



Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest

Emrehan Kutlug Sahin¹ Received: 22 April 2020 / Accepted: 11 June 2020 / Published online: 30 June 2020
© Springer Nature Switzerland AG 2020

Abstract

Decision tree-based classifier ensemble methods are a machine learning (ML) technique that combines several tree models to produce an effective or optimum predictive model, and that allows well-predictive performance especially compared to a single model. Thus, selecting a proper ML algorithm help us to understand possible future occurrences by analyzing the past more accurate. The main purpose of this study is to produce landslide susceptibility map of the Ayancik district of Sinop province, situated in the Black Sea region of Turkey using three featured regression tree-based ensemble methods including gradient boosting machines (GBM), extreme gradient boosting (XGBoost), and random forest (RF). Fifteen landslide causative factors and 105 landslide locations occurred in the region were used. The landslide inventory map was randomly divided into training (70%) and testing (30%) dataset to construct the RF, XGBoost and GBM prediction models. Symmetrical uncertainty measure was utilized to determine the most important causative factors, and then the selected features were used to construct susceptibility prediction models. The performance of the ensemble models was validated using different accuracy metrics including Area under the curve (AUC), overall accuracy (OA), Root mean square error (RMSE), and Kappa coefficient. Also, the Wilcoxon signed-rank test was used to assess differences between optimum models. The accuracy results showed that the model of XgBoost_Opt model (the model created by optimum factor combination) has the highest prediction capability (OA = 0.8501 and AUC = 0.8976), followed by the RF_opt (OA = 0.8336 and AUC = 0.8860) and GBM_Opt (OA = 0.8244 and AUC = 0.8796). When the Wilcoxon sign-rank test results were analyzed, XgBoost_Opt model, which is the best subset combinations, were confirmed to be statistically significant considering other models. The results showed that, the XGBoost method according to optimum model achieved lower prediction error and higher accuracy results than the other ensemble methods.

Keywords Ensemble tree · Machine learning · Extreme gradient boosting · Symmetrical uncertainty · Feature selection

1 Introduction

In many parts of the world, natural disasters like landslides are major natural hazard and cause threat to human's life, economic losses and the environment. Over the last 2 decades or so, international organizations in disaster management including government and research institutions have been focused on produced assessment methodologies

and to portray its spatial distribution in maps [41]. In any type of landslide hazard assessment methodology, there is a need to consider several processes such as landslide inventory map (LIM) production, determination of optimum factors combination, selection of method for the preparation of landslide susceptibility maps (LSMs) and performance analysis [49, 72, 78]. For produce reliable and accurate map showing the susceptibility of a particular

✉ Emrehan Kutlug Sahin, emrehansahin@ibu.edu.tr; emrehans@gmail.com | ¹Department of Civil Engineering, Bolu Abant Izzet Baysal University, Bolu, Turkey.



region to landslide, a prerequisite is to have information regarding the spatial and temporal frequency of landslides [34, 80]. Furthermore, in the LSM studies, landslide inventory is also useful to production of models, accuracy assessment of the resulting map and validation of output scores.

Another challenge faced by researchers is the nature of the Earth is not the same and the factors triggering the landslide is not consistent [67]. Researchers also have to decide the best combination of factors to create the desired prediction model for each study area [24, 52, 56]. Chen et al. [16], were applied to principal component analysis to select significant and independent factors through 17 contributing factors. As a result, six factors (slope, distance from fault, aspect, lithology, elevation, and settlement density) were selected as the explanatory features that contribute to landslide occurrence for this study area. Tsangaratos and Ilia [99] stated that multicollinearity analysis can be used in order to determine the conditional independence among variables for the feature selection process. Also, feature selection algorithm as Fisher score [40, 56], gain ratio [94, 107], χ^2 [46, 56], learning vector quantization [66], correlation based feature selection [84] and information gain [11, 79] was used to determine the best feature combination or feature ranking by several researchers. Due to the complexity of mechanism of system structure of landslide more flexible nonlinear ML algorithms [14, 63, 64] have also been utilized to find more efficient or irrelevant factors for producing landslide related applications.

While the quality of the data is an important issue for performance of landslide modelling, modeling approaches to be selected is another issue in LSM. Although, there is no universally accepted specific methodology or procedure for assessment and prediction pattern of landslide hazard, to date, there are several approaches that have been utilized for the prediction of LSM [41, 78]. These approaches generally can be divided into two broad categories as qualitative and quantitative methods from logical, experience-based analysis, extending to complex mathematical and computer-based systems [30]. Generally, qualitative methods called heuristic approaches are based on expert opinions. These methods depend on previous landslide occurrence to determine locations in disaster-prone areas. There are a number of qualitative methods that have been employed to produce LSMs, e.g., analytical hierarchy process [51], weighted linear combination [6] and frequency ratio [2]. Although qualitative studies have been carried out for landslide hazard management, it has been gradually abandoned due to the risk that expert judgment and become more subjective than quantitative methods [20]. Quantitative methodology such as logistic regression (LR), decision tree, artificial neural network and, support vector machine

is broadly construed and refers to subjective view involved in statistical/mathematical modeling, sampling methods and model development. In recent years, ensemble learning has been widely applied to improve accuracy in classification as well as regression problems beyond the level reached by separate input data [41, 20, 30]. On the other hand, a series of ensemble-based methods (e.g. random forest, rotation forest and canonical correlation forest) have recently attracted great attention within the LSM community [18, 19, 48, 50, 82, 90].

The use of multiple learning algorithms instead of a single algorithm to unravel a given difficult problem has always been an accepted solution. Ensemble methods have been theoretically and empirically obtained better performance than single weak learners, especially while handling computational with high dimensional, complex regression and classification problems [15]. Bagging and Boosting, are the classical tree-based ensemble approaches including classification and regression, that have been improved with the objective of reducing the over-fitting problem and while leading to better accuracy [97, 101]. Also, in terms of ensemble, various approaches have been proposed and utilized in application of hazard and risk assessment, such as random forest [14, 45], rotation forest [18, 19, 72], AdaBoost [48, 64, 73], multiboost [83, 98] and Random Subspace [73, 85]. Alternatively, for LSM, Gradient Boosting Machines (GLM) has been used [61, 76, 95], which is an extremely powerful ML algorithm and that produces a strong learner in the form of an ensemble of weak learners/models. In more recent years, Extreme Gradient Boosting (XGBoost) has become a popular decision tree-based ensemble ML technique and that has been dominating applied ML for structured or tabular data. Although only very few studies are available in LSMs applications, XGBoost has been successfully used for regression or classification by researchers since its release [28, 38].

The primal objective of this work is to investigate of the three different state-of-art tree-based ensemble methods, namely RF, GBM and XGBoost for landslide susceptibility mapping at Ayancik district, Sinop province in the Black Sea region of Turkey. A number of papers have been published about the LSM assessment in the literature and most of them focus only on the feature ranking process for the detection of the factor importance or correlation with each other [29, 79, 87, 88], and only a few cases take into account detailed analysis factor relation with landslide susceptibility process [62, 64]. The novelty of this work is to search the combination of the optimal factors based on a hybrid approach using Symmetrical uncertainty (SU) for feature ranking algorithm, LR for model producing, and the Wilcoxon sign-rank test for detection of the significance of

the differences in LR model. Unlike similar studies, this study evaluates to analyze the effect of models including varying number of factors model will be tried to determine for this study area. Moreover, to analyze the performance of these models OA, AUC, Root mean square error (RMSE) and Kappa coefficient were used. Moreover, the level of performance difference of the models was confirmed by Wilcoxon sign-rank test considering to each of the stages in modelling process. The computation process was executed using R language environment software (ver. 3.5.3) and spatial data were analyzed in ArcGIS 10.

2 Study area and data used

The selected study area covered the central part of Ayancik District, Sinop province in the Black Sea region of Turkey. The research site is was cropped to a rectangular area of about 544 km², which is bounded by 34° 30' 27" to 34° 45' 45" west–east longitudes and 41° 43' 8" to 41° 57' 20" north–south latitudes (Geographic Lat/Lon WGS 84 Projection). The type of landslide activities in the study area are mainly shallow landslides and are triggered by extreme precipitation [23]. The mean annual precipitation ranged between 750 and 900 mm/year and the mean temperature is about 14 °C. Analysis of mean annual precipitation is investigated, it should be clearly said that, the spatial distribution of the mean annual precipitation in the region that appears to be significantly higher than the mean annual precipitation value given for Turkey [70]. Also, several studies [69, 74] have indicated for the black sea region covering the study area that extreme rainfall events can potentially lead to landslide activities. During February and March 1985, the heavy rains that were effective in the Middle and Western Black Sea region caused many landslide events including Sinop region. In these events, many houses were destroyed or damaged, roads were closed, and many families had to abandon the houses. Fortunately, no loss of life occurred, but hundreds of acres of farmland have been unusable due to natural hazards.

