Article

Computational Machine Learning Approach for Flood Susceptibility Assessment Integrated with Remote Sensing and GIS Techniques from Jeddah, Saudi Arabia

Ahmed M. Al-Areeq [1] **, S. I. Abba** [1,*]**, Mohamed A. Yassin** [1]**, Mohammed Benaafi** [1]**, Mustafa Ghaleb** [2]
and Isam H. Aljundi [1,3]

1 Interdisciplinary Research Center for Membranes and Water Security (IRC-MWS), King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia
2 Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS), King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia
3 Department of Chemical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
* Correspondence: sani.abba@kfupm.edu.sa

Abstract: Floods, one of the most common natural hazards globally, are challenging to anticipate and estimate accurately. This study aims to demonstrate the predictive ability of four ensemble algorithms for assessing flood risk. Bagging ensemble (BE), logistic model tree (LT), kernel support vector machine (k-SVM), and k-nearest neighbour (KNN) are the four algorithms used in this study for flood zoning in Jeddah City, Saudi Arabia. The 141 flood locations have been identified in the research area based on the interpretation of aerial photos, historical data, Google Earth, and field surveys. For this purpose, 14 continuous factors and different categorical are identified to examine their effect on flooding in the study area. The dependency analysis (DA) was used to analyse the strength of the predictors. The study comprises two different input variables combination (C1 and C2) based on the features sensitivity selection. The under-the-receiver operating characteristic curve (AUC) and root mean square error (RMSE) were utilised to determine the accuracy of a good forecast. The validation findings showed that BE-C1 performed best in terms of precision, accuracy, AUC, and specificity, as well as the lowest error (RMSE). The performance skills of the overall models proved reliable with a range of AUC (89–97%). The study can also be beneficial in flash flood forecasts and warning activity developed by the Jeddah flood disaster in Saudi Arabia.

Keywords: artificial intelligence; flood; Saudi Arabia; machine learning; GIS; remote sensing

1. Introduction

Millions of people worldwide are affected by natural disasters every year. Floods are the most severe phenomenon among all types of natural disasters that cause significant losses in human lives and economics. Every year about 200 million people are affected by flood hazards [1]. In the past few decades, the number of such disasters has increased due to climate change [2,3]. Jonkman [3] reported that floods affected 1.4 billion people and killed over 100,000 individuals during the last decade of the twentieth century. Because of their unique qualities, flooding in semi-arid and arid regions is more dangerous than in wet regions.

Studying floods in semi-arid and arid regions is a challenging task. Saudi Arabia (KSA) is considered within the semi-arid regions according to the classification of the World Map of Kopper–Geiger Climate [4]. Many cities in Saudi Arabia suffer from annual floods [5]. The flood of 2009 caused loss of life and economic loss of more than 121 deaths, 20,000 displaced households [6,7], and billions of dollars [7]. The Jeddah flood disaster has worsened due to urban changes, rainfall, climatic changes, and network and watershed

factors. Thus, flood vulnerability research is mandatory to accurately identify flood risk zones to reduce the effect of floods by developing preventive measures [8].

Various studies about flood occurrence probability have been done using different techniques, such as rainfall runoff, pattern classification, and traditional analyses [9]. Rainfall-runoff models (e.g., GSSHA [10], MIKE by DHI [11], HEC-HMS [12], and SWAT [13–15]) are most commonly used to predict the temporal and spatial variation of floods based on establishing a relation between runoff and rainfall. However, they require field observation data to achieve a high accuracy level of prediction [16–18]. Pattern classification used "on-off" classification, which does not require field observation data. In this model, the flood-prone area is classified into non-flood and flood zones based on geo-environmental data and historical floods [19]. Various models have been suggested and proposed based on the remote sensing data to predict floods [20–23], but in mountainous areas, these models could not accurately predict flash floods [24,25]. In traditional analyses, regression models are generated based on the field observation of historical data to forecast discharge [25–28]. However, this kind of model is mostly applied to specific areas.

In order to properly prepare for any natural hazard susceptibility mapping, it is essential to control the risk factors that cause the hazard [29]. Floods are caused by various connected conditions or factors [29,30]. The terrain significantly impacts the surface runoff in terms of direction and rate. According to earlier studies [31,32], elevation is one of the most critical criteria in mapping flood susceptibility. It is, however, inversely associated with flooding; the lower the elevation, the higher the likelihood that a flood might occur. The slope angle determines the trends of physiography and soil moisture patterns [33]. As a result, it is crucial to determine the status of hydrologic settings that affect infiltration rate, runoff, and subsurface drainage [32]. Hydrologic processes such as the direction of frontal precipitation, evapotranspiration, vegetation development, and weathering processes are significantly impacted by the slope aspects, particularly in dry environmental conditions [34]. Another key component in controlling flooding is lithology. The type of geological formation greatly influences the permeability of soils. The porosity will decrease with increasing topsoil fineness, generating more runoff flow [35]. Higher permeability of the upper soil layer on the terrain increased infiltration capacity and decreased runoff [36].

The most important climate parameter influencing hydrological processes and flood risk assessment is precipitation, which includes rainfall and snowfall [37]. The amount and rate of precipitation significantly impact the flood risk. Notably, the crucial factors are the frequency and intensity of rainfall events [38]. Various hydrological responses are significantly impacted by changes in land cover and land use. In this context, several studies have been devoted to evaluating the effects of land use and land cover (LULC) trends at various scales on flooding assessment and management. They reported the crucial role of LULC plays in the runoff rate and volume [39]. Many landscape changes are primarily caused by LULC changes, such as transitions from forest to agriculture, forestry to arable land, rain and/or groundwater-fed farmland to irrigated agriculture, or forest use to urbanised regions [40].

In the last few years, flood susceptibility evaluation research has significantly increased [35,38,41,42]. In recent years, artificial intelligence models have been integrated with GIS and remote sensing techniques to predict the spatial variability of flood susceptibility. Examples of these models include: kernel logistic regression (KLR) [43], evidential belief function and decision trees (DT) [44], support vector machine (SVM) [45,46], deep learning neural network [47], logistic regression [48,49], artificial neural networks (ANN) [49–51], rotation forest [52], WELLSVM [53], Naïve Bayes [54], random forest [55], QUEST and GARP [56], and classification and regression trees (CART) [56]. However, researchers are not consensus about selecting the best-performed models. Accordingly, other studies found that hybrid models can predict flood susceptibility at a high level of accuracy, such as metaheuristic algorithms and neuro-fuzzy systems [57–59], support vector machines with an ensemble of weights-of-evidence [28], bagging ensembles, and logistic model tree [60], the ensemble of multi-criteria decision making [61], SVM, CART, and an ensemble of multi-

variate discriminant analysis [62], swarm optimised neural networks [63], fuzzy rule-based ensembles [9,64], and hybrid Bayesian framework [9].

Similarly, Mosavi et al. [65] provide a detailed review of the performance of machine learning models in flood prediction. It was evidenced that several technical pieces of literature have been published on flood assessment. However, the estimation of flood risk and vulnerability is valuable in preventing loss of life and damages. Computational algorithms such as ML models are a subdivision of artificial intelligence in which the machine learns from machine-readable data and information. It uses data, learns the pattern, and predicts new outcomes. Its popularity is growing because it helps us to understand the trend and provides a solution that can be either a model or a product. Applications of ML algorithms have increased drastically in GIS and remote sensing in recent years. The role of geoinformation technology, for instance, GIS and remote sensing in flood analysis and land use changes, cannot be overlooked. This role further generated several interests of researchers to employ data-driven approaches and sensed data as well as other source data to create the vector or raster inputs.

The application of machine learning to analyse floods in Jeddah city received less attention despite other published studies along the same line. This study explored machine learning models viz: bagging ensemble (BE), logistic model tree (LT), kernel support vector machine (k-SVM), and k-nearest neighbour (KNN) for flood susceptibility mapping and prediction in Jeddah City. For this purpose, dependency analysis (DA) was employed to feature categorization and selection. The major motive behind this study is attributed to the flood scenarios of 2009 and 2011, where more than 113 persons died, in addition to damaged records of buildings, roads, cars, and the loss of several properties.

2. Description of the Study Area and Input Data

2.1. Study Area

Jeddah city, located within three major sub-basins (northern, middle, and southern), which are the main source of flash floods, was selected as a case study (Figure 1). The northern sub-basin includes wadi Daghbaj, wadi Brayman, wadi Muraygh, wadi Quraa, wadi Ghaia, and wadi Um Hablain. The middle sub-basin includes Wadi Mraikh and Wadi Bani Malik. The southern sub-basin includes wadi Qaws, Wadi Wadi Methweb, Asheer, Wadi Al Khomra, and Wadi Ghulail. In 2022, an estimated 4.78 million people occupy Jeddah city [66]. The city is boarded west by the Red Sea and east by mountain chains with a maximum altitude of 675 m. The drainage area, as delineated by 30 m DEM is 1821 km^2. The residential area is in the coastal plain, exposing it to the effects of flash floods from the mountain chains. Figure 1 shows the topography of the Jeddah watershed, which shows two geomorphological units: the coastal plain and the mountains which dominate the city. Although Jeddah is arid, it suffers from flash floods that have hit the city several times. The urban areas were attacked on 25 November 2009, by flash floods causing much damage to infrastructures, buildings, cars, and roads, and about 113 people died. Huge damage was also caused by another event in 2011 [67]. The watershed drains in many neighbourhoods, such as Al-Harazat, King Abdel Aziz University, Al-Haramin Highway, Al-Mesaid, Queza, and Al-Sawaid, which were significantly affected due to the 2009 flash flood event.
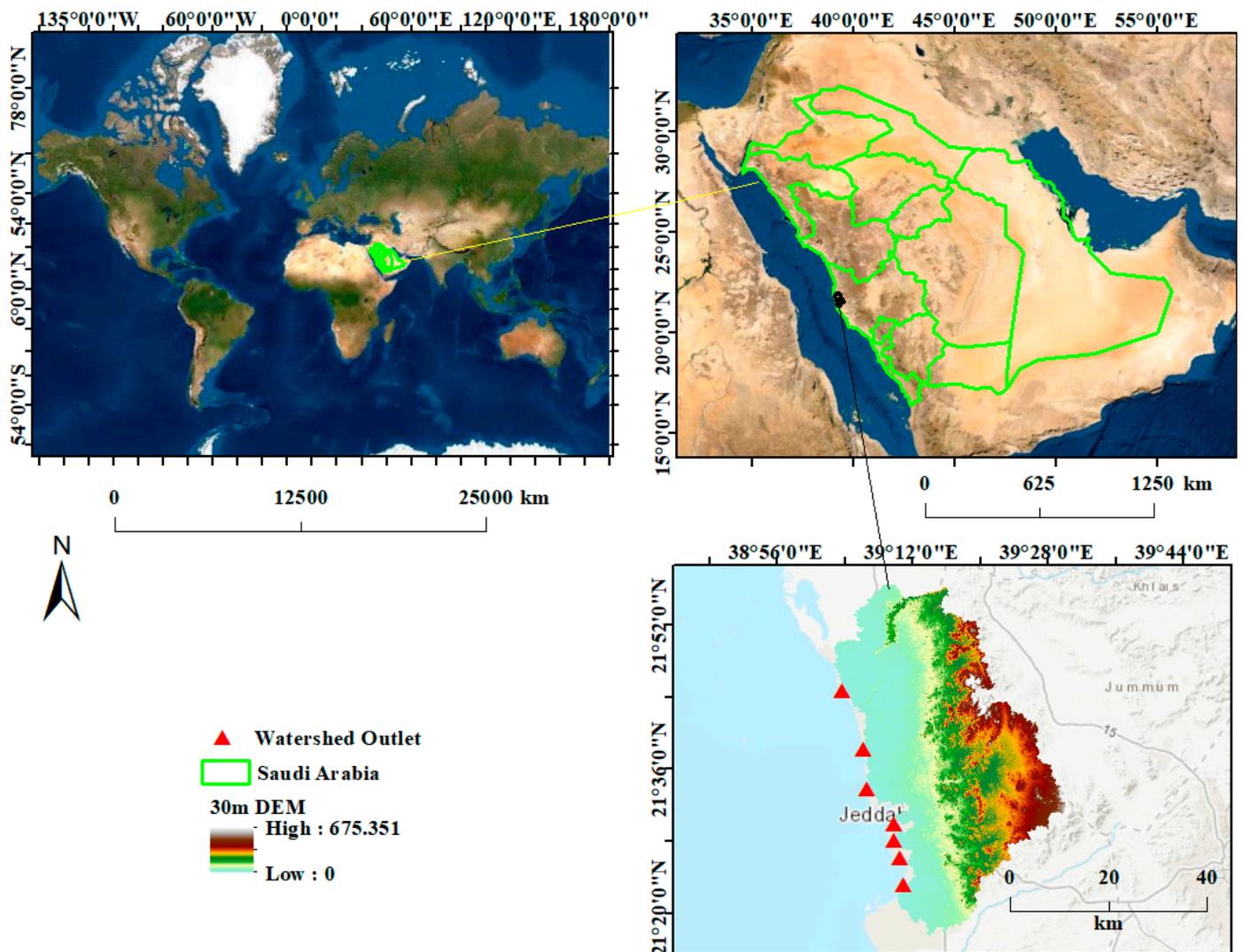
Figure 1. Study area employed in this work.

