 19 | 684,320.1 | 18.46 | 730,546.04 | 19.70 | 620,507.84 | 16.73 |

Fig. 8 Flash Flood Susceptibility Maps a: Logistic Regression Map, b: Support Vector Machine Map, c: Multi-Layer Perceptron map d: ensemble map

4.7 Sensitivity analysis result

For flood susceptibility, choosing a suitable conditioning factor is very critical (Kourgialas and Karatzas, 2012). Therefore, jackknife test was used in this study to calculate the sensitivity of eleven conditioning factors. It is a simple method and reduces the bias in the estimator, which could cause if we apply a complex method. Figure 9 indicate the relative importance of each factor. Land Use was the most critical factor in this study; it had the highest contribution in all model’s predictions, followed by elevation and
rainfall. Overall, LULC, Rainfall, Elevation and NDVI are the most important contributor in all the models.

5 Discussion

Identifying flood-prone areas and their zonation is crucial for reducing damage caused by flooding. There are several methodologies suggested by different researchers for the development of flood susceptibility maps around the world (Tehrany et al. 2015b; Chen et al. 2020). Remote Sensing and GIS applications provide powerful tools for predicting and analysing many multidimensional incidents like flooding, which are influenced and controlled by multiple factors (Arabameri et al. 2020a, b; Arabameri et al. 2019). But machine learning models, primarily ensemble of single algorithms with each other or with some other statistical techniques, gives a more remarkable performance, especially in the case of flood susceptibility analysis (Hong et al. 2018; Arabameri et al. 2019; Pham et al. 2017; Tien Bui et al. 2012; Towfiqul Islam et al. 2020; Wang et al. 2021).

For better visualisation of susceptibility areas, the final map was classified into different classes (Pham et al. 2017; Tehrany et al. 2014b). There are different techniques of reclassification, such as standard deviation, geometric intervals, equal intervals, and quantiles (Francis et al. 2015). All these methods show different results, and therefore it is crucial to analyse which methods best suit the particular study (Tehrany et al. 2015a). The equal intervals are suitable for the data with standard data distribution; natural breaks better suit certain jumps (Tehrany et al. 2015b). The geometric intervals are suitable for not normally distributed data, because it reduces variance (Russell et al. 2012). However, the literature shows that the quantile method performs best in the

---

Fig. 9 Sensitivity analysis results from the Jackknife test: a Logistic Regression (LR), b support vector machine (SVM), c Multi-Layer Perceptron (MLP), and d Ensemble
case of susceptibility maps (Tehrany et al. 2015b; Chapi et al. 2017). Therefore, final maps were generated using the quantile method of classification.

In all the maps, it was observed that the high susceptibility areas were located on the southern side, which is the low elevation zone. Also, the results of sensitivity analysis showed that Landuse, elevation, rainfall and NDVI are major contributing factors. Similar results were obtained by some previous research (Arabameri et al., 2020a, b; Santos et al., 2019; Zhao et al., 2018). Most of the settlement area of Karachi is in the southern part Raza et al. (2019), due to which there is very few barren and green areas which result in less seepage of water. This area also receives the highest rainfall compared to the other areas. Along with intense settlements, which reduces seepage, this area faces the problem of the poor drainage system, clogged main holes, and nullah’s heaped with garbage, due to which low laying areas of Karachi is always at the risk of flooding Arif Hasan (2020).

In comparison, the Northern sides, which are mountainous regions of Karachi, fell in the lowest susceptibility zone. The comparison of susceptibility maps showed that the ensemble model calculated the susceptibility areas more precisely and accurately as compared to single machine learning models (Chapi et al. 2017; Arabameri et al. 2020a, b; Prasad et al. 2021).