Based on the Mineral Research and Exploration of Turkey (MTA), the generalized surface geology mainly consists of clastic and carbonate rocks (coverage about 66% of study area), undifferentiated quaternary, granitoid and neritic limestone. Shallow landslides can often occur in study area and that happen frequently in most of the generalized geologic units especially with the exception of the neritic limestone and granitoid. Altitude and slope angles in the field of study ranged between around 0 and 1670 m and 0° to 72°, respectively.

2.1 Landslide inventory map

The LIM of the study was achieved from the map provided by the MTA. The main scarp of each landslides which includes depletion and accumulation areas was obtained as a vector features (i.e. polygons) and build in a geodatabase containing a series of polygon shapefiles. In this study area, the landslide activities of mass movements are classified into two groups as active and inactive, and only active slides that were classified as shallow landslide (depth < 5 m) were considered [32]. Also, all landslide areas were carefully edited and cleaned considering that based on two basic facts [37]: landslide activity is not likely to happen on river channels and on terrains with slope angles lower than 5°. After all, totally 105 polygons (represented either by landslide area) covering 6460 pixels (i.e. grid cells) were available and each pixel of the landslide area with a 30 m spatial resolution. A different landslide (1) and non-landslide (0) group samples are required for the application of LSM using ML. In region excluding landslides occurred site, areas on the river channel and places having slope angles between 0° and 5° were considered as non-landslide areas. Many sources such as digital orthophoto, high-resolution satellite image, and updated map from a filed survey using stratified random sampling have also been utilized in this process [27, 53, 55]. Finally, 63 non-landslide polygons covering 3730 pixels were determined and, all non-landslide and landslide area each other were stored in LIM. There is no rule for specifically the determination of training and validation sample size [42], but the most common and preferred ratio is 70–30 training-validation split [93]. Landslide inventory dataset was divided randomly into 70% of the total number of landslides and 70% of the total number of non-landslides for the training (74 landslides and 49 non-landslides areas) and the rest 30% (31 landslide and 14 non-landslide areas) for validation data. It should be noted that the training and testing set were selected at random, with an equal number of pixels in each subset in the assessment, validation, and recognition of model learning. Study area including inventory map was given Fig. 1.

2.2 Landslide conditioning factors

When the previous studies were analyzed, it could be said that there is still no universal guidelines or agreement for selecting landslide causative factors and conditioning factors are usually selected based on the landslide types, data availability and the characteristics of the study area [5, 29]. In this work, 15 landslide causative factors are determined to apply LSM models and to predict the distribution of landslide susceptibility. The causative factors can be categorized into four major types,

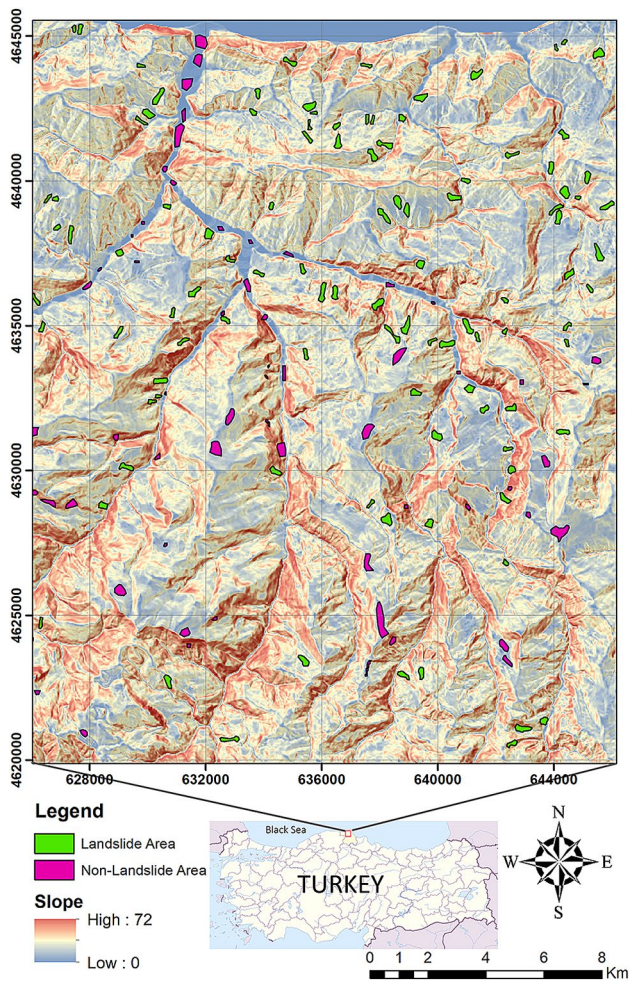


Fig. 1 Location and the map of the study area showing the general slope range and landslide inventory including training and validation data

topographic factors [slope, aspect, elevation, plan curvature, profile curvature, Topographical Position Index (TPI), Topographical Roughness Index (TRI)], geological factors (lithology), hydrological factors (Topographical Wetness Index (TWI), drainage density, Stream Power Index (SPI), Sediment Transport Index (STI), distance to river), and land cover (land use/land cover (LULC) and Normalized Difference Vegetation Index [NDVI]). Table 1 contained detailed information used in this study is given. Some of the factors can be classified as continuous or categorical form. Hence, each input factor data was converted to the same scale and intervals depending on their data structure. In addition, continuous format of maps was divided into discrete classes for standardize the maps. Hence, continuous data factors were reclassified into eight intervals by using natural break classifier [56, 68].

2.2.1 Topographic factors

Digital Elevation Model (DEM) is a crucial component to understand the nature of the terrain and it is an important input for topography for the accurate modelling and prediction of LSMs [57]. The quality of DEM has direct impacts of the derived factors (e.g. slope, aspect, curvatures, etc.) [12]. In this study, the DEM from digitized elevation contours was converted to raster format with 30 m pixel resolution. The following derived factors [slope, aspect, TRI (Topographic roughness index), TPI (Topographic position index), Topographical wetness index (TWI), plan and profile curvatures, STI and Stream power index (SPI)] were utilized for preparing the susceptibility models. Slope angle is one of the most important and the main controlling factor in LSM [24, 51]. Because the landslide occurrences and slope angles are directly associated with each other and parameter of slope is frequently used in many LSM by researchers [10, 22, 32, 104]. Slope angles were divided into 8 susceptibility level using natural break (NB) classifier from 0° to 72.3° . Altitude or elevation above sea level can express both topographic conditions and, indirectly, the role of thermo-pluviometric conditions [24]. Elevation map, range between 0 and 1670 m, was classified into eight classes using NB method. Slope aspect, which is associated with solar radiation, the wind, and rainfall, cloud play an accelerated role in the landslide activities [92, 104]. Aspect was reclassified nine classes for each of the main directions of the compass (i.e., N, NE, E, SE, S, SW, W, NW and flat). Topographic curvatures (e.g. plan and profile curvatures) control on both continuous and impulsive fluxes of water and sediment through the landscape which both surface runoff and gravitative stresses acting on shallow failure surfaces can converge or diverge. Plan curvature and profile curvature were divided into eight classes using NB strategy. TRI is a useful measure in the characterization of landslide morphology and was reclassified in 8 sub-classes by NB. TPI measures the difference between elevation at the central point of a neighborhood pixel and the average elevation around this pixel [103] and it was reclassified in 8 natural break classes.

2.2.2 Hydrological factors

SPI, a measure of the flow erosion power of the stream at the given point of the topographical surface, which can control the initiation of landslides. SPI was arranged in eight classes with natural break from -8.7 to 11.9 . TWI exhibits a topographic effect at the site of the saturated area size of runoff generation and so it can be said that lower TPI values and higher slope length and SPI values represent higher landslide susceptibility [65, 75]. TWI was reclassified into eight sub-classes using NB strategy.