2.2. Flood Conditioning Factors/Predictors

The study's flooding conditioning factors, including elevation, slope, aspect, topography, lithology, precipitation, land cover, and land use, were taken into account based on the study region and the available data. Hydrology, geography, environment, and anthropology form the four categories of these flood conditioning factors. In this study, the following variables were extracted from 30 m shuttle radar topographic mission (SRTM) to be considered slope angle (SA), topographic position index (TPI), stream power index (SPI), plan curvature (PC), topographic wetness index (TWI), distance river (DR), rainfall (P), lithology, land use (LD), soils, convergence index (CI), flow accumulation (FA), elevation, topographic ruggedness index (TRI), F-NF (flood and non-flood) and aspect (Figure 2).

A major factor in the occurrence of floods is rainfall. It should be mentioned that the vast majority of research pertaining to the assessment of flood susceptibility employed the yearly average rainfall figure. All rainfall values were interpolated using the spline approach. When there are only a few data points, as there are in this study, this approach is advised. Due to its primary connection to the fluctuation of soil moisture, the aspect was discovered to be a predictor of floods. A single flat zone and the ninth divisions of north, north-east, east, south-east, south-west, west, and north-west and north were created using the values of the aspect raster. The elevation of the surrounding cells was taken into consideration by the topographic position index (TPI), which successfully distinguishes a

cell from them (Jenness, 2000). In the current study, the following classes for the TPI map were defined: (-79.38)–(-15.1); (-15.1)–(-3.72); (-3.72)–5.36; 5.36–22.76; 22.76–113.56. Another morphometric indicator that will be employed as a flood prediction is the stream power index (SPI). When determining water's values, erosive force, and transport ability are taken into account. The following expert judgment-based classifications for SPI maps were created: 0–50; 50–100; 100–400; 400–1000; >1000.
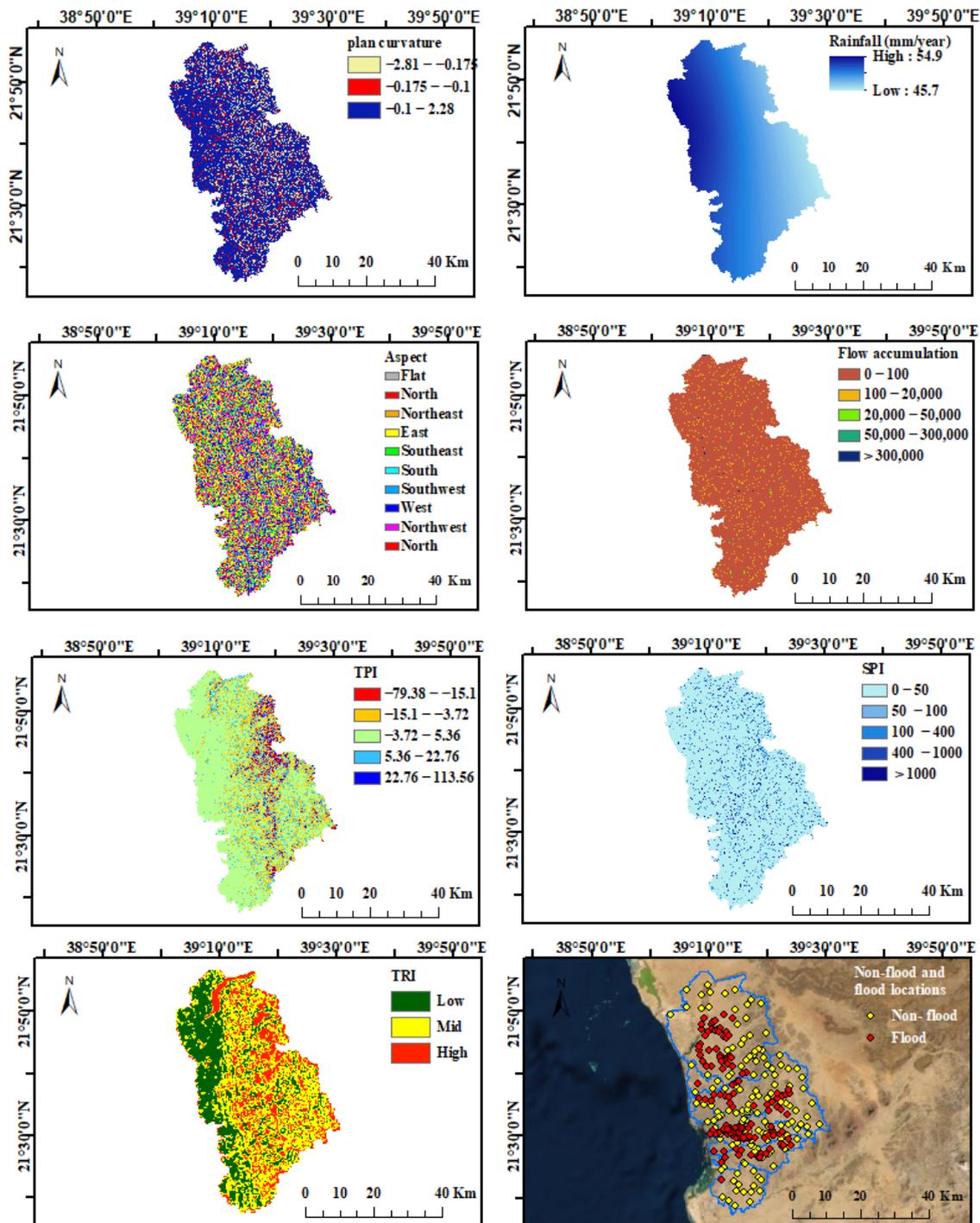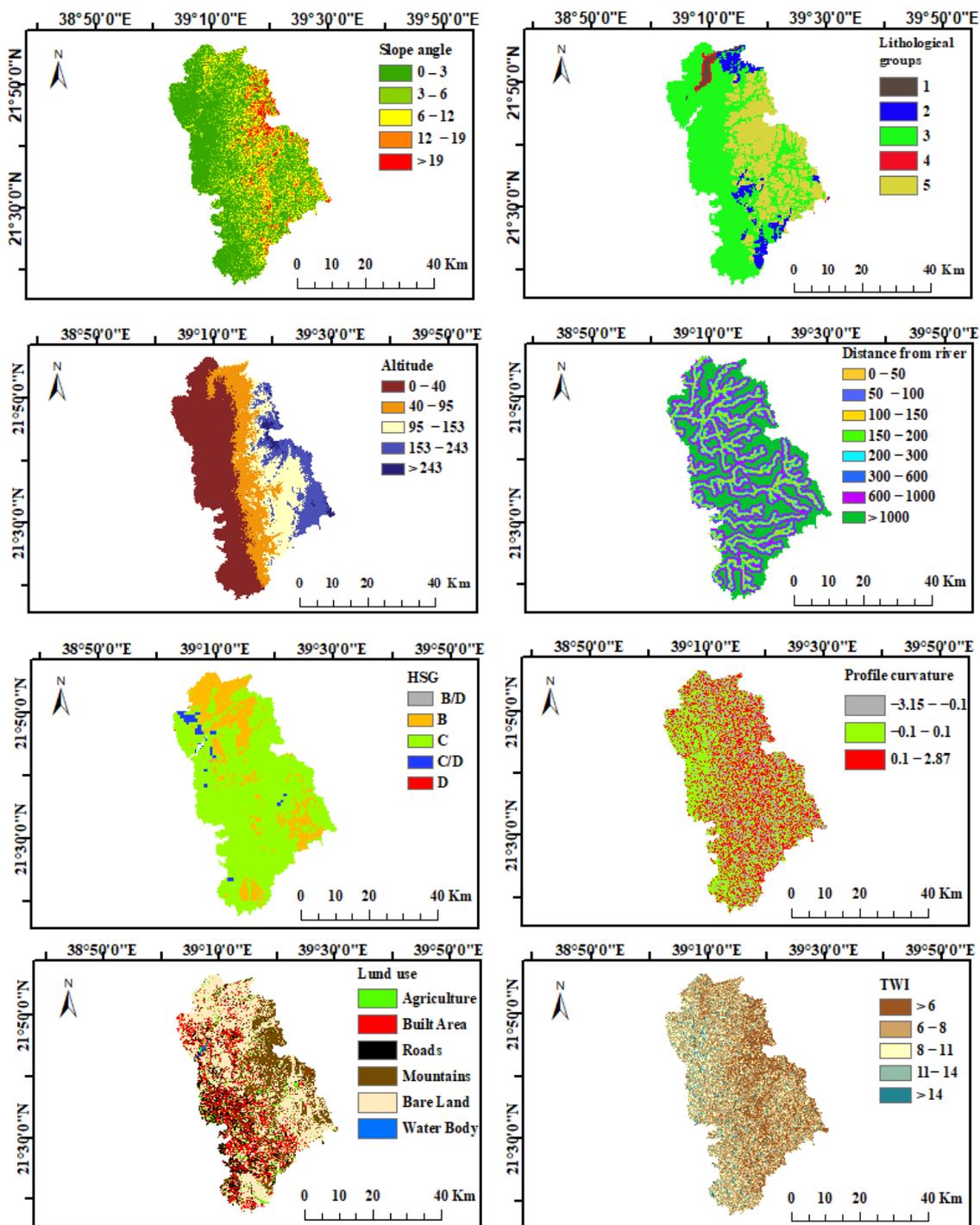

Figure 2. *Cont.*

Figure 2. Flood condition factors based on the study area.