The precise and accurate framework for flood modelling is a fundamental task because unpredicted floods are responsible for damage in terms of human and infrastructure damage and huge economic loss (Tehrany et al. 2013; Youssef et al. 2016a, b). Therefore, the precision of the proposed model was calculated on both training and validation datasets. Much previous research has shown that statistical metrics, for example, False Positive Rate (FP), False Negative (FN), True Positive (TP) and True Negative (TN), Receiver Operating Curve (ROC), is the best statistical metrics, for example, False Positive Rate (FP), True Positive (TP) and True Negative (TN). The comparison showed that the accuracy of the ensemble model was 99%, which is almost 5%, 2%, and 1% more than LR, SVM, and MLP, respectively. In the case of the ROC curve ensemble model, it gave way better results than single models, which shows that the total number of true positive and true negative cases is more in the ensemble model compared to individual models. This increased and better performance of ensemble and hybrid models are confirmed by using these statistical measures in many previous studies (Chapi et al. 2017; Arabameri et al. 2020a, b; Prasad et al. 2021).

Also, this ensemble model gave a much better performance than previously applied ensemble techniques; in the case of Chapi et al. (2017) bagging ensemble was applied with the logistic model tree, which gave an accuracy of 95%. Similarly, the ensemble approach using a weighted average by Choubin et al. 2019 showed an AUC of 0.91, Towfiqul Islam et al. 2020., used bagging ensemble, which showed RMSE 0.189.

### 6 Conclusion

For flood susceptibility mapping, an ensemble model (LR-SVM-MLP) was used, which was formed by a stacking ensemble of LR, SVM, and MLP. This model was then compared with its benchmarked individual models, which showed that the ensemble model performed best and has higher reliability than all other models.

This research contributes several ways; for example, there is no previous record of any research done on this region for flood forecasting using machine learning models. No previous research has been conducted to evaluate the conditioning or influencing factor contributing to this region’s flood, although this region suffers floods almost every year or every second year. This research gives some of the best models which can be used individually to predict flood in this region, but the proposed ensemble model outweigh all other models, and it highlights high and very high flood-prone regions in the watershed.

The main limitation of this study is that it did not consider the sanitary condition of Karachi, as blockage of nullah’s, is one of the significant reasons for Karachi’s flood. As future work, this research can be enhanced if these models are combined with 2D and 3D modelling systems for better and real-world visualization. The model’s effectiveness can be strengthened if it combines with other models that can tell lag time as lag time is one of the most critical factors in the case of flash flood monitoring and prediction.

Field-based surveys are costly and time-consuming work, while the produced maps have fine details of areas; therefore, they can help management, policymakers, government, and other relevant authorities to provide better flood prevention measures and to mitigate any damage that can be caused by the flood.

### Funding

This work was funded by the National Key Research and Development Program (2018YFC1506506), the Frontier Project of the Applied Foundation of Wuhan (2019020701011502), the Key Research and Development Program of Jiangxi Province (20201BBG71002), and the LIESMARS Special Research Funding.

### Availability of data and material

Not Applicable.

### Code availability

Not Applicable.
References

Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejad H, 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. https://doi.org/10.1080/10106049.2018.1474276

Althuwaynee OF, Pradhan B, Park HJ, Lee JH (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

Arabameri A, Rezaei K, Cerdi A, Conoscenti C, Kalantari Z (2019) A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. Sci Total Environ 660:443–458. https://doi.org/10.1016/j.scitotenv.2019.01.021

Arabameri A, Chen W, Loche M, Zhao X, Li Y, Lombardo L, Cerda A, Pradhan B, Bui DT (2020a) Comparison of machine learning models for gully erosion susceptibility mapping. Geoscient Front 11:1609–1620. https://doi.org/10.1016/j.gsf.2019.11.009

Arabameri A, Saha S, Mukherjee K, Blaschke T, Chen W, Ngo PTT, Band SS (2020b) Modeling spatial flood using novel ensemble artificial intelligence approaches in northern Iran. Remote Sens 12:1–30. https://doi.org/10.3390/rs12030423

Arif Hasan (2020) Urban flooding: the case of Karachi. International Institute for Environment and Development. https://www.iied.org/urban-flooding-case-karachi. Accessed 26 December 2021