Table 1 Landslide casual factor maps used in study and information of subclasses

| Major factors | Sub-factors | Sub-Classes | Source | Scale/resolution | |
|---------------|-------------------|--|--|------------------|--|
| Geology | Lithology | (1) k_2s ; (2) e_{1-2} ; (3) k_1 ; (4) Q; (5) g_4 ; (6) k_2e ; (7) j_3k_1 | General directorate of mineral research and exploration of Turkey https://www.mta.gov.tr | 1:100,000 | |
| Geomorphology | Elevation (m)—DEM | 0–163; 163–360; 360–543; 543–726; 726–903; 903–1080; 1080–1296; 1296–1670 | General directorate of mapping (Turkey) https://www.harita.gov.tr/english/ | 1:25,000 | |
| | Slope (°) | 0–5.39; 5.39–12.20; 12.20–17.30; 17.30–21.85; 21.85–26.39; 26.39–30.93; 30.93–36.32; 36.32–72.36 | DEM | 30 m | |
| | Aspect | (1) Flat; (2) north; (3) northeast; (4) east; (5) southeast; (6) south; (7) southwest; (8) west; (9) northwest | | | |
| | TRI | 0–53; 53–112; 112–171; 171–230; 230–293; 293–367; 367–473; 473–794 | | | |
| | Plan Curvature | –20.09 to –8.53; 8.53–0.83; –0.83 to –0.32; –0.32 to 0.01; 0.01–0.67; 0.67–1.68; 1.68–8.21; 8.21–22.61 | | | |
| | TPI | 0–0.14; 0.14–0.29; 0.29–0.40; 0.40–0.49; 0.49–0.58; 0.58–0.67; 0.67–0.81; 0.81–1.00 | | | |
| | Profile Curvature | –17.80 to –8.81; 8.81 to –1.63; –1.63 to –0.79; –0.79 to 0.23; 0.23–0.19; 0.19–0.75; 0.75–6.79; 6.79–18.04 | | | |
| Hydrology | Drainage density | 0–0.21; 0.21–0.42; 0.42–0.62; 0.62–0.82; 0.82–1.05; 1.05–1.23; 1.23–1.36; 1.36–1.51 | DEM | 30 m | |
| | Distance to river | 0–100; 100–200; 200–300; 300–400; 400–500; 500–750; 750–1000; 1000–> 1000 | | | |
| | STI | 0–7.92; 7.92–31.67; 31.67–71.27; 71.27–142.54; 142.54–308.84; 308.84–871.10; 871.10–1496.7; 1496.7–2019.4 | | | |
| | SPI | –8.72 to –3.53; –3.53 to –2.55; –2.55 to –1.65; –1.65 to –0.59; –0.59 to 0.37; 0.37–1.11; 1.11–5.57; 5.57–12 | | | |
| | TWI | 2.25–4.99; 4.99–5.98; 5.98–7.13; 7.13–8.57; 8.57–10.40; 10.40–12.53; 12.53–15.65; 15.65–21.67 | | | |
| Land cover | LULC | (1) deciduous; (2) coniferous; (3) bare soil; (4) bare rock; (5) urban; (6) water | Landsat-8 operational land imager (OLI) multispectral image (2016) | 30 m | |
| | NDVI | –0.18 to 0.05, 0.05–0.22, 0.22–0.34, 0.34–0.42, 0.42–0.47, 0.47–0.51, 0.51–0.86, 0.56–0.68 | | | |

Although water is not always directly involved as the transporting mass movement processes such as landslide, it could be said that it does play an important role. In a situation of distance to rivers, water can lead to the saturation sediments, reducing the integrity of the slope and allowing it to slope movement or mass movement [11, 44]. As mentioned above distance to rivers should be used for LSMs. Distance to rivers, which ranges from 0 to 1100 m, map was divided into eight categories. Drainage density, which indicates how well or how poorly a watershed is overflowed by stream channels, is the total

length of channels in a drainage basin divided by its drainage area [59]. The drainage density was considered to use LSM which ranges from 0.0 to 1.508 km^{-1} and is classified into eight classes using NB classifier. STI reflecting the erosive power of overland flow and deposition [26]. A high STI signifies for the effect of topography on erosion and the probability shows high values of STI is more vulnerable to the landslide [87, 88]. STI (0–2019) eight classes were constructed using NB algorithm for analysis.

2.2.3 Land related factors

Another conditioning parameter associated with landslide occurrences is the land use and land cover (LULC). LULC changes are affected by local, regional, global climate processes and human activities. LULC map was produced from Landsat 8 OLI (Operational Land Imager) data acquired in 2016. In general, six classes, including bare soil, coniferous, deciduous, bare rock, water and urban were identified. Vegetation cover is an important factor affecting the occurrence of shallow landslides, and changes to vegetation cover can affect landslide behavior [35]. Normalized difference vegetative index (NDVI), which is a measure of the vegetative properties, was used to estimate the density of biomass and nitrogen phenology [31]. NDVI is created from red and near-infrared bands of Landsat 8 OLI image was divided into eight sub-classes using NB.

2.2.4 Geological factor

Lithology is the most important feature of landslide phenomena which influence on the geo-mechanical characteristics of terrains [24, 58]. Lithology factor has been widely preferred by the researchers contributing the literature [1, 4, 14]. Most of the landslides that occurred in Sinop province were laid over the Cretaceous and Eocene flysch structures [77]. According to Turan et al. [100], North of the Black Sea region including this study area formed during the Cretaceous Period behind and north of the Pontide magmatic extrusive in consequence of the subduction of the northern Neo-Tethys Ocean. The researchers also reported that the Black Sea region has a rough, irregular and very heterogeneous morphology that includes steep slopes that were shaped by tectonic plate movements manifested as mostly North East-South West and North West-South East directed folding and fault systems. Geological formation of the area is represented by seven lithological units (i.e. geological explanations/age) including (1) k_2s (Clastic and carbonate rocks (flysch)—Upper Senonian); (2) e_{1-2} (Clastic and carbonate rocks—Lower Middle Eocene); (3) k_1 (Clastic and carbonate rocks—Upper Cretaceous); (4) Q (Undifferentiated quaternary—Quaternary); (5) g_4 (Granitoid—Dogger); (6) k_2e (Clastic and carbonate rocks—Upper Cretaceous—Eocene); and (7) j_3k_1 (Pelagic limestone—Upper Jurassic—Lower Cretaceous).

3 Methodology

In order to develop a model for LSM, there are many methodology or strategies can be found in the research area. The fundamental processing stages of the adopted methodology in this study comprises the six main steps:

(1) obtaining data related to landslides, preparation of landslide inventory and construction of a database of landslides, (2) preparation of the training and testing dataset, (3) feature scoring using symmetrical uncertainty, (4) determination of the best feature set based on the logistic regression model, (5) producing of LSM using RF, GBM and XGBoost models and (6) comparison of performance differences among LSMs.

The Wilcoxon signed-rank test is a non-parametric test that is used to compare two sets of scores that come from the same population [102]. Therefore, the Wilcoxon signed-rank test was employed to evaluate the statistically significant differences between the model performances. Feature selection methods (i.e. symmetrical uncertainty and logistic regression prediction) and model prediction processes (i.e. RF, GLM and XGBoost) were utilized in R Programming using the “FSelector”, “xgboost”, “gbm”, “randomForest”, and “caret” packages, respectively. ArcGIS software was used for classification of final maps and visual analysis. Flowchart of the methodology followed was given in Fig. 2.