Similarly, the slope angle is a crucial morphometric component that significantly affects the flooding process. It is common knowledge that a level location is more vulnerable to floods than an area with a steep slope that promotes the manifestation of surface runoff. To construct the slope angle map, the slope angle was divided into the following ranges: <3; 3–7; 8–15; 16–25; and >25. Lithology primarily regulates the penetration of water due to the permeability of the rock, which in turn impacts the flooding phenomenon. In the research region, a total of five lithological classifications were discovered (igneous extrusive rocks, igneous intrusive rocks, Polylithologic rocks, sedimentary rocks, and sedimentary surficial deposits). As a result of the fact that flood occurrences are more

likely to occur in low-lying locations, elevation is a second morphometric element that is extremely important in determining flood vulnerability. In order to create the elevation map for the case study at hand, the following 8 altitude classes were used: <50 m; 50–100 m; 100–200 m; 200–300 m; 300–400 m; 400–500 m; 500–600 m; >600 m. The Hydrological Soil Group, a flood control parameter, has a significant impact on water infiltration. Particularly, the soil texture affects infiltration directly because it affects hydraulic conductivity. Three hydrological soil categories are located inside the watershed of Jeddah. The vertical plan's slope direction affects the profile's curvature. Values greater than 0 imply enhanced surface runoff, whereas negative values suggest decelerating surface runoff. The profile curvature map was made using the following three classes: −3.15 to −0.1; −0.1 to 0.1; and 0.1 to 2.87. Surface runoff and water storage processes are significantly influenced by land use, a component in flood prediction. The Jeddah catchment is divided into the following six land-use categories: agricultural zones, built areas, roads, mountains, bare lands, and water bodies. The topographic wetness index (TWI) is determined using the slope angle values and the particular catchment region. This indicator emphasises how geography affects the phenomena of water accumulation. The natural breaks technique was used to create the classes listed below in order to design the TWI map: 3.5–6.58, 6.58–8.6, 8.6–11.1, 11.1–14.66, and 14.66–28.05.

2.3. Flood Locations Inventory

Understanding the flooding inventory is very essential for successful flood management and mitigations. For computing the flood susceptibility, the past event locations were used to generate the crucial input variables. According to Costache et al. [41], Costache et al. [68], Sammen et al. [69], the probability of floods could be attributed to areas with the same features and other characteristics. The source of flood inventory differs from one geographical location to another but generally can be from past technical, scientific work, government achieves, newspapers, field surveys, or recently emerging technologies. The current study developed the flood inventory map based on information from past published articles, aerial photos, historical data, Google Earth, and news from the government database in Saudi Arabia. Hence, a total of 282 flood events were used for the occurred flood in the prone zone of Saudi Arabia (Figure 2). The points were proved as the most important flood zone locations; thus, we considered these points to reflect the complex problems of the Jeddah region. In several works of literature, for example [60,61,70,71], flood susceptibility-dependent variables have been used as points for flood locations.

3. Background of Methods Used

The complicated phenomena of natural events such as floods proved to be due to several factors, including climate change and human activities. This study proposes different machine learning algorithms integrated with remote sensing and GIS to control this phenomenon. The flood pattern falls into binary classification; hence the susceptibility mapping procedure includes the non-flood's location (141). As a tradition of reliable models, the training and testing performance was validated using several indicators. The total sample of the data comprises both flood and non-flood samples with a ratio of 70% and 30% for the training and testing phase, respectively. To ensure a subjective sampling process, the external validation of random sampling for all the locations was conducted using the ArcGIS 10.8 software. The overall proposed methodology is presented in Figure 3.
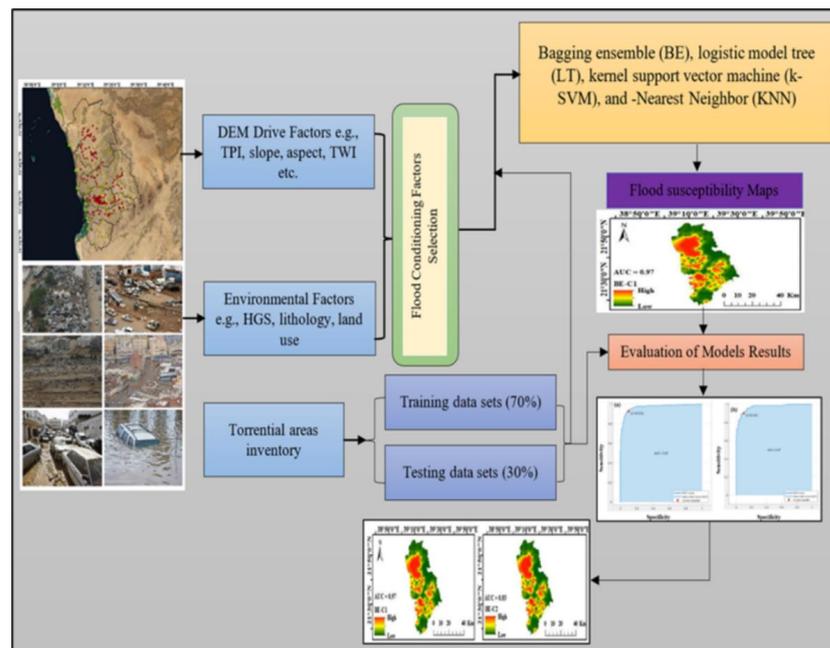
Figure 3. The overall proposed methodology.

3.1. Bagging Ensemble (BE)

For more stable, reliable, and accurate models, a technique known as "bagging" or "Bootstrap Aggregating" is used [72,73] (Figure 4). One reliable ensemble learning technique used to resample the training dataset is bagging. The raw data samples that make up the multiple sets of training data are bootstrapped in the first stage. These training datasets are used to construct a variety of models. The continual training processes for datasets and numerous models produce predictions. The basic idea behind the bagging technique is simple. Multiple models are developed to characterise the relationship between the input and output variables instead of a single model making appropriate predictions for the actual data. Then, several models are linked to create a single output using the weighted average in the bagged algorithm [74,75]. The potential uncertainties in the modelling process can be successfully reduced with this tactic. As demonstrated by earlier publications, bagging is a good option for ensemble modelling of various environmental problems [76].
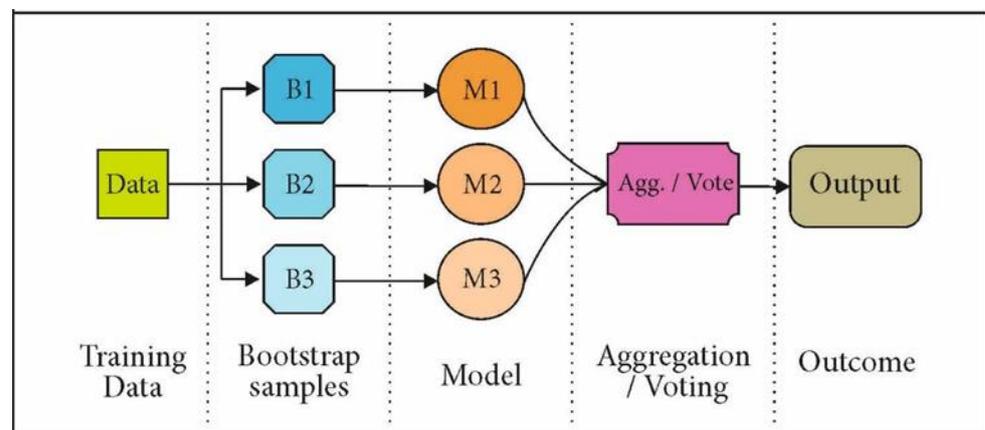

Figure 4. Bagging ensemble processes flow.

3.2. Logistic Model Tree

The logistic model tree is a classification model that combines logistic regression (LR) with decision tree learning techniques [77,78]. While the efficiency of categorization is not greatly affected, this exploratory strategy significantly improves time. The key advantages

of LMT models are their rapid construction and simple interpretation. The LogitBoost algorithm is used in the logistic variation to create an LR model at each node in the tree, and the CART algorithm is used to prune the tree. Information gain is employed for splitting [79]. To avoid overfitting with training data, the LMT finds several LogitBoost rounds using cross-validation. Generally, LMT constructs compact tree structures using high-tech, low-cost pruning techniques [80,81].

3.3. Kernel SVM Algorithms

The k-SVM model (Figure 5) was developed based on the concept of a support vector machine (SVM), which is generally used in solving problems through regression and classification approach [82–86]. SVR is an established computational technique with various merits, such as good noise-tolerating, superior generalization ability, and high learning speed [87]. Generally, the input variables from the datasets were mapped into a compacted-spatial elements filter architecture via a nonlinear kernel operator using the SVM [87–89]. This regression technique can convert a complex process into a simple one via understanding the learning complexity of the interaction between the predictors and responses [90,91]. Different kernel functions have been used to solve chaotic problems in science and engineering, for example, linear, multinomial, and radial basis function (RBF). In the current study, RBF was used owing to its robustness to handle the complex nonlinear process.
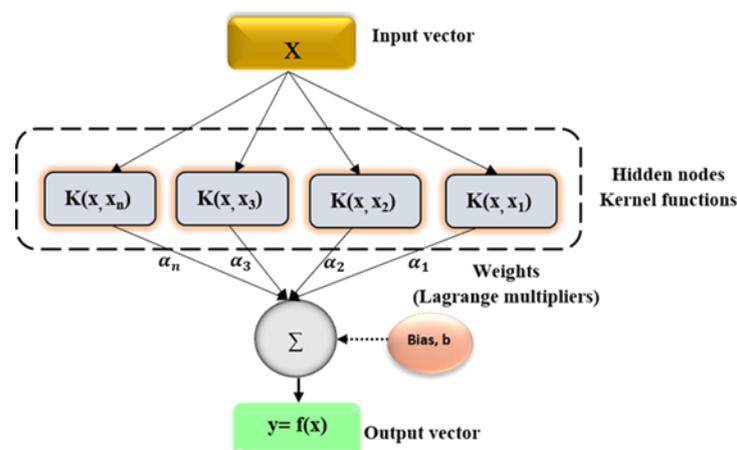

Figure 5. Structure of the k-SVM model.

3.4. K-Nearest Neighbor (KNN)

As non-parametric soft computing supervised learning algorithm, KNN is employed in both problems related to classification and regression model that classifies a viewpoint in n-spatial space [92]. However, the k-adjoining bordering features are used in the exercise calibration, which relies mainly on distance pattern classification [93]. It is based on the concept that elements which the same geographic coordinate or site will eventually occupy the same attribute and characteristic if they are situated near each other. This algorithm has been promising in forecasting problems such as flooding with the voting process nature to spatial objects. As indicated in several works, for instance, distance has been defined in [68,94,95].

4. Model Validation Techniques

For all the developed models, the results from the proposed approach were validated using several performance matrices such as sensitivity, specificity, precision, and accuracy. According to [96,97], performance indices are considered significant if the spatial correlation exists between the measured flood and non-flood and predicted flood susceptible zone.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where TP = true positive, TN = true negative, FP = false positive, FN = false negative. Similarly, another common index called the receiver operating characteristic (ROC) curve was used in the analysis. As the most frequently employed, the ROC defines the reliability of the predictive models by considering the area under the curve (AUC). In addition, root mean squared error (RMSE) and mean absolute error (MAE) were also applied to compute the flood susceptibility mapping. These two indices have been employed in several scientific works.

$$\text{AUC} = \frac{(\sum \text{TP} + \sum \text{TN})}{(\text{P} + \text{N})} \tag{5}$$

where TP is true positive, TN is true negative, P is the total number of pixels with torrential phenomena, and N is the total number of pixels without torrential phenomena

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(X_{\text{predicted}} - X_{\text{actual}} \right)^2} \tag{6}$$

$$\text{MAE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left| X_{\text{predicted}} - X_{\text{actual}} \right|} \tag{7}$$

where n is the total samples in the training or testing phase, the predicted value is $X_{\text{predicted}}$, the observed value is X_{actual} from the flood susceptibility model.