Bandos AI, Guo B, Gur D (2017) Jackknife variance of the partial area under the empirical receiver operating characteristic curve. Stat Methods Med Res 26(2):525–541

Beven KJ, Kirkby MJ (1979) A physically based, variable contribution model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). J Hydrol. 540:317–330. https://doi.org/10.1016/j.jhydrol.2016.06.027

Bui DT, Ngo P-TT, Pham TD, Jaafari A, Minh NQ, Hoa PV, Samui P (2019) A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. CATENA 179:184–196. https://doi.org/10.1016/j.catena.2019.04.009

Çelik HE, Coskun G, Giczegolu HG, Akgıralıoğlu N, Aydin A, Esin AI (2012) The analysis of 2004 flood on Kozdere Stream in Istanbul. Nat Hazards 63:461–477. https://doi.org/10.1007/s11069-012-0165-x

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci Model Dev 7:1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chapi K, Singh VP, Shrzardi A, Shahabi H, Bui DT, Pham BT, Khorasvi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ Model Softw 95:229–245. https://doi.org/10.1016/j.envsoft.2017.06.012

Chen FW, Liu CW (2012) Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy and Water Environ 10(3):209–222. https://doi.org/10.1007/s10333-012-0319-1

Chen W, Pourghasemi HR, Zhao Z (2017) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto Int 32:367–385. https://doi.org/10.1080/10106049.2016.1140824

Chen W, Zhao X, Tsangaratos P, Shahabi H, Ilia I, Xue W, Wang X, Bin AB (2020) Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. J Hydrol 583:124602. https://doi.org/10.1016/j.jhydrol.2020.124602

Choi IH, Koo J, Bin JW, Son JAB, Yoon DY, YG, Oh TK, (2020) Damage evaluation of porcelain insulators with 154 kV transmission lines by various support vector machine (SVM) and ensemble methods using frequency response data. Appl Sci. https://doi.org/10.3390/app10010084

Choubin B, Moradi E, Golshan M, Adamowski J, Sadiei-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. https://doi.org/10.1016/j.scitotenv.2018.10.064

Costache R, Pham QB, Avand M, Linh NTT, Vojtek M, Vojteková J, Lee S, Khoi DN, Nhi PTT, Dung TD (2020) Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. J Environ Manag 265: 110485.

Das I, Sahoo S, van Westen C, Stein A, Hack R (2010) Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). Geomorphology 114:627–637. https://doi.org/10.1016/j.geomorph.2009.09.023

DeLeo JM (1993) Receiver operating characteristic laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty. Proc. - 2nd Int. Symp Uncertain Model Anal ISUMA 1993:318–325. https://doi.org/10.1109/ISUMA.1993.366790

Dodangeh E, Choubin B, Eidigir AN, 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. https://doi.org/10.1016/j.scitotenv.2019.135983

Dou J, Yunus AP, Bui DT et al (2020) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed. Japan Landslides 17:641–658. https://doi.org/10.1007/s10346-019-01286-5
Yang L, Cervone G (2019) Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event. Soft Comput. https://doi.org/10.1007/s00500-019-03878-8

Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci 61(4):821–836. https://doi.org/10.1016/j.scitotenv.2018.10.064

Youssef AM, Pradhan B, Hassan AM (2011) Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. Environ Earth Sci 62:611–623. https://doi.org/10.1007/s12665-010-0551-1

Youssef AM, Pradhan B, Sefry SA (2016a) Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. Environ Earth Sci 75:1–16. https://doi.org/10.1007/s12665-015-4830-8

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016b) 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. https://doi.org/10.1007/s10346-015-0614-1

Zhao G, Pang B, Xu Z, Yue J, Tu T (2018) Mapping flood susceptibility in mountainous areas on a national scale in China. Sci Total Environ 615:1133–1142. https://doi.org/10.1016/j.scitotenv.2017.10.037

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.