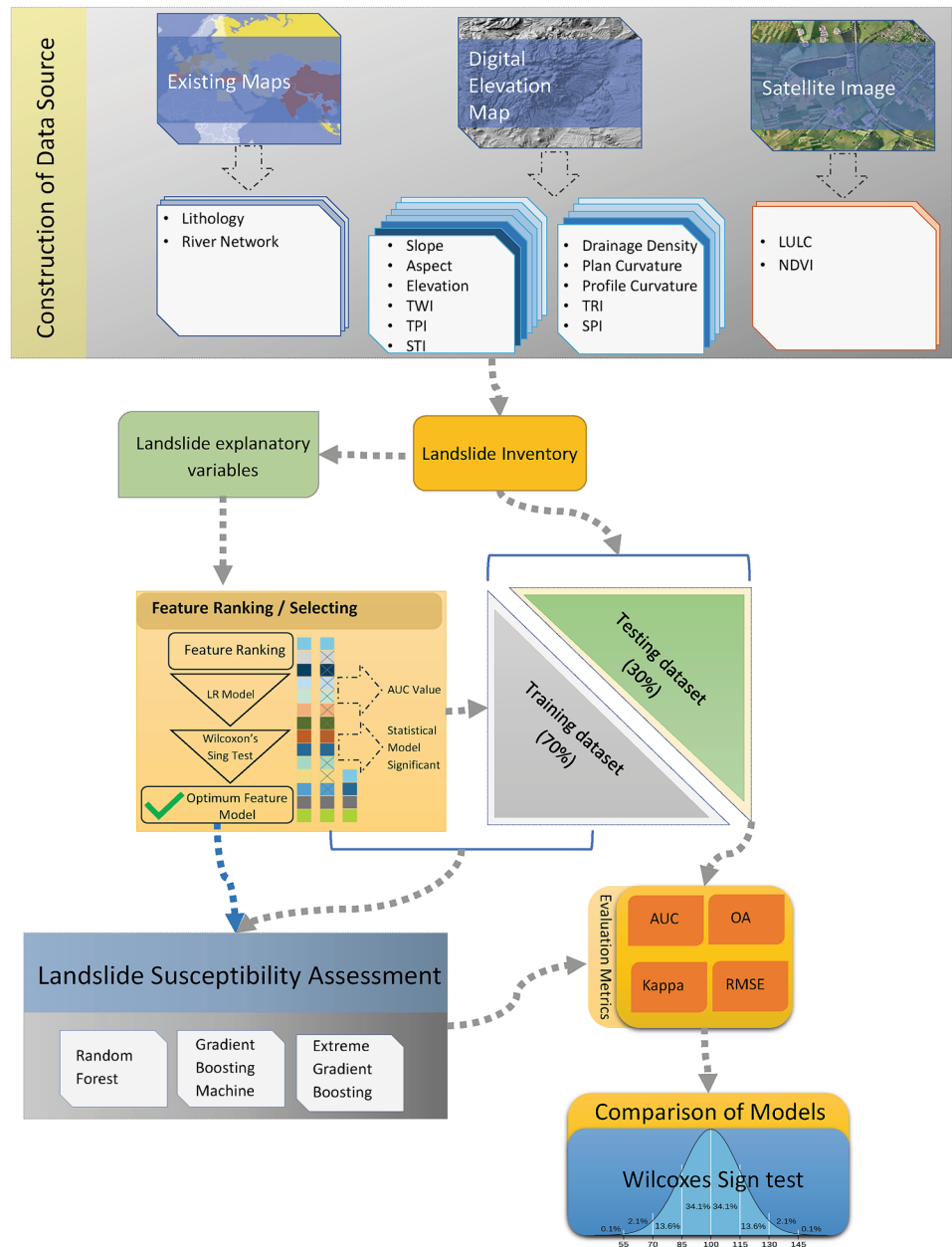
3.1 Evaluating factor importance using symmetrical uncertainty

High dimensionality problem can emerge when the features are more, or data points are very large [86]. To manage this issue feature selection (FS) techniques can be employed. In the perspective of this study, FS can be defined as the selection of the best landslide contributing factor subset which can represent the original data set. One of the most important characteristics of FS is that it is possible to conceive the proper predictive model. Hence, it helps the decision makers by preferring features that can launch the preferable performance with limited features [86]. To overcome the high dimensionality problem in landslide related factor set, Symmetrical Uncertainty (SU) is applied and reduced the feature set. SU is applied to analyze the degree of association between discrete features based on hypothesis that a good feature subset is on that contains features highly correlate with the class, yet uncorrelated to each other. SU is derived from entropy. SU is defined as follows:

$$SU = 2 * \frac{H(X) + H(Y) - (X, Y)}{H(X) + H(Y)} \quad (1)$$

This equation indicates that $H(X)$ and $H(Y)$ represent the entropy of features X and Y . This values within the range of 0 and 1. The numeric value of 1 state that one variable (either X or Y) completely predicts the other variable [60].

Fig. 2 Flow chart of the overall methodology



3.2 Determination of optimum feature subset

Generally, a smaller number of features are encouraged to generate the model, because it reduces the complexity of the computation, and also simple model is easy to understand [86]. Feature importance values for the SU method was calculated for each factor, and factors have listed in descending order according to their weights. After obtaining SU values for each factor, model performances including 2–15 factors were evaluated using LR to analyze the effect of varying number of factors on susceptibility prediction. The highest accuracy of the model was investigated by using the overall accuracy with the mentioned

analysis. After that, the goodness of the LR-based models was confirmed by the Wilcoxon signed-rank test and the most effective factor subset was determined [52, 61].

LR, which has been widely preferred as a benchmark method in the literature, was used to produce LSMs from the factor ranking obtained by SU application. Another important reason for the used LR method was an easy operation tool for LSM assessment [105]. Regression analysis is a statistical technique for building a model prediction which investigates the relationship between the presence or absence of dependent variable (i.e. landslide inventory) and a set of independent variables (landslide causative factors such as slope, lithology, etc.).

4 Ensemble methods for LSM

In general, ensemble methods that build multiple models to produce better predictive performance compared to a single model. This is done to make more accurate solutions than a single model would which incorporates the predictions from all the base learners. In this paper, methods of ensemble learning including RF, GBM, and XGBoost were used, and their results compared. When landslide susceptibility indexes have positive and negative skewness, the best classification methods are quantile or natural break approaches [3, 20]. In this present study, the produced LSMs were subdivided into five susceptibility levels by the quantile-based classification approach according to the histogram of data distribution.

4.1 Random forest

Random forest (RF), introduced by Breiman [7]. One of the important points to be mentioned for the RF algorithm can be used to perform both classification and regression tasks. The algorithm starts at the root node of a tree considering the all data. Each predictor variable is estimated to see how well it separates two different nodes. The tree-based method is usually a pruning process to cut the tree down to a size that is less likely to overfit the data and that process generally accomplished by cross-validation [8, 39]. For the implementation of the RF, it is required to be set two main parameters: the number of trees (*n_{tree}*), and the number of randomly selected predictor variables (*m_{try}*).

4.2 Gradient boosting machine

As similar to RF, gradient boosting (GBM) is another technique used for performing supervised ML applications including various classification and regression problems. GBM produced a prediction model in the form of an ensemble of weak prediction models, typically decision trees [96]. GBM includes tree main components: (1) a loss function that is to be optimized, (2) a weak learner to predict; (3) an additive model to add weak learners to optimize the loss function [47]. For GBM model, there are three main tuning parameters. These parameters include maximum number of trees “*n_{tree}*”, maximum number of interactions between independent variables “*tree depth*” and learning rate also known as shrinkage [54].

4.3 Extreme gradient boosting

XGBoost is a ML system to scale up tree boosting algorithms and that has recently been the most popular and

dominating method applied in recent years. XGBoost produces a prediction model in the form of a boosting ensemble of weak classification trees by a gradient descent that optimizes the loss function [25]. The algorithm is highly effective in reducing the process time and that also can be used for both regression and classification task. The parameters of XGBoost algorithm can be separated into three categories as suggested by Chen et al. [17], namely, *General Parameters*, *Task Parameters*, and *Booster Parameters*. In this study, there are three general parameters chosen to adjust in XGBoost algorithm for application of LSM: *colsample_bytree* (subsample ratio of columns when constructing each tree), *subsample* (subsample ratio of the training instance) and *nrounds* (max number of boosting iterations).

5 Results and discussion

5.1 Landslide conditioning factor analysis

In accordance with the proposed methodology, first, factor database including 15 causative factors and a inventory map (including landslide and non-landslide area) was constructed. Correlation between the landslide conditioning factors in LSM analysis can be high due to the most of them derived from same source (i.e. DEM). Therefore, feature selection procedure was utilized to eliminate correlated features and to construct the best subset including selected factors. For this purpose, to detect and select the best subset factors among the used 15 selected variables, the SU measure and related rank values was used in this study. As it is mentioned, feature importance values for each factor estimated by the SU were given in Table 2. In the feature importance analyses, the high value estimated by SU in the table indicates the high importance level of a feature for LSM. When the results are analyzed, the slope has the highest value (0.540), followed by elevation (0.354), TWI (0.334), STI (0.329), drainage density (0.326), lithology (0.312), NDVI (0.237) and LULC (0.226), respectively. According to estimated importance value, on the other hand, SPI, aspect, distance to river, TRI, TPI, plan curvature and profile curvature were selected as the less effective factors for LSM problem considered dataset used in this study.