5. Results and Discussion

Due to the different flood predictors used in this study, it is evident that conditioning factors affect flood variability. The strength of these predictors needs to be understood again the target flood occurrence. Different methods to analyse the strength of the predictors have been utilised in various technical works of literature, for example, the IGR method, average merit (AM), and dependency analysis (DA). The dependency analysis was employed in this study due to its popularity in science and engineering and uniqueness novelty in risk and flooding problems. Figure 6 shows the DA with respect to another flood-predicted occurrence. The analysis indicates that TWI, FA, SPI, and P are directly related to the target parameters, while all other variables, including TPI, TRI, DEM, aspect, PC, slope, lithology, soils, LU, and DR, were correlated indirectly.

The five most strongly factors affecting the target variables irrespective of their direction according to the DA were DR (−0.79007), TWI (0.5619), slope (−0.5114), DEM (−0.4563), and lithology (−0.4145). This justification has been reported in some of the literature; however, historical investigation of flood data depicted that the common occurrence of floods is attributed to the low slopes and distance to the rivers, this is in line with the study of [60]. The factors that notably affected the overall flood analysis and their absolute weight were ranked in Figure 7. As stated above, the study comprises two different input variables combination (C1 and C2) based on the features sensitivity selection. The C1 and C2 define the dominancy towards target variables for instance C1 (DR + TWI + S lope + DEM + Lithology + TRI + TPI), and C1 (FA + P + SPI + PC + LU + Soil + Aspect). It is quite important to note that formulating these variables is quite complex hence feature optimization is crucial to ease the complex process.
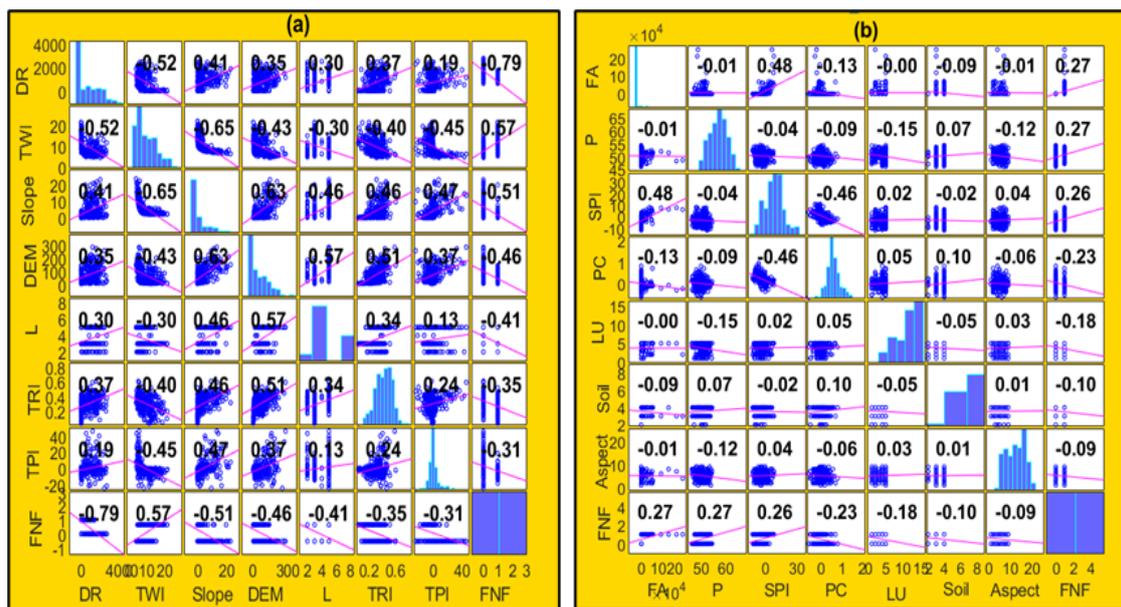
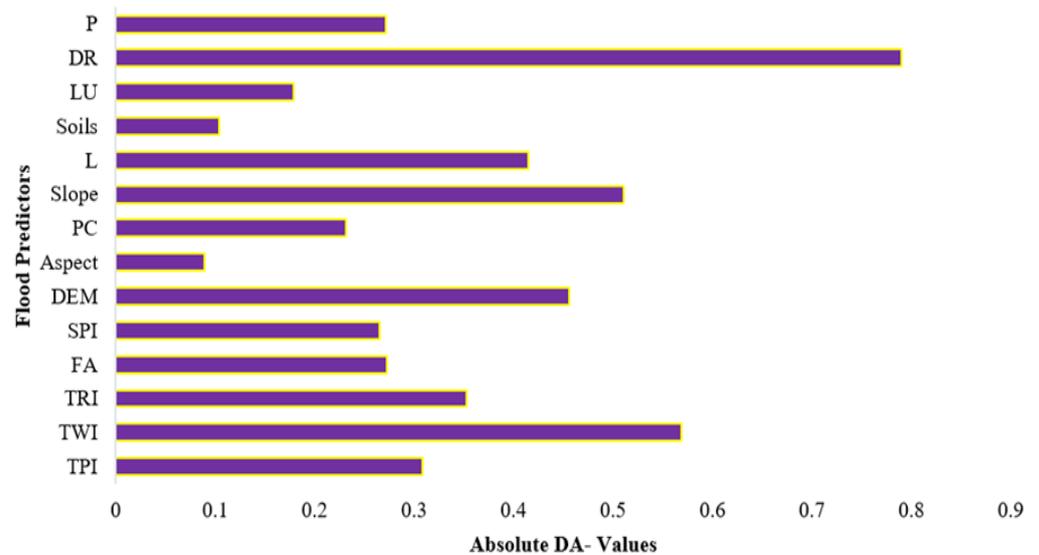**Figure 6.** Dependency analysis for (**a**) Combo 1 (**b**) Combo 2.

Figure 7. Flash-flood conditioning factors vs. predictive strength.

The results for this study were constructed using a bagging ensemble (BE), logistic tree (LT), kernel support vector machine (k-SVM), and k-nearest neighbour (KNN), and the training and validation dataset was divided into 70% and 30%, respectively. In addition, 10-k-fold cross-validation was applied for all the combinations. The validation phase results were recorded for the combination 1 and 2, as presented in Tables 1 and 2, respectively. For all the classifiers, the optimal parameters such as the number of iterations, prediction speed (obs/s) and training time (s) were attained, and the best stooping criteria were selected. Different performance indicators were used to estimate flood susceptibility in this study. The under-the-receiver operating characteristic curve (AUC) was utilised to determine the accuracy of a good forecast; for example, if AUC is ranged between 0.5 and 1, where 0.5 implies a poor estimate, while an excellent forecast is when AUC approaches or is equal to 1. Figure 8a–d indicates the efficiency performance of BE, LT, k-SVM, and KNN algorithms during the validation phase.

Table 1. Performance of the models based on the calibration dataset.

Validation Phase	BE-C1	LT-C1	k-SVM-C1	KNN-C1
True positive (TP)	92.9	90.1	93.6	88.7
True negative (TN)	90.1	90.1	83.7	88.7
False positive (FP)	9.9	9.9	16.3	11.3
False negative (FN)	7.1	9.9	6.4	11.3
Precision	0.9037	0.901	0.85168	0.887
Sensitivity	0.929	0.901	0.936	0.887
Specificity	0.901	0.901	0.837	0.887
Accuracy	0.915	0.901	0.8865	0.887
RMSE	0.00501	0.017006	0.01504	0.005
AUC	0.97	0.97	0.93	0.89
Prediction speed (obs/s)	1700	4100	4300	3300
Training time (s)	5.4626	3.8231	2.9381	2.2905

Table 2. Performance of the models based on the validation dataset.

Validation Phase	BE-C2	LT-C2	k-SVM-C2	KNN-C2
True positive (TP)	73	76.6	67.4	65.2
True negative (TN)	75.9	63.1	67.4	65.2
False positive (FP)	42.1	36.9	32.6	34.8
False negative (FN)	27	23.4	32.6	34.8
Precision	0.634231	0.67489	0.674	0.652
Sensitivity	0.73	0.766	0.674	0.652
Specificity	0.390738	0.38806	0.5	0.5
Accuracy	0.683028	0.6985	0.674	0.652
RMSE	0.007092	0.031518	0.025324	0.005015
AUC	0.83	0.80	0.75	0.65
Prediction speed (obs/s)	690	1800	1800	1700
Training time (s)	14.376	11.8	8.9565	5.2392