As the second step of the proposed methodology, for the purpose of analyzing the best feature subset proposed by the SU ranking results, variations in the LR performances of each factor sets (including 2–15 factors) were analyzed using AUC values (Table 3). The table was four section namely Model, Number of Factors, Selected factors and AUC. Model section in table was indicated that model produced by LR and number (i.e. *_2*, *_3* etc.) represents

Table 2 Feature importance values estimated by SU algorithm

| Factors | Value |
|-------------------|----------|
| Slope | 0.54086 |
| Elevation | 0.353783 |
| TWI | 0.334336 |
| STI | 0.328842 |
| Drainage density | 0.326181 |
| Lithology | 0.311614 |
| NDVI | 0.237236 |
| LULC | 0.226365 |
| SPI | 0.174027 |
| Aspect | 0.143255 |
| Distance to river | 0.069686 |
| TRI | 0.026366 |
| TPI | 0.024626 |
| Plan curvature | 0.010801 |
| Profile curvature | 0.004386 |

results within the 95% confidence interval. Results from the Wilcoxon signed-rank test are shown in Table 4. When the test results were evaluated, it was seen that the difference between the Model_8 and the following models (Model_9, Model_10, etc.) was not statistically significant. The Wilcoxon test results revealed that Model_8 including 9 factors was not statistically significant. Therefore, it can be said that instead of using whole data set (15 factors), 9 factors combination selected by the SU could be used as it produced similar prediction performance. Thus slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC and SPI were found to be main conditioning factors among all other factors (aspect, distance to river, TRI, TPI, plan curvature and profile curvature) for this study area. As a result, Model_8 has selected the best factor combination and it was used to produce LSMs by RF, GBR and XGBoost methods.

the number of factor size. For example, Model_1 indicates 2-factor combination ranked by the SU (e.g. Model_1: slope and elevation).

In addition to these accuracy measures, significance of the differences in LR model performances was analyzed through the Wilcoxon signed-rank test. If the test statistic value is less than the critical value (i.e. p value is less than 0.05 at 95% confidence interval), the difference in performances in terms of model accuracy is said to be statistically significant. In this statistical approach, if the p value is greater than or equal to the value, it was represented as bold in the table which means there are no statistically significant differences between the two

5.2 Application of random forest model (RF) for landslide susceptibility mapping

In the RF method, the number of trees (n_{tree}) and the node split ($mytry$) has been set by the analyst. The dataset contained of nine factors selected by SU analysis, so the number of input variables was set to be square root of number of input features ($k = \sqrt{m}$ variables at each split and m represents the number of input variables). There is a need to calculate or estimate the minimum number of trees to reduce the Out-Of-Bag error. The input data, which is used to generate the tree model, was classified using a set of individual trees to estimate of the test error in OOB

Table 3 All models produced by LR methods and AUC results

| Model | Number of factors | Selected factors | AUC |
|----------|-------------------|--|--------|
| Model_1 | 2 | Slope, elevation | 0.5643 |
| Model_2 | 3 | Slope, elevation, TWI | 0.6067 |
| Model_3 | 4 | Slope, elevation, TWI, STI | 0.6065 |
| Model_4 | 5 | Slope, elevation, TWI, STI, drainage density | 0.7281 |
| Model_5 | 6 | Slope, elevation, TWI, STI, drainage density, lithology | 0.7278 |
| Model_6 | 7 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI | 0.7301 |
| Model_7 | 8 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC | 0.7302 |
| Model_8 | 9 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI | 0.7326 |
| Model_9 | 10 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect | 0.7326 |
| Model_10 | 11 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river | 0.7349 |
| Model_11 | 12 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river, TRI | 0.7350 |
| Model_12 | 13 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river, TRI, TPI | 0.7463 |
| Model_13 | 14 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river, TRI, TPI, plan curvature | 0.7468 |
| Model_14 | 15 | Slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river, TRI, TPI, plan curvature, profile curvature | 0.7473 |

Table 4 Wilcoxon's signed-rank test statistic for landslide susceptibility models

| | Model_1 | Model_2 | Model_3 | Model_4 | Model_5 | Model_6 | Model_7 | Model_8 | Model_9 | Model_10 | Model_11 | Model_12 | Model_13 | Model_14 |
|----------|---------|---------------|---------------|---------------|---------|---------|---------------|---------|---------------|---------------|---------------|---------------|---------------|---------------|
| Model_1 | - | 0.3787 | 0.0451 | 0.1258 | 0.0002 | 0.0184 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Model_2 | | - | 0.3451 | 0.3195 | 0.0053 | 0.4402 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0012 |
| Model_3 | | | - | 0.6863 | 0.0046 | 0.0068 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Model_4 | | | | - | 0.0102 | 0.0235 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0004 |
| Model_5 | | | | | - | 0.0125 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0003 |
| Model_6 | | | | | | - | 0.3657 | 0.0000 | 0.9816 | 0.9911 | 0.0001 | 0.0000 | 0.0000 | 0.0001 |
| Model_7 | | | | | | | - | 0.0000 | 0.0016 | 0.1320 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |
| Model_8 | | | | | | | | - | 0.8044 | 0.7109 | 0.5054 | 0.5031 | 0.5054 | 0.8183 |
| Model_9 | | | | | | | | | - | 0.8768 | 0.4286 | 0.4255 | 0.4301 | 0.8168 |
| Model_10 | | | | | | | | | | - | 0.1270 | 0.1211 | 0.1030 | 0.5537 |
| Model_11 | | | | | | | | | | | - | 0.8895 | 0.9112 | 0.1001 |
| Model_12 | | | | | | | | | | | | - | 0.3049 | 0.0868 |
| Model_13 | | | | | | | | | | | | | - | 0.1487 |
| Model_14 | | | | | | | | | | | | | | - |
| AUC | 0.5643 | 0.6067 | 0.6065 | 0.7281 | 0.7278 | 0.7301 | 0.7302 | 0.7326 | 0.7326 | 0.7349 | 0.7350 | 0.7463 | 0.7468 | 0.7473 |

with increasing *ntree*, and then, the number of 210 trees (*ntree*) was set.

5.3 Application of gradient boosting machines (GBM) model for landslide susceptibility mapping

GBM model was utilized in the R package “gbm” in this study and is a freely available package for regression and classification purpose. The total number of trees (*ntree*) to fit is equal to 500 and in this paper, a grid search algorithm was employed to find this parameter. The shrinkage parameter which also known as the learning rate or step-size reduction in GBM, was set in this work to 0.1. Maximum number of interactions between independent variables “tree depth”, which controls the complexity of the boosted ensemble. In this method performs best, *tree depth* was selected 4. On the other hand, the remaining parameters have not been changed and default values were accepted recommended by “gbm” package.

5.4 Application of extreme gradient boosting (XGBoost) model for landslide susceptibility mapping

It could be said that some XGBoost parameters have a significant effect on the output model, including *colsample_bytree*, *subsample*, *max_depth* and *nrounds* [13]. XGBoost was implemented using the ‘xgboost’ package in the R programming and the remaining parameters were used to the default value with *nrounds* set to 10. The main parameters setup for XGBoost are: Subsample ratio of columns (*colsample_bytree*): 0.5, maximum tree depth (“*max_depth*”): 5 and proportion of data instances to grow tree (“*subsample*”): 0.7. It should be noted that, to guarantee the robustness of the results, each experiment was run ten times and reported.

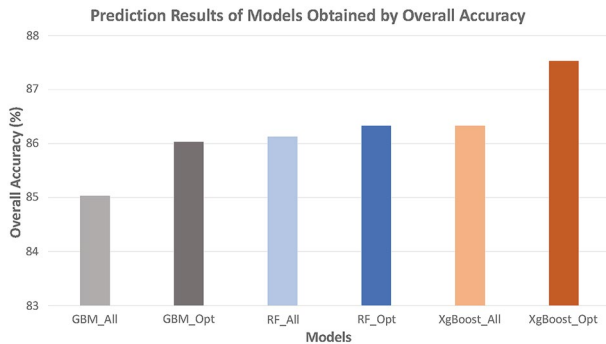
5.5 Evaluating the performance of LSM models

Performance evaluation is a curial step for comparison of several models in LSM. Without a validation, no interpretation is possible, no support of the method or of the input information can be provided and have no scientific significance [21]. Therefore, for the validation of LSMs two type of dataset as best model conditioning factors selected by SU with all factor combination were used and compared each other. In this purpose, the several metrics (OA, AUC values, RMSE and kappa coefficient) were used to evaluate models. Also, Wilcoxon signed-rank test was utilized for statistical differences between models.

Using both factor datasets (Model_8 and all factor combinations), GML, RF, and XGBoost algorithm were

Table 5 Model performance results of OA, AUC, Kappa and RMSE metrics

| | OA (%) | AUC (%) | Kappa | RMSE |
|-------------|---------|---------|---------|--------|
| GBM_All | 85.0299 | 91.8323 | 0.8163 | 0.3030 |
| GBM_Opt | 86.0279 | 92.1486 | 0.8283 | 0.2929 |
| RF_All | 86.1277 | 93.7736 | 0.8602 | 0.2643 |
| RF_opt | 86.9301 | 94.1529 | 0.8762 | 0.2487 |
| XGBoost_All | 86.3273 | 94.8689 | 0.8922 | 0.2321 |
| XGBoost_Opt | 87.5249 | 95.6816 | 0.91216 | 0.2095 |

**Fig. 3** A graphic illustration of OA results for prediction models

built using the training dataset and model's performance results were performed using the test dataset. The results are shown in Table 5 in the rows which depends on the type of percentage and value.

It could be observed that for all samples, XGBoost_Opt model (the model created by optimum factor combination) evaluated the highest performance. They are followed by the RF and GBM models for optimum factor set. For the three models constructed with SU selected subsets and the whole dataset was also applied to measure the prediction accuracy in terms of OA. OA is calculated as (sensitivity + specificity)/total number of training pixels. Figure 3 shows the accuracy test results from a confusion matrix based on OA analysis. It should be noted that very high and high classified area were considered as landslide zones, and the rest of the classified levels (i.e. moderate, low and very low) was considered as non-landslide in the accuracy assessment. The achieved OA results for the all prediction models of both landslide and non-landslide locations was above 85% (overall). The results showed that XGBoost_Opt model had a higher accuracy value (87.52%) that did all models. When the comparison results between optimum and all models were analyzed, it was shown that only RF_All and RF_Opt were had the closest accuracy results between each other as 86.12% and 86.93%, respectively. On the other hand, GBM and XGBoost models considering both data

sets had considerable accuracy differences as approximately 1% and 1.2%, respectively.

Also, RMSE and Kappa coefficient were applied to assess and compare the difference between the performance of ensemble learning methods. Kappa index value shows model compatibility powers and the reliability of the landslide models. The kappa index for all models varied from approximately 0.82–0.93 and this shows a substantial agreement between the models and the reality. RMSE is the most popular performance metrics for continuous variables can give a clear objective idea of how good a set of predictions. The lower RMSE value indicates the more accurate prediction model. RMSE values for the whole data set of GBM, RF and XGBoost ensemble learning methods were calculated as 0.3030, 0.2643 and 0.2321, respectively. Furthermore, the calculated RMSE values for optimum dataset were 0.2929, 0.2487 and 0.2095 for GBM, RF and XGBoost, respectively. When the performance of models was analyzed according to RMSE results, it could be said that XGBoost method was showed superior prediction performance both whole and optimum data sets. In addition, AUC values, which is the most widely used evaluation metric for calculating the performance, was utilized to measure the statistical differences in comparing the accuracies of LSMs. The AUC varies between 0.5 and 1 as the more the AUC, a score closes to 0.5 would represent a poor model and an AUC near to the 1 which means it has good measure of separability. AUC values obtained from ensemble models (Table 5) when analyzed it was clearly showed that both datasets (i.e. all and optimum factor sets) for XGBoost method had higher AUC values (94.87% and 95.68%, respectively). And it followed by RF_All (93.77%) and GBM_Opt (92.15%). As a result, regardless of the data structure used, these results show that the performance of models is quite acceptable.

Wilcoxon test, rank test used in nonparametric statistics, was applied to measure the statistical significance of the difference in the performance of two models. The Wilcoxon test was used in the comparison of the equality of the prediction models of two different subsets especially when the performance metric results of the LSMs is very similar and became difficult to determine exactly which model is the best decision. Results from the Wilcoxon test for the models are illustrated in Table 6. Within this confidence level, p value less than 0.05 indicates a statistically significant difference between the compared groups, otherwise, the calculated statistic is greater than the p value it is mean that there are no statistically significant differences in the parameters from groups at a 95% confidence interval. As the Table 6 shows, if the value of the calculated test statistic is greater than the p value, the value was set in bold in the table. When the Wilcoxon test results were evaluated, XGBoost_Opt model, which is the best subset

Table 6 Wilcoxon’s signed-rank test statistic for LSMs

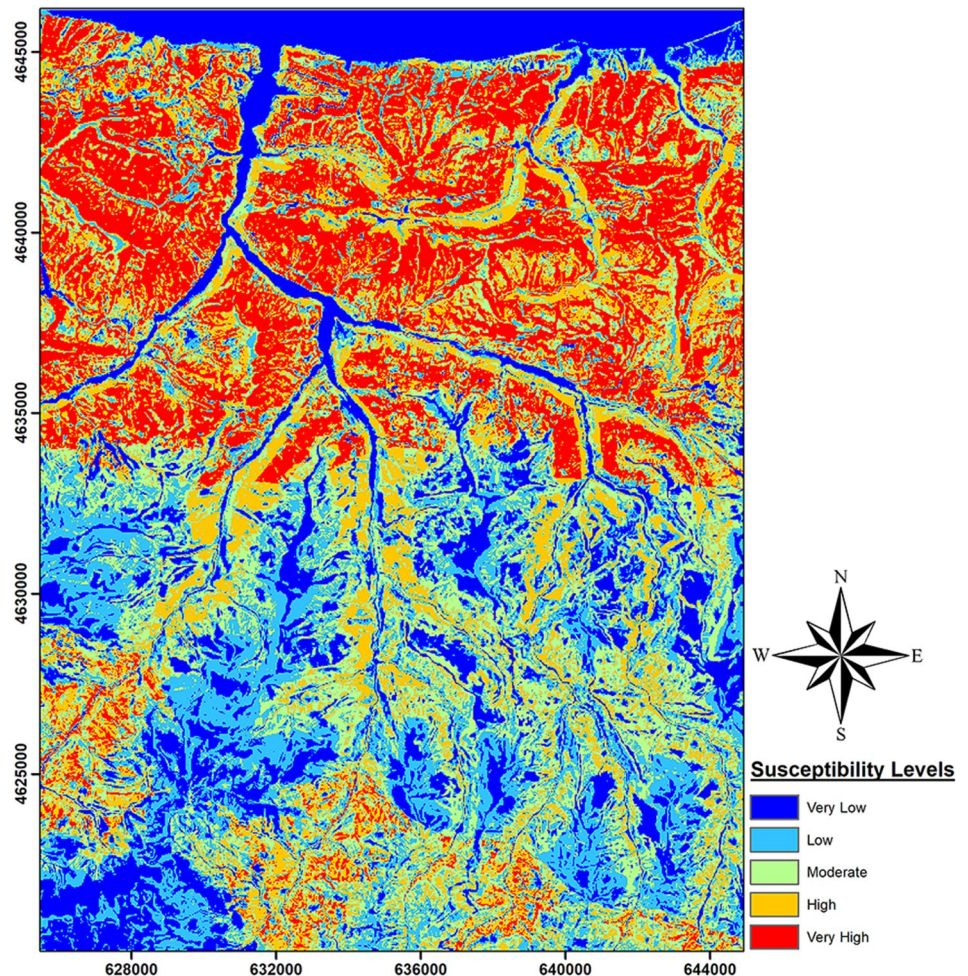
| | GBM_All | GBM_Opt | RF_All | RF_opt | XGBoost_All | XGBoost_Opt |
|-------------|---------|---------------|--------|--------|-------------|----------------|
| GBM_All | – | 0.0012 | 0.1144 | 0.0796 | 0.06153 | 0.0000 |
| GBM_Opt | | – | 0.3797 | 0.5636 | 0.9510 | 0.0002 |
| RF_All | | | – | 0.6888 | 0.2273 | 0.0001 |
| RF_opt | | | | – | 0.4047 | 0.00011 |
| XGBoost_All | | | | | – | 0.0047 |
| XGBoost_Opt | | | | | | – |

combinations, were confirmed to be statistically significant considering others model and subset combinations. These findings supported that XGBoost method with factor combination estimated by SU algorithm serves a valuable contribution according to statistically significant in producing accurate LSMs compared to GBM and RF approach considering the whole and selected dataset used in the study.

In this study, the LMS continuous map of XGBoost_Opt was divided by the quantile classification into 5 susceptibility levels including very low, low, moderate, high and very high which are shown in Fig. 4. According to XGBoost_Opt derived LSM, the very high susceptible

zones yielded about 21.92% of the total area, while about 18.48% was classified as a high susceptibility and 19.84% of the area as a moderate susceptibility. Also, 39.76% of the study area is classified as a low and very low susceptibility zone. When this LSM was analyzed visually, almost all north parts of the area were mainly located on the very high and high susceptible zones. In addition, a small part to the south of the study area was mainly covered by high and very high susceptibility zone. In addition to this, the remaining susceptible areas (i.e. very low and low) were mainly located along the from east to west-central part of the map.