The best iteration algorithms were chosen and reported, as indicated in Table 1. From the Table results, BE-C1 attained the best results based on the sensitivity (0.929) and specificity (0.901). The quantitative comparison regarding the accuracy shows that BE-C1 (0.915), LT-C1 (0.901), k-SVM-C1 (0.886), and KNN-C1 (0.887) demonstrated a promising ability of ensemble boosting. For combination 1 (C1), the validation results depicted that BE-C1 had the highest performance in terms of accuracy, precision, AUC, and specificity and the lowest error (RMSE). This indicated the probability of the machine ensemble models in classifying the flood. The accuracy of the models' combination (C2) is also presented in Table 2; this combination includes an absolute strength ranging from 23–35% with the target variables. The highest sensitivity was attributed to the LT-C2 (0.763) followed by BE-C2 (0.7), k-SVM (0.67), and KNN-C2 (0.65) model. The C2 shows marginal convergence and accuracy owing to the less correlated strength with the target variables. The numerical comparison for the two-combination revealed that k-SVM- C1 models had promising sensitivity, with more than 93% of the flood pixels being acceptably classified into the classes of flooding. Similarly, BE-C1 models were found to have the highest specificity, indicating that 90% of the validation datasets were classified as non-floods (Figure 9).
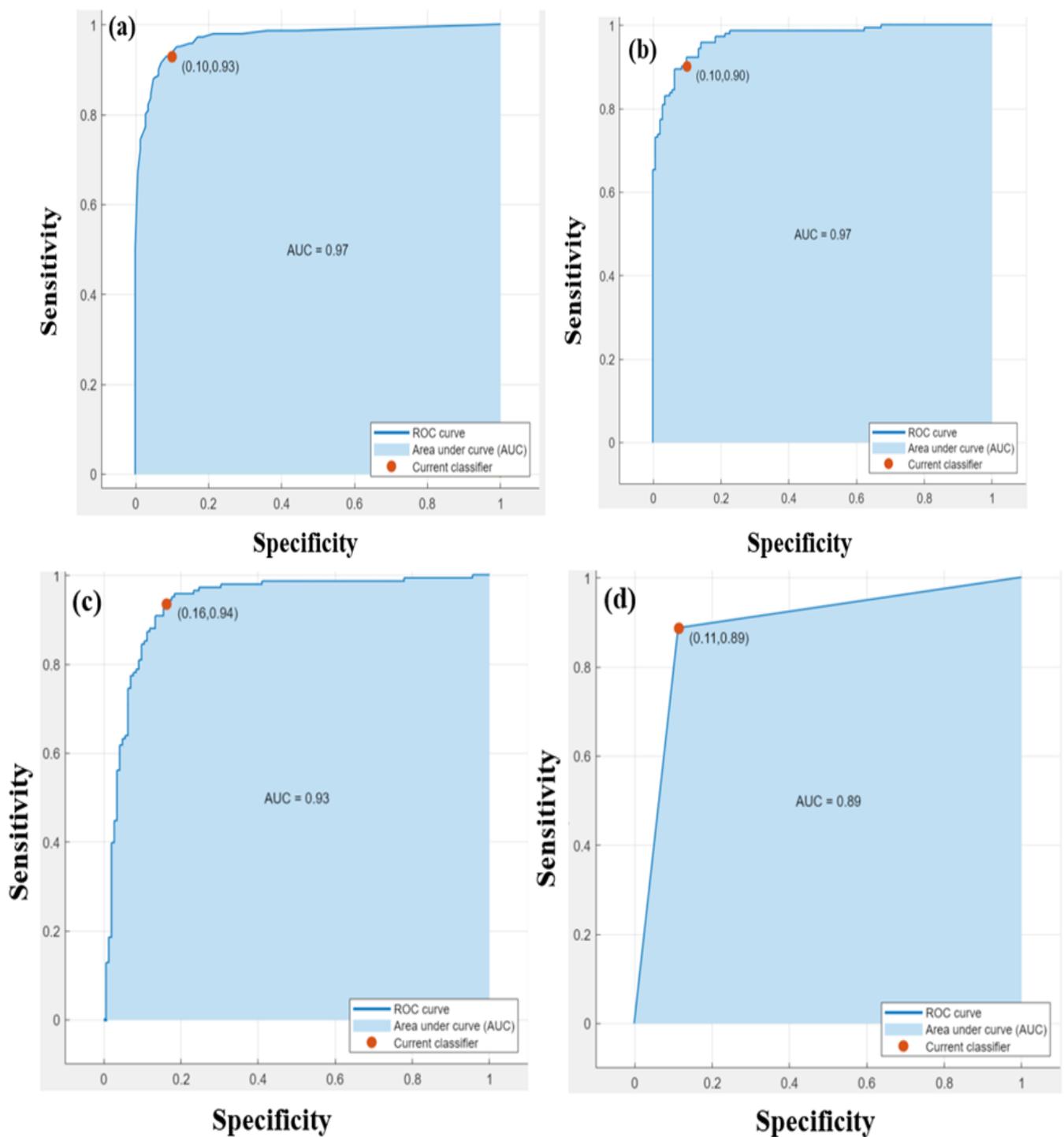
Figure 8. Performance of the model for the spatial prediction of flash floods using the ROC curve technique and AU for (**a**) EB-C1, (**b**) LT-C1, (**c**) k-SVM (**d**) KNN-C2 algorithms.

Besides the specificity and sensitivity criteria, the statistical comparisons were generated using the RMSE, as demonstrated in Tables 1 and 2. The results showed that the RMSE for the best combination is 0.00501, 0.0315, 0.01504, and 0.0050 for EB-C1, LT-C1, k-SVM-C1, and KNN-C1, respectively. The overall ranking of RMSE can be indicated in Figure 10. The values of RMSE for EB-C1 depicted the perfect agreement between the observed and simulated values; hence the estimated probability of susceptibility was attained.
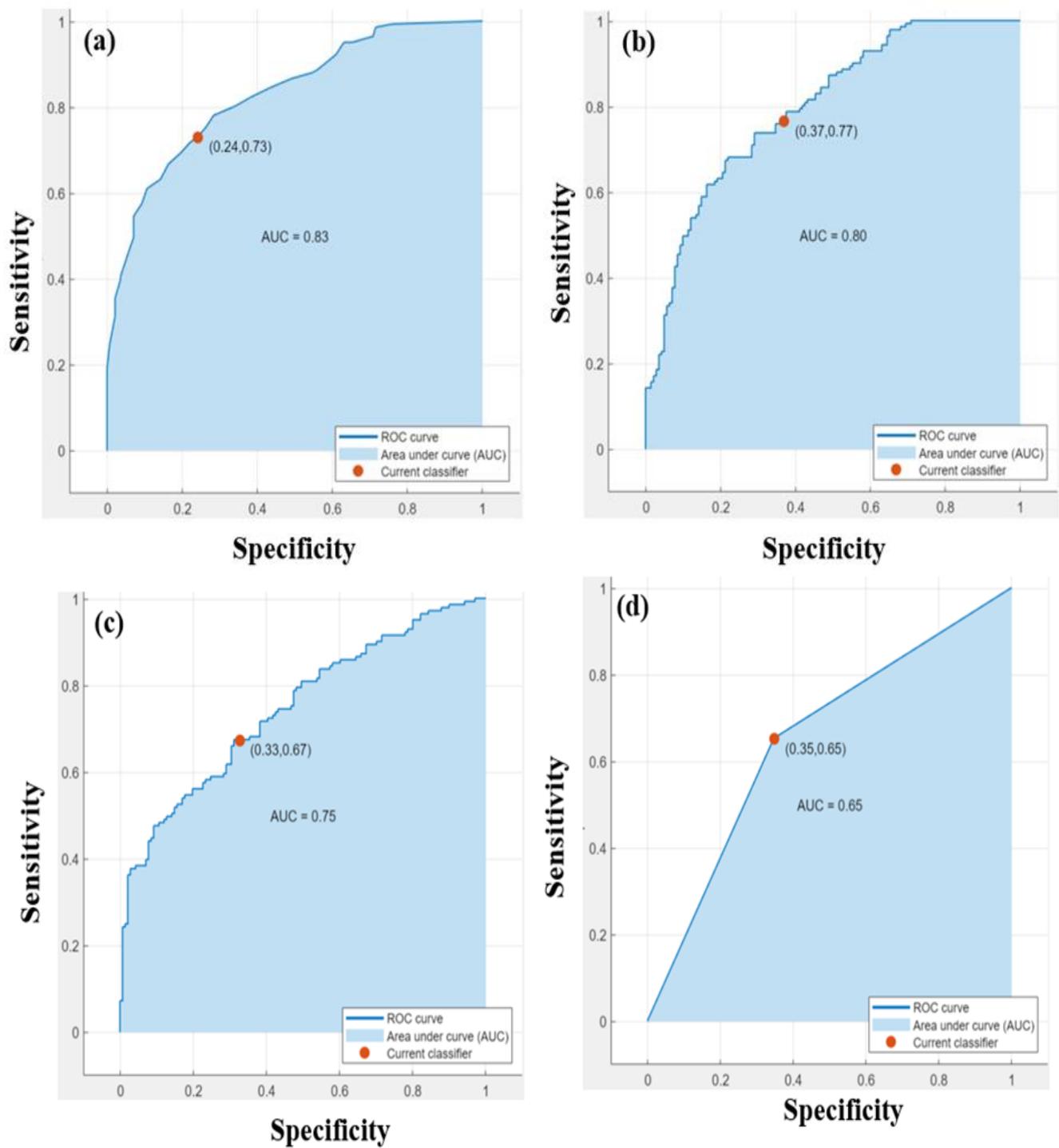
Figure 9. Performance of the model for the spatial prediction of flash floods using the ROC curve technique and AU for (**a**) EB-C2, (**b**) LT-C2, (**c**) k-SVM, (**d**) KNN-C2 algorithms.
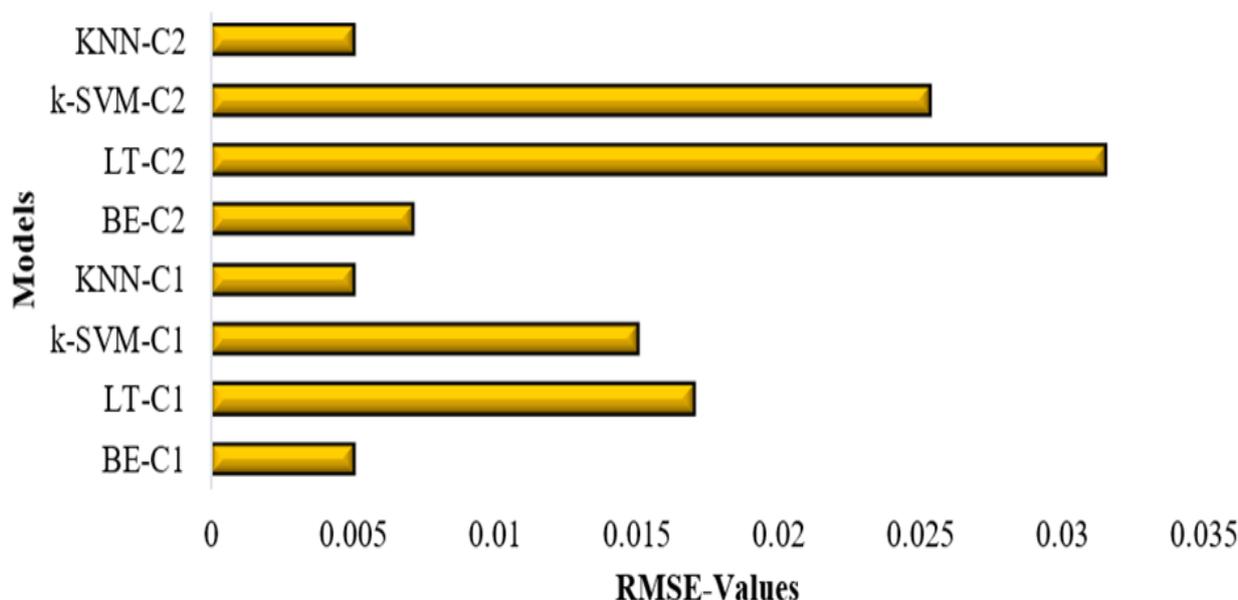
Figure 10. Performance of the model prediction of floods using the RMSE values.

Generating Flood Susceptibility Map

In a complex prone flood area such as Jeddah, it is very essential to develop and categories flood susceptibility probability maps for both the failure and success rate. In most of the literature, training datasets were used to efficiently create flood model reliability performance. According to Towfiqul Islam et al. [98] spatial flood prediction was attributed to the probability of several flood vulnerability maps. The flood maps in this study were generated using GIS software after the model's calibration. Researchers such as [46] categorised the various approaches of flood susceptibility indices (FSI), for example, natural breaks, standard deviation, interval techniques, etc., depending upon the nature and objective of the specific method. Moreover, the previous literature recommended the quantile-based approach as the most widely used classification method for generating susceptibility maps. The classifications were verified using the validation results, and the outcomes show that predominant locations towards the distance to the river served as highly susceptible to flooding with high intensity of overestimation.