Fig. 4 LSM prepared by the XGBoost model and produced from 9-factor model determined by SU algorithm



In recent years, ensemble machine learning algorithms have been proposed in the literature as alternatives to traditional statistical methods for assessment of LSM. The past studies are carefully investigated, machine learning-based approaches have shown outperformed compared to others [36, 89]. Additionally, recently proposed ensemble tree learning methods such as Random Subspace [9, 73], Rotation Forest [71, 85], Boosting and Bagging [38, 98] help to improve machine learning prediction process by combining several approaches. In this study, a new member of ensemble tree methods as XGBoost and GBM was proposed. Also, traditional and mostly used RF method was used as a benchmark algorithm. XGBoost method has outperformed the other models for the prediction of LSM as similar to different research studies [33, 106]. The performance of the RF and GBM results are in agreement with several studies [43, 61, 95]. Also, the findings of the present study have shown similar performance and parallel to previous findings given by previous studies. For example, Nhu et al. [72] applied AdaBoost, bagging, and random subspace algorithms in Shoor River watershed in northwestern Iran. The results indicated that all ensembles models provided a high goodness-of-fit and prediction accuracy. Pham et al. [81] was proposed various ensemble predictive ML models (i.e. Random Subspace with Best First Decision Tree, Functional Tree (RSSFT), J48 Decision Tree, Naive Bayes Tree and Reduced Error Pruning Trees). Mentioned study results showed that the RSSFT model achieved the highest performance in terms of predicting future landslides and it followed by the other ensemble methods. In another study proposed by Sevgen et al. [91], statistical and machine learning methodologies (i.e. LR, artificial neural network (ANN) and RF) have been applied for the production of landslide susceptibility maps. Among the LR and ANN methods, ensemble learning method based on RF is showed better prediction performance to the other two for predicting the future landslides.

6 Conclusions

In this present study introduced a conceptual framework of tree-based ensemble methods in three different approaches, (i.e. XGBoost, GBM, and RF) for the LSM in Ayancik district, Sinop province in the Black Sea region of Turkey. In order to achieve a good quality of LSM model, selecting landslide causal factors based on their importance was found by SU method. For this present study, fifteen landslide conditioning factors for the analysis (slope, elevation, TWI, STI, drainage density, lithology, NDVI, LULC, SPI, aspect, distance to river, TRI, TPI, plan curvature, profile curvature) was used and Model_8 including 9 factors (slope, elevation, TWI, STI, drainage density, lithology,

NDVI, LULC and SPI) was evaluated using the Wilcoxon Sign Test integrated with LR method to produce landslide susceptibility maps. Tree-Based Ensembles Methods constructed with SU-selected subset and whole dataset were applied to the testing data to measure the prediction accuracy in terms of performance metrics such as OA, AUC, RMSE and Kappa coefficient. According to this case study, the performance metrics demonstrates that there was generally little differentiation in prediction performance between method applied to the Model_8 and whole data set. In terms of validation results, all the produced models (i.e. XGBoost, GBM, and RF) show reasonably good performances and results also stated that the XGBoost has the best predictive capability among the other models (i.e. Rf and GBM). Therefore, due to differences in accuracy of predictions, the Wilcoxon test was applied to determine the optimal map production method for factor subset options. As a result, result map of the XGBoost_Opt was significantly different from the GBM and RF maps (including all and optimum factor sets). Also, feature selection approach as a SU was proved to be an effective solution for this study. The novelty of the research is the use of the SU algorithm in factor ranking and to obtain new models by using the LR method according to factor sets selected by SU and to determine the optimum model among these models by Wilcoxon sign-rank test. In summary, results revealed the robustness of produced landslide susceptibility maps using the tree-based ensemble methods when using the optimal landslide causative factors among too many factors.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest regarding the publication of this article.

References

1. Akgun A (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at Izmir, Turkey. *Landslides* 9:93–106. <https://doi.org/10.1007/s10346-011-0283-7>
2. Akgun A, Dag S, Bulut F (2008) Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ Geol* 54:1127–1143. <https://doi.org/10.1007/s00254-007-0882-8>
3. Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci* 36:23–34. <https://doi.org/10.1016/j.cageo.2011.04.012>
4. Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the

- Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15–31. <https://doi.org/10.1016/j.geomorph.2004.06.010>
5. Ayalew L, Yamagishi H, Marui H, Kanno T (2005) Landslides in Sado Island of Japan: part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng Geol* 81:432–445. <https://doi.org/10.1016/j.enggeo.2005.08.004>
 6. Ayalew L, Yamagishi H, Ugawa N (2004) Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides* 1:73–81. <https://doi.org/10.1007/s10346-003-0006-9>
 7. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
 8. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. Wadsworth, Belmont
 9. Bui DT et al (2019) New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. *Forests*. <https://doi.org/10.3390/f10090743>
 10. Bui DT, Tuan TA, Hoang ND, Thanh NQ, Nguyen DB, Liem NV, Pradhan B (2017) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* 14:447–458. <https://doi.org/10.1007/s10346-016-0711-9>
 11. Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) 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* 13:361–378. <https://doi.org/10.1007/s10346-015-0557-6>
 12. Burrough PA (1988) Principles of geographical information-systems for land resources assessment. Clarendon Press, Oxford
 13. Cao K, Guo H, Zhang Y (2019) Comparison of approaches for urban functional zones classification based on multi-source geospatial data: a case study in Yuzhong District, Chongqing, China. *Sustainability*. <https://doi.org/10.3390/su11030660>
 14. Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazard Earth Syst* 13:2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
 15. Caves R (1982) Multinational enterprise and economic analysis. Cambridge University Press, Cambridge
 16. Chen G, Meng XM, Tan L, Zhang FY, Qiao L (2014) Comparison and combination of different models for optimal landslide susceptibility zonation. *Q J Eng Geol Hydrogeol* 47:283–306. <https://doi.org/10.1144/qjgeh2013-071>
 17. Chen T, He T, Benesty M (2016) XGBoost: extreme gradient boosting. R package version 4-3
 18. Chen W, Pourghasemi HR, Kornejady A, Zhang N (2017) Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 305:314–327. <https://doi.org/10.1016/j.geoderma.2017.06.020>
 19. Chen W, Shirzadi A, Shahabi H, Bin Ahmad B, Zhang S, Hong HY, Zhang N (2017) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naive Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat Nat Hazard Risk* 8:1955–1977. <https://doi.org/10.1080/19475705.2017.1401560>
 20. Chen W, Sun ZH, Han JC (2019) Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl Sci Basel*. <https://doi.org/10.3390/app9010171>
 21. Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30:451–472. <https://doi.org/10.1023/B:NHAZ.00000007172.62651.2b>
 22. Clerici A, Perego S, Tellini C, Vescovi P (2002) A procedure for landslide susceptibility zonation by the conditional analysis method. *Geomorphology* 48:349–364. [https://doi.org/10.1016/S0169-555x\(02\)00079-X](https://doi.org/10.1016/S0169-555x(02)00079-X)
 23. Comert R, Avdan U, Gorum T, Nefeslioglu HA (2019) Mapping of shallow landslides with object-based image analysis from unmanned aerial vehicle data. *Eng Geol*. <https://doi.org/10.1016/j.enggeo.2019.105264>
 24. Costanzo D, Rotigliano E, Irigaray C, Jimenez-Peralvarez JD, Chacon J (2012) Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain). *Nat Hazard Earth Syst* 12:327–340. <https://doi.org/10.5194/nhess-12-327-2012>
 25. Cui Y, Cai M, Stanley HE (2017) Comparative analysis and classification of cassette exons and constitutive exons. *Biomed Res Int*. <https://doi.org/10.1155/2017/7323508>
 26. Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR, Althuwaynee OF (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Nat Hazards* 65(1):135–165. <https://doi.org/10.1007/s11069-012-0347-6>
 27. Dhakal AS, Amada T, Aniya M (2000) Landslide hazard mapping and its evaluation using GIS: an investigation of sampling schemes for a grid-cell based quantitative method. *Photogram Eng Remote Sens* 66:981–989
 28. Dong H, Xu X, Wang L, Pu FL (2018) Gaofen-3 PolSAR image classification via XGBoost and polarimetric spatial information. *Sensors*. <https://doi.org/10.3390/s18020611>
 29. Dou J et al (2015) Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0133262>
 30. Ercanoglu M (2005) Landslide susceptibility assessment of SE Bartin (West Black Sea region, Turkey) by artificial neural networks. *Nat Hazard Earth Syst* 5:979–992. <https://doi.org/10.5194/nhess-5-979-2005>
 31. Erener A, Duzgun HSB (2010) Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides* 7:55–68. <https://doi.org/10.1007/s10346-009-0188-x>
 32. Erener A, Duzgun HSB (2012) Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environ Earth Sci* 66:859–877. <https://doi.org/10.1007/s12665-011-1297-0>
 33. Gholami H, Mohamadifar A, Collins AL (2020) Spatial mapping of the provenance of storm dust: application of data mining and ensemble modelling. *Atmos Res*. <https://doi.org/10.1016/j.atmosres.2019.104716>
 34. Glade T (2001) Landslide hazard assessment and historical landslide data—an inseparable couple? In: Use of historical data in natural hazard assessments, vol 17, pp 153–168. 10.1007/978-94-017-3490-5_12
 35. Glade T (2003) Landslide occurrence as a response to land use change: a review of evidence from New Zealand. *CATENA* 51:297–314. [https://doi.org/10.1016/S0341-8162\(02\)00170-4](https://doi.org/10.1016/S0341-8162(02)00170-4)
 36. Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>
 37. Gomez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Eng Geol* 78:11–27. <https://doi.org/10.1016/j.enggeo.2004.10.004>
 38. Gomez-Rios A, Luengo J, Herrera F (2017) A study on the noise label influence in boosting algorithms: AdaBoost, GBM and