The flood susceptibility maps are presented in Figure 11. For comparison with Youssef et al. [67] proposed bivariate and multivariate statistical models for flood assessment in Jeddah city. The obtained conclusion indicated that high susceptibility was associated with TWI and DR along the wadis and western catchments. This conclusion is in line with our findings. Besides, the outcomes and the behaviour of flood predictors depicted that the topography of the Jeddah watershed is expected to have frequent flooding even with the small number of return periods. Generally, in the Kingdom of Saudi Arabia, for example, so much research has been taken to map out the flood-vulnerable areas and zones to take accurate mitigate its future occurrence; still, in an ongoing flood event, the adoption of the policies and strategies are not fully taking cognizance by the decision-makers. Information on flood characteristics and their effects is vital for flood defence authorities to help formulate policies and, at the same time, in flood decision-making management strategies such as building flood protection structures to strengthen flood emergency response settlement plans [99].
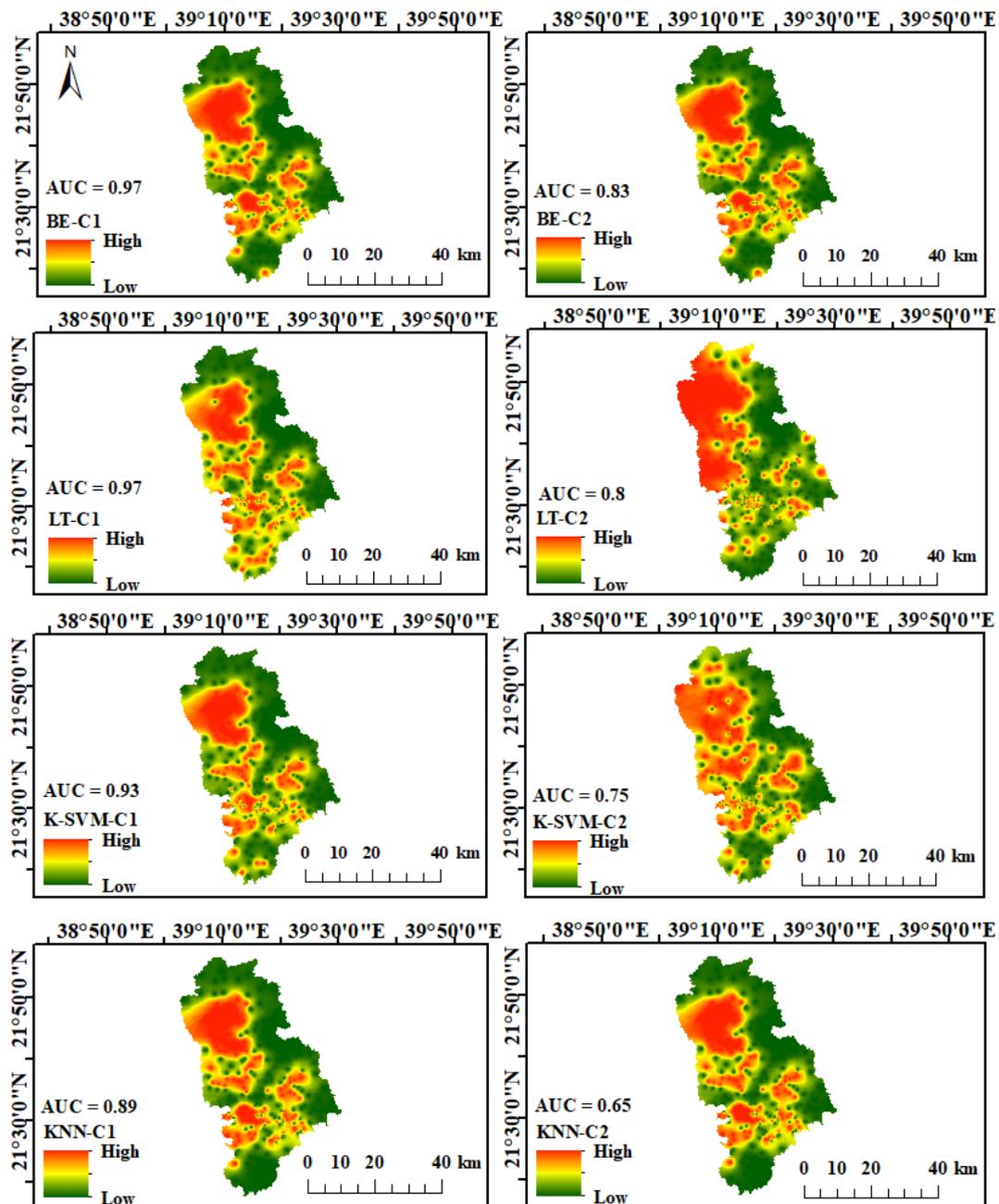
Figure 11. Flood susceptibility maps derived from BE, LT, k-SVM, and KNN models.

6. Conclusions

Flash flood management is crucial to preventing human casualties and economic losses, particularly in high residential areas like Jeddah. Therefore, high-accuracy mapping of flood susceptibility is seen to be crucial for developing flood management strategies, especially in light of climate change. The major motive behind this study is attributed to the flood scenarios of 2009 and 2011, where more than 113 persons died in addition to damaged records of buildings, roads, cars, and the loss of several properties. This study explored machine learning models viz: bagging ensemble (BE), logistic model tree (LT), kernel support vector machine (k-SVM), and k-nearest neighbour (KNN) for flood susceptibility mapping and prediction in Jeddah City. For this purpose, dependency analysis (DA) was employed to feature categorization and selection. The AUC values of the BE-C1, LT-C1, k-SVM-C1, and KNN-C1 are 97%, 97%, 93%, and 89%, respectively. While AUC values of the BE-C2, LT-C2, k-SVM-C2, and KNN-C2 are 0.83%, 0.8%, 0.75%, and 0.65%, respectively.

In general, the introduced models in this paper can be used as alternate methods for spatial flood model prediction. However, owing to the emerging knowledge in the field of AI-based models it is suggested that other feasible alternative approaches such as feature selection methods, hybrid learning techniques, and optimization algorithms should also be practised. The major limitation of this study is based on data mining and gathering using highly sophisticated high-resolution satellites.

Author Contributions: Conceptualization, A.M.A.-A. and S.I.A.; methodology, S.I.A. and A.M.A.-A.; software, S.I.A., M.G. and A.M.A.-A.; validation, S.I.A. and A.M.A.-A.; formal analysis, A.M.A.-A. and S.I.A.; investigation A.M.A.-A. and S.I.A.; resources, A.M.A.-A., S.I.A. and M.B.; data curation, A.M.A.-A.; writing—original draft preparation, A.M.A.-A., S.I.A., M.A.Y., M.B., M.G. and I.H.A.; writing—review and editing, A.M.A.-A., S.I.A. and I.H.A.; visualization, A.M.A.-A., S.I.A. and M.G.; supervision, I.H.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data are available based on the request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ali, S.A.; Parvin, F.; Pham, Q.B.; Vojtek, M.; Vojteková, J.; Costache, R.; Linh, N.T.T.; Nguyen, H.Q.; Ahmad, A.; Ghorbani, M.A. GIS-Based Comparative Assessment of Flood Susceptibility Mapping Using Hybrid Multi-Criteria Decision-Making Approach, Naïve Bayes Tree, Bivariate Statistics and Logistic Regression: A Case of Topľa Basin, Slovakia. *Ecol. Indic.* **2020**, *117*, 106620. [CrossRef]
2. Alfieri, L.; Dottori, F.; Betts, R.; Salamon, P.; Feyen, L. Multi-Model Projections of River Flood Risk in Europe under Global Warming. *Climate* **2018**, *6*, 6. [CrossRef]
3. Jonkman, S.N. Global Perspectives on Loss of Human Life Caused by Floods. *Natural. Hazards* **2005**, *34*, 151–175. [CrossRef]
4. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated World Map of the Köppen-Geiger Climate Classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [CrossRef]
5. Youssef, A.M.; Maerz, N.H. Overview of Some Geological Hazards in the Saudi Arabia. *Environ. Earth Sci.* **2013**, *70*, 3115–3130. [CrossRef]
6. Maghrabi, K. Impact of Flood Disaster on the Mental Health of Residents in the Eastern Region of Jeddah Governorate, 2010: A Study in Medical Geography. *Life Sci. J.* **2012**, *9*, 95–110.
7. Momani, N.M.; Fadil, A.S. Changing Public Policy Due to Saudi City of Jeddah Flood Disaster. *J. Soc. Sci.* **2010**, *6*, 424–428. [CrossRef]
8. Tien Bui, D.; Hoang, N.-D.; Pham, T.-D.; Ngo, P.-T.T.; Hoa, P.V.; Minh, N.Q.; Tran, X.-T.; Samui, P. A New Intelligence Approach Based on GIS-Based Multivariate Adaptive Regression Splines and Metaheuristic Optimization for Predicting Flash Flood Susceptible Areas at High-Frequency Tropical Typhoon Area. *J. Hydrol. (Amst.)* **2019**, *575*, 314–326. [CrossRef]
9. Tien Bui, D.; Hoang, N.-D. A Bayesian Framework Based on a Gaussian Mixture Model and Radial-Basis-Function Fisher Discriminant Analysis (BayGmmKda V1.1) for Spatial Prediction of Floods. *Geosci. Model. Dev.* **2017**, *10*, 3391–3409. [CrossRef]
10. Downer, C.W.; Ogden, F.L. GSSHA: Model To Simulate Diverse Stream Flow Producing Processes. *J. Hydrol. Eng.* **2004**, *9*, 161–174. [CrossRef]
11. Zhou, Q.; Mikkelsen, P.S.; Halsnæs, K.; Arnbjerg-Nielsen, K. Framework for Economic Pluvial Flood Risk Assessment Considering Climate Change Effects and Adaptation Benefits. *J. Hydrol. (Amst.)* **2012**, *414–415*, 539–549. [CrossRef]
12. Scharffenberg, W. *Hydrologic Modeling System HEC-HMS—User's Manual*; US Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center: Davis, CA, USA, 2013; 442p.
13. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: Model development1. *JAWRA J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89. [CrossRef]
14. Neitsch, S.L.; Arnold, J.G.; Kiniry, J.R.; Srinivasan, R.; Williams, J.R. *Soil and Water Assessment Tool Input/Output File Documentation Version 2005*; Texas A&M University System: College Station, TX, USA, 2004.
15. Neitsch, S.L.; Arnold, J.G.; Kiniry, J.R.; Williams, J.R. *College of Agriculture and Life Sciences Soil and Water Assessment Tool Theoretical Documentation Version 2009*; Texas A&M University System: College Station, TX, USA, 2011.

16. Al-Areeq, A.M.; Al-Zahrani, M.A.; Sharif, H.O. Physically-Based, Distributed Hydrologic Model for Makkah Watershed Using GPM Satellite Rainfall and Ground Rainfall Stations. *Geomat. Nat. Hazards Risk* **2021**, *12*, 1234–1257. [CrossRef]
17. Al-Areeq, A.M.; Al-Zahrani, M.A.; Sharif, H.O. The Performance of Physically Based and Conceptual Hydrologic Models: A Case Study for Makkah Watershed, Saudi Arabia. *Water* **2021**, *13*, 1098. [CrossRef]
18. Al-Zahrani, M.; Al-Areeq, A.; Sharif, H.O. Estimating Urban Flooding Potential near the Outlet of an Arid Catchment in Saudi Arabia. *Geomat. Nat. Hazards Risk* **2016**, *8*, 672–688. [CrossRef]
19. Tien Bui, D.; Pradhan, B.; Nampak, H.; Bui, Q.-T.; Tran, Q.-A.; Nguyen, Q.-P. Hybrid Artificial Intelligence Approach Based on Neural Fuzzy Inference Model and Metaheuristic Optimization for Flood Susceptibilitgy Modeling in a High-Frequency Tropical Cyclone Area Using GIS. *J. Hydrol. (Amst.)* **2016**, *540*, 317–330. [CrossRef]
20. Khosravi, K.; Panahi, M.; Tien Bui, D. Spatial Prediction of Groundwater Spring Potential Mapping Based on an Adaptive Neuro-Fuzzy Inference System and Metaheuristic Optimization. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 4771–4792. [CrossRef]
21. de Musso, N.M.; Capolongo, D.; Refice, A.; Lovergine, F.P.; D'Addabbo, A.; Pennetta, L. Spatial Evolution of the December 2013 Metaponto Plain (Basilicata, Italy) Flood Event Using Multi-Source and High-Resolution Remotely Sensed Data. *J. Maps* **2018**, *14*, 219–229. [CrossRef]
22. Tong, X.; Luo, X.; Liu, S.; Xie, H.; Chao, W.; Liu, S.; Liu, S.; Makhinov, A.N.; Makhinova, A.F.; Jiang, Y. An Approach for Flood Monitoring by the Combined Use of Landsat 8 Optical Imagery and COSMO-SkyMed Radar Imagery. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 144–153. [CrossRef]
23. Lim, J.; Lee, K. Flood Mapping Using Multi-Source Remotely Sensed Data and Logistic Regression in the Heterogeneous Mountainous Regions in North Korea. *Remote Sens.* **2018**, *10*, 1036. [CrossRef]
24. Elkiran, G.; Ergil, M. The Assessment of a Water Budget of North Cyprus. *Build. Environ.* **2006**, *41*, 1671–1677. [CrossRef]
25. Osinowo, A.A.; Okogbue, E.C.; Ogungbenro, S.B.; Fashanu, O. Analysis of Global Solar Irradiance over Climatic Zones in Nigeria for Solar Energy Applications. *J. Sol. Energy* **2015**, *2015*, 819307. [CrossRef]
26. Pham, Q.B.; Abba, S.I.; Usman, A.G.; Linh, N.T.T.; Gupta, V.; Malik, A.; Costache, R.; Vo, N.D.; Tri, D.Q. Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall. *Water Resour. Manag.* **2019**, *33*, 5067–5087. [CrossRef]
27. Sajedi-Hosseini, F.; Malekian, A.; Choubin, B.; Rahmati, O.; Cipullo, S.; Coulon, F.; Pradhan, B. A Novel Machine Learning-Based Approach for the Risk Assessment of Nitrate Groundwater Contamination. *Sci. Total Environ.* **2018**, *644*, 954–962. [CrossRef]
28. Tehrany, M.S.; Lee, M.-J.; Pradhan, B.; Jebur, M.N.; Lee, S. Flood Susceptibility Mapping Using Integrated Bivariate and Multivariate Statistical Models. *Environ. Earth Sci.* **2014**, *72*, 4001–4015. [CrossRef]
29. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood Susceptibility Mapping Using Frequency Ratio and Weights-of-Evidence Models in the Golastan Province, Iran. *Geocarto Int.* **2015**, *31*, 42–70. [CrossRef]
30. Gudiyangada Nachappa, T.; Tavakkoli Piralilou, S.; Gholamnia, K.; Ghorbanzadeh, O.; Rahmati, O.; Blaschke, T. Flood Susceptibility Mapping with Machine Learning, Multi-Criteria Decision Analysis and Ensemble Using Dempster Shafer Theory. *J. Hydrol. (Amst.)* **2020**, *590*, 125275. [CrossRef]
31. Avand, M.; Moradi, H.; Lasboyee, M.R. Using Machine Learning Models, Remote Sensing, and GIS to Investigate the Effects of Changing Climates and Land Uses on Flood Probability. *J. Hydrol. (Amst.)* **2021**, *595*, 125663. [CrossRef]
32. Avand, M.; Moradi, H.R.; Ramazanzadeh Lasboyee, M. Spatial Prediction of Future Flood Risk: An Approach to the Effects of Climate Change. *Geosciences* **2021**, *11*, 25. [CrossRef]
33. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens.* **2019**, *11*, 196. [CrossRef]
34. Lee, S.; Kim, J.-C.; Jung, H.-S.; Lee, M.J.; Lee, S. Spatial Prediction of Flood Susceptibility Using Random-Forest and Boosted-Tree Models in Seoul Metropolitan City, Korea. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1185–1203. [CrossRef]
35. Costache, R. Flood Susceptibility Assessment by Using Bivariate Statistics and Machine Learning Models-a Useful Tool for Flood Risk Management. *Water Resour. Manag.* **2019**, *33*, 3239–3256. [CrossRef]
36. Yariyan, P.; Avand, M.; Abbaspour, R.A.; Torabi Haghighi, A.; Costache, R.; Ghorbanzadeh, O.; Janizadeh, S.; Blaschke, T. Flood Susceptibility Mapping Using an Improved Analytic Network Process with Statistical Models. *Geomat. Nat. Hazards Risk* **2020**, *11*, 2282–2314. [CrossRef]
37. Rahman, M.; Ningsheng, C.; Islam, M.M.; Dewan, A.; Iqbal, J.; Washakh, R.M.A.; Shufeng, T. Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-Criteria Decision Analysis. *Earth Syst. Environ.* **2019**, *3*, 585–601. [CrossRef]
38. Vilasan, R.T.; Kapse, V.S. Evaluation of the Prediction Capability of AHP and F-AHP Methods in Flood Susceptibility Mapping of Ernakulam District (India). *Natural. Hazards* **2022**, *112*, 1767–1793. [CrossRef]
39. Fabio, D.N.; Abba, S.I.; Pham, B.Q.; Towfiqul Islam, A.R.M.; Talukdar, S.; Francesco, G. Groundwater Level Forecasting in Northern Bangladesh Using Nonlinear Autoregressive Exogenous (NARX) and Extreme Learning Machine (ELM) Neural Networks. *Arab. J. Geosci.* **2022**, *15*, 647. [CrossRef]
40. Akter, T.; Quevauviller, P.; Eisenreich, S.J.; Vaes, G. Impacts of Climate and Land Use Changes on Flood Risk Management for the Schijn River, Belgium. *Environ. Sci. Policy* **2018**, *89*, 163–175. [CrossRef]
41. Costache, R.; Trung Tin, T.; Arabameri, A.; Crăciun, A.; Ajin, R.S.; Costache, I.; Towfiqul Islam, A.R.M.; Abba, S.I.; Sahana, M.; Avand, M.; et al. Flash-Flood Hazard Using Deep Learning Based on H2O R Package and Fuzzy-Multicriteria Decision-Making Analysis. *J. Hydrol. (Amst.)* **2022**, *609*, 127747. [CrossRef]
42. Hadian, S.; Shahiri Tabarestani, E.; Pham, Q.B. Multi Attributive Ideal-Real Comparative Analysis (MAIRCA) Method for Evaluating Flood Susceptibility in a Temperate Mediterranean Climate. *Hydrol. Sci. J.* **2022**, *67*, 401–418. [CrossRef]

43. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.-X.; Li, S. Novel Hybrid Artificial Intelligence Approach of Bivariate Statistical-Methods-Based Kernel Logistic Regression Classifier for Landslide Susceptibility Modeling. *Bull. Eng. Geol. Environ.* **2018**, *78*, 4397–4419. [CrossRef]

44. Rahmati, O.; Pourghasemi, H.R. Identification of Critical Flood Prone Areas in Data-Scarce and Ungauged Regions: A Comparison of Three Data Mining Models. *Water Resour. Manag.* **2017**, *31*, 1473–1487. [CrossRef]

45. Liu, J.; Wang, J.; Xiong, J.; Cheng, W.; Li, Y.; Cao, Y.; He, Y.; Duan, Y.; He, W.; Yang, G. Assessment of Flood Susceptibility Mapping Using Support Vector Machine, Logistic Regression and Their Ensemble Techniques in the Belt and Road Region. *Geocarto Int.* **2022**. [CrossRef]

46. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood Susceptibility Assessment Using GIS-Based Support Vector Machine Model with Different Kernel Types. *Catena (Amst.)* **2015**, *125*, 91–101. [CrossRef]

47. Ahmadlou, M.; Al-Fugara, A.; Al-Shabeeb, A.R.; Arora, A.; Al-Adamat, R.; Pham, Q.B.; Al-Ansari, N.; Linh, N.T.T.; Sajedi, H. Flood Susceptibility Mapping and Assessment Using a Novel Deep Learning Model Combining Multilayer Perceptron and Autoencoder Neural Networks. *J. Flood Risk Manag.* **2020**, *14*, e12683. [CrossRef]

48. Nandi, A.; Mandal, A.; Wilson, M.; Smith, D. Flood Hazard Mapping in Jamaica Using Principal Component Analysis and Logistic Regression. *Environ. Earth Sci.* **2016**, *75*, 465. [CrossRef]

49. Khoirunisa, N.; Ku, C.-Y.; Liu, C.-Y. A GIS-Based Artificial Neural Network Model for Flood Susceptibility Assessment. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1072. [CrossRef] [PubMed]

50. Sahoo, G.B.; Ray, C.; de Carlo, E.H. Use of Neural Network to Predict Flash Flood and Attendant Water Qualities of a Mountainous Stream on Oahu, Hawaii. *J. Hydrol. (Amst.)* **2006**, *327*, 525–538. [CrossRef]

51. Youssef, A.M.; Pradhan, B.; Hassan, A.M. Flash Flood Risk Estimation along the St. Katherine Road, Southern Sinai, Egypt Using GIS Based Morphometry and Satellite Imagery. *Environ. Earth Sci.* **2010**, *62*, 611–623. [CrossRef]

52. Costache, R.; Tien Bui, D. Spatial Prediction of Flood Potential Using New Ensembles of Bivariate Statistics and Artificial Intelligence: A Case Study at the Putna River Catchment of Romania. *Sci. Total Environ.* **2019**, *691*, 1098–1118. [CrossRef]

53. Zhao, G.; Pang, B.; Xu, Z.; Peng, D.; Xu, L. Assessment of Urban Flood Susceptibility Using Semi-Supervised Machine Learning Model. *Sci. Total Environ.* **2019**, *659*, 940–949. [CrossRef]

54. Tang, X.; Li, J.; Liu, M.; Liu, W.; Hong, H. Flood Susceptibility Assessment Based on a Novel Random Naïve Bayes Method: A Comparison between Different Factor Discretization Methods. *Catena (Amst.)* **2020**, *190*, 104536. [CrossRef]

55. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B.; et al. Modeling Flood Susceptibility Using Data-Driven Approaches of Naïve Bayes Tree, Alternating Decision Tree, and Random Forest Methods. *Sci. Total Environ.* **2020**, *701*, 134979. [CrossRef] [PubMed]

56. Darabi, H.; Choubin, B.; Rahmati, O.; Torabi Haghighi, A.; Pradhan, B.; Kløve, B. Urban Flood Risk Mapping Using the GARP and QUEST Models: A Comparative Study of Machine Learning Techniques. *J. Hydrol. (Amst.)* **2019**, *569*, 142–154. [CrossRef]

57. Razavi-Termeh, S.V.; Khosravi, K.; Sadeghi-Niaraki, A.; Choi, S.-M.; Singh, V.P. Improving Groundwater Potential Mapping Using Metaheuristic Approaches. *Hydrol. Sci. J.* **2020**, *65*, 2729–2749. [CrossRef]

58. Razavi Termeh, S.V.; Kornejady, A.; Pourghasemi, H.R.; Keesstra, S. Flood Susceptibility Mapping Using Novel Ensembles of Adaptive Neuro Fuzzy Inference System and Metaheuristic Algorithms. *Sci. Total Environ.* **2018**, *615*, 438–451. [CrossRef] [PubMed]

59. Hong, H.; Pradhan, B.; Bui, D.T.; Xu, C.; Youssef, A.M.; Chen, W. Comparison of Four Kernel Functions Used in Support Vector Machines for Landslide Susceptibility Mapping: A Case Study at Suichuan Area (China). *Geomat. Nat. Hazards Risk* **2016**, *8*, 544–569. [CrossRef]

60. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A Novel Hybrid Artificial Intelligence Approach for Flood Susceptibility Assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [CrossRef]

61. Wang, Y.; Hong, H.; Chen, W.; Li, S.; Pamučar, D.; Gigović, L.; Drobnjak, S.; Bui, D.T.; Duan, H. A Hybrid GIS Multi-Criteria Decision-Making Method for Flood Susceptibility Mapping at Shangyou, China. *Remote Sens.* **2018**, *11*, 62. [CrossRef]

62. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An Ensemble Prediction of Flood Susceptibility Using Multivariate Discriminant Analysis, Classification and Regression Trees, and Support Vector Machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [CrossRef]

63. Ngo, P.-T.T.; Hoang, N.-D.; Pradhan, B.; Nguyen, Q.K.; Tran, X.T.; Nguyen, Q.M.; Nguyen, V.N.; Samui, P.; Tien Bui, D. A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data. *Sensors* **2018**, *18*, 3704. [CrossRef]

64. Bui, D.T.; Tsangaratos, P.; Ngo, P.-T.T.; Pham, T.D.; Pham, B.T. Flash Flood Susceptibility Modeling Using an Optimized Fuzzy Rule Based Feature Selection Technique and Tree Based Ensemble Methods. *Sci. Total Environ.* **2019**, *668*, 1038–1054. [CrossRef] [PubMed]

65. Mosavi, A.; Sajedi Hosseini, F.; Choubin, B.; Taromideh, F.; Ghodsi, M.; Nazari, B.; Dineva, A.A. Susceptibility Mapping of Groundwater Salinity Using Machine Learning Models. *Environ. Sci. Pollut. Res.* **2020**, *28*, 10804–10817. [CrossRef] [PubMed]

66. GAS. *Population*; General Authority for Statistics: Riyadh, Saudi Arabia, 2020.

67. Youssef, A.M.; Pradhan, B.; Sefry, S.A. Flash Flood Susceptibility Assessment in Jeddah City (Kingdom of Saudi Arabia) Using Bivariate and Multivariate Statistical Models. *Environ. Earth Sci.* **2015**, *75*, 12. [CrossRef]

68. Costache, R.; Pham, Q.B.; Sharifi, E.; Linh, N.T.T.; Abba, S.I.; Vojtek, M.; Vojteková, J.; Nhi, P.T.T.; Khoi, D.N. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. *Remote Sens.* **2019**, *12*, 106. [CrossRef]

69. Sammen, S.S.; Mohammed, T.A.; Ghazali, A.H.; Sidek, L.M.; Shahid, S.; Abba, S.I.; Malik, A.; Al-Ansari, N. Assessment of Climate Change Impact on Probable Maximum Floods in a Tropical Catchment. *Theor. Appl. Climatol.* **2022**, *148*, 15–31. [CrossRef]

70. Mosavi, A.; Ozturk, P.; Chau, K. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [CrossRef]

71. Costache, R.; Arabameri, A.; Costache, I.; Crăciun, A.; Md Towfiqul Islam, A.R.; Abba, S.I.; Sahana, M.; Pham, B.T. Flood Susceptibility Evaluation through Deep Learning Optimizer Ensembles and GIS Techniques. *J. Environ. Manag.* **2022**, *316*, 115316. [CrossRef]

72. Chen, T.; Ren, J. Bagging for Gaussian Process Regression. *Neurocomputing* **2009**, *72*, 1605–1610. [CrossRef]

73. Adnan, R.M.; Jaafari, A.; Mohanavelu, A.; Kisi, O.; Elbeltagi, A. Novel Ensemble Forecasting of Streamflow Using Locally Weighted Learning Algorithm. *Sustainability* **2021**, *13*, 5877. [CrossRef]

74. Azhari, M.; Abarda, A.; Alaoui, A.; Ettaki, B.; Zerouaoui, J. Detection of Pulsar Candidates Using Bagging Method. *Procedia Comput. Sci.* **2020**, *170*, 1096–1101. [CrossRef]

75. Xue, X.; Zhang, K.; Tan, K.C.; Feng, L.; Wang, J.; Chen, G.; Zhao, X.; Zhang, L.; Yao, J. Affine Transformation-Enhanced Multifactorial Optimization for Heterogeneous Problems. *IEEE Trans. Cybern.* **2022**, *52*, 6217–6231. [CrossRef] [PubMed]

76. Tuyen, T.T.; Jaafari, A.; Yen, H.P.H.; Nguyen-Thoi, T.; van Phong, T.; Nguyen, H.D.; van Le, H.; Phuong, T.T.M.; Nguyen, S.H.; Prakash, I.; et al. Mapping Forest Fire Susceptibility Using Spatially Explicit Ensemble Models Based on the Locally Weighted Learning Algorithm. *Ecol. Inform.* **2021**, *63*, 101292. [CrossRef]

77. Landwehr, N.; Hall, M.; Frank, E. Logistic Model Trees. *Mach. Learn.* **2005**, *59*, 161–205. [CrossRef]

78. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. *Landslides* **2015**, *13*, 361–378. [CrossRef]

79. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017; ISBN 1315139472.

80. Shah, K.; Patel, H.; Sanghvi, D.; Shah, M. A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. *Augment. Hum. Res.* **2020**, *5*, 12. [CrossRef]

81. Chen, W.; Shahabi, H.; Shirzadi, A.; Li, T.; Guo, C.; Hong, H.; Li, W.; Pan, D.; Hui, J.; Ma, M.; et al. A Novel Ensemble Approach of Bivariate Statistical-Based Logistic Model Tree Classifier for Landslide Susceptibility Assessment. *Geocarto Int.* **2018**, *33*, 1398–1420. [CrossRef]

82. Usman, A.G.; Işik, S.; Abba, S.I. Hybrid Data-Intelligence Algorithms for the Simulation of Thymoquinone in HPLC Method Development. *J. Iran. Chem. Soc.* **2021**, *18*, 1537–1549. [CrossRef]

83. Veenaas, C.; Linusson, A.; Haglund, P. Retention-Time Prediction in Comprehensive Two-Dimensional Gas Chromatography to Aid Identification of Unknown Contaminants. *Anal. Bioanal. Chem.* **2018**, *410*, 7931–7941. [CrossRef]

84. Olson, R.S.; la Cava, W.; Mustahsan, Z.; Varik, A.; Moore, J.H. Data-Driven Advice for Applying Machine Learning to Bioinformatics Problems. *Pac. Symp. Biocomput.* **2018**, *2018*, 192–203. [CrossRef]

85. Tewari, S.; Dwivedi, U.D. Ensemble-Based Big Data Analytics of Lithofacies for Automatic Development of Petroleum Reservoirs. *Comput. Ind. Eng.* **2019**, *128*, 937–947. [CrossRef]

86. Chuma, G.B.; Bora, F.S.; Ndeko, A.B.; Mugumaarhahama, Y.; Cirezi, N.C.; Mondo, J.M.; Bagula, E.M.; Karume, K.; Mushagalusa, G.N.; Schimtz, S. Estimation of Soil Erosion Using RUSLE Modeling and Geospatial Tools in a Tea Production Watershed (Chiseke in Walungu), Eastern Democratic Republic of Congo. *Model. Earth Syst. Environ.* **2021**, *8*, 1273–1289. [CrossRef]

87. ArunKumar, K.E.; Kalaga, D.V.; Sai Kumar, C.M.; Chilkoor, G.; Kawaji, M.; Brenza, T.M. Forecasting the Dynamics of Cumulative COVID-19 Cases (Confirmed, Recovered and Deaths) for Top-16 Countries Using Statistical Machine Learning Models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Appl. Soft Comput.* **2021**, *103*, 107161. [CrossRef] [PubMed]

88. Bagherzadeh, F.; Mehrani, M.-J.; Basirifard, M.; Roostaei, J. Comparative Study on Total Nitrogen Prediction in Wastewater Treatment Plant and Effect of Various Feature Selection Methods on Machine Learning Algorithms Performance. *J. Water Process Eng.* **2021**, *41*, 102033. [CrossRef]

89. Zeng, J.; Chai, Q.; Peng, X.; Li, S. Geographical Origin Identification for Tetrastigma Hemsleyanum Based on High Performance Liquid Chromatographic Fingerprint. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 1816–1820. [CrossRef]

90. Agrawal, P.; Ganesh, T.; Mohamed, A.W. A Novel Binary Gaining–Sharing Knowledge-Based Optimization Algorithm for Feature Selection. *Neural. Comput. Appl.* **2020**, *33*, 5989–6008. [CrossRef]

91. Yaseen, Z.M.; Deo, R.C.; Hilal, A.; Abd, A.M.; Bueno, L.C.; Salcedo-Sanz, S.; Nehdi, M.L. Predicting Compressive Strength of Lightweight Foamed Concrete Using Extreme Learning Machine Model. *Adv. Eng. Softw.* **2018**, *115*, 112–125. [CrossRef]

92. Kombo, O.; Kumaran, S.; Sheikh, Y.; Bovim, A.; Jayavel, K. Long-Term Groundwater Level Prediction Model Based on Hybrid KNN-RF Technique. *Hydrology* **2020**, *7*, 59. [CrossRef]

93. Thi Thuy Linh, N.; Pandey, M.; Janizadeh, S.; Sankar Bhunia, G.; Norouzi, A.; Ali, S.; Bao Pham, Q.; Tran Anh, D.; Ahmadi, K. Flood Susceptibility Modeling Based on New Hybrid Intelligence Model: Optimization of XGboost Model Using GA Metaheuristic Algorithm. *Adv. Space Res.* **2022**, *69*, 3301–3318. [CrossRef]

94. Sakizadeh, M.; Mirzaei, R. A Comparative Study of Performance of K-Nearest Neighbors and Support Vector Machines for Classification of Groundwater. *J. Min. Environ.* **2016**, *7*, 149–164. [CrossRef]

95. Sami, N.A.; Ibrahim, D.S. Forecasting Multiphase Flowing Bottom-Hole Pressure of Vertical Oil Wells Using Three Machine Learning Techniques. *Pet. Res.* **2021**, *6*, 417–422. [CrossRef]

96. Costache, R.; Tien Bui, D. Identification of Areas Prone to Flash-Flood Phenomena Using Multiple-Criteria Decision-Making, Bivariate Statistics, Machine Learning and Their Ensembles. *Sci. Total Environ.* **2020**, *712*, 136492. [CrossRef]

97. Costache, R. Flash-Flood Potential Assessment in the Upper and Middle Sector of Prahova River Catchment (Romania). A Comparative Approach between Four Hybrid Models. *Sci. Total Environ.* **2019**, *659*, 1115–1134. [CrossRef] [PubMed]

98. Towfiqul Islam, A.R.M.; Talukdar, S.; Mahato, S.; Kundu, S.; Eibek, K.U.; Pham, Q.B.; Kuriqi, A.; Linh, N.T.T. Flood Susceptibility Modelling Using Advanced Ensemble Machine Learning Models. *Geosci. Front.* **2021**, *12*, 101075. [CrossRef]

99. Desalegn, H.; Mulu, A. Flood Vulnerability Assessment Using GIS at Fetam Watershed, Upper Abbay Basin, Ethiopia. *Heliyon* **2021**, *7*, e05865. [CrossRef] [PubMed]