- XGBoost. Hybrid Artif Intell Syst Hais 10334:268–280. https://doi.org/10.1007/978-3-319-59650-1_23
39. Gould LA (2014) Statistical methods for evaluating safety in medical product development. Wiley, Hoboken
 40. Gupta SK, Shukla DP, Thakur M (2018) Selection of weightages for causative factors used in preparation of landslide susceptibility zonation (LSZ). *Geomat Nat Hazard Risk* 9:471–487. <https://doi.org/10.1080/19475705.2018.1447027>
 41. Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216. [https://doi.org/10.1016/S0169-555x\(99\)00078-1](https://doi.org/10.1016/S0169-555x(99)00078-1)
 42. Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT (2012) Landslide inventory maps: new tools for an old problem. *Earth Sci Rev* 112:42–66. <https://doi.org/10.1016/j.earsci.2012.02.001>
 43. Hamzehpour N, Shafizadeh-Moghadam H, Valavi R (2019) Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. *CATENA*. <https://doi.org/10.1016/j.catena.2019.104141>
 44. Highland L, Bobrowski PT (2008) The landslide handbook: a guide to understanding landslides. US Geological Survey, Reston
 45. Hong HY, Miao YM, Liu JZ, Zhu AX (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *CATENA* 176:45–64. <https://doi.org/10.1016/j.catena.2018.12.035>
 46. Hong HY, Pradhan B, Sameen MI, Chen W, Xu C (2017) Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China). *Geomat Nat Hazard Risk* 8:1997–2022. <https://doi.org/10.1080/19475705.2017.1403974>
 47. Huynh XP, Park SM, Kim YG (2017) Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine computer vision. In: *Accv 2016 workshops*, Pt Iii, vol 10118, pp 134–145. [10.1007/978-3-319-54526-4_10](https://doi.org/10.1007/978-3-319-54526-4_10)
 48. Kadavi PR, Lee CW, Lee S (2018) Application of ensemble-based machine learning models to landslide susceptibility mapping. *Remote Sens*. <https://doi.org/10.3390/rs10081252>
 49. Kanungo DP, Arora MK, Sarkar S, Gupta RP (2009) Landslide susceptibility zonation (LSZ) mapping: a review. *J South Asia Disaster Stud* 2:81–105
 50. Kavzoglu T, Colkesen I, Sahin EK (2019) Machine learning techniques in landslide susceptibility mapping: a survey and a case study. *Landslides Theory Pract Model* 50:283–301. https://doi.org/10.1007/978-3-319-77377-3_13
 51. Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11:425–439. <https://doi.org/10.1007/s10346-013-0391-7>
 52. Kavzoglu T, Sahin EK, Colkesen I (2015) Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Eng Geol* 192:101–112. <https://doi.org/10.1016/j.enggeo.2015.04.004>
 53. Khamehchiyan M, Abdolmaleki P, Rakei B (2011) Landslide susceptibility mapping using backpropagation neural networks and logistic regression: the Sephidargole case study, Semnan, Iran. *Geomech Geoeng* 6:237–250. <https://doi.org/10.1080/17486025.2011.560289>
 54. Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York
 55. Kutlug Sahin E, Colkesen I (2019) Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping. *Geocarto Int*. <https://doi.org/10.1080/10106049.2019.1641560>
 56. Kutlug Sahin E, Ipbuker C, Kavzoglu T (2017) Investigation of automatic feature weighting methods (Fisher, Chi-square and Relief-F) for landslide susceptibility mapping. *Geocarto Int* 32:956–977. <https://doi.org/10.1080/10106049.2016.1170892>
 57. Liu XY, Zhang ZY, Peterson J, Chandra S (2008) Large area DEM generation using Airborne LiDAR data and quality control. In: *Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, vol Ii, pp 79–85
 58. Magliulo P, Di Lisio A, Russo F, Zelano A (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. *Nat Hazards* 47:411–435. <https://doi.org/10.1007/s11069-008-9230-x>
 59. Mandal S, Mandal K (2018) Bivariate statistical index for landslide susceptibility mapping in the Rorachu river basin of eastern Sikkim Himalaya, India. *Spat Inf Res* 26:59–75. <https://doi.org/10.1007/s41324-017-0156-9>
 60. Meesad P, Sodsee S, Unger H (2017) Recent advances in information and communication technology 2017. In: *Proceedings of the 13th international conference on computing and information technology (IC2IT)*, Bangkok, Thailand, 6–7 July, 2017. *Advances in intelligent systems and computing*, vol 566. Springer. <https://doi.org/10.1007/978-3-319-60663-7>
 61. Merghadi A, Abderrahmane B, Bui DT (2018) Landslide susceptibility assessment at Mila Basin (Algeria): a comparative assessment of prediction capability of advanced machine learning methods. *ISPRS Int Geo Inf*. <https://doi.org/10.3390/ijgi7070268>
 62. Meten M, PrakashBhandary N, Yatabe R (2015) Effect of landslide factor combinations on the prediction accuracy of landslide susceptibility maps in the Blue Nile Gorge of Central Ethiopia. *Geoenviromn Disasters* 2:9. <https://doi.org/10.1186/s40677-015-0016-7>
 63. Mezaal MR, Pradhan B (2018) Data mining-aided automatic landslide detection using airborne laser scanning data in densely forested tropical areas Korean. *J Remote Sens* 34:45–74. <https://doi.org/10.7780/kjrs.2018.34.1.4>
 64. Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2014) Machine learning feature selection methods for landslide susceptibility mapping. *Math Geosci* 46:33–57. <https://doi.org/10.1007/s11004-013-9511-0>
 65. Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling—a review of hydrological, geomorphological, and biological. *Appl Hydrol Process* 5:3–30. <https://doi.org/10.1002/hyp.3360050103>
 66. Motevalli A, Pourghasemi HR, Zabihi M (2018) Assessment of GIS-based machine learning algorithms for spatial modeling of landslide susceptibility: case study in Iran. *Compr Geogr Inf Syst*. <https://doi.org/10.1016/b978-0-12-409548-9.10461-0>
 67. Mutasem SA, Bt Ngah UK, Tay LT, Mat Isa NAB (2012) Landslide susceptibility hazard mapping techniques review. *J Appl Sci* 12(9):802–808. <https://doi.org/10.3923/jas.2012.802.808>
 68. Naghbi SA, Moghaddam DD, Kalantar B, Pradhan B, Kisi O (2017) A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *J Hydrol* 548:471–483. <https://doi.org/10.1016/j.jhydrol.2017.03.020>
 69. Nefeslioglu HA, Gokceoglu C (2011) Probabilistic risk assessment in medium scale for rainfall-induced earthflows: Catakli catchment area (Cayeli, Rize, Turkey). *Math Probl Eng*. <https://doi.org/10.1155/2011/280431>
 70. Nefeslioglu HA, San BT, Gokceoglu C, Duman TY (2012) An assessment on the use of Terra ASTER L3A data in landslide

- susceptibility mapping. *Int J Appl Earth Obs* 14:40–60. <https://doi.org/10.1016/j.jag.2011.08.005>
71. Nguyen VV et al (2019) Hybrid machine learning approaches for landslide susceptibility modeling. *Forests*. <https://doi.org/10.3390/f10020157>
 72. Nhu V-H et al (2020) Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran. *Forests* 11:421. <https://doi.org/10.3390/f11040421>
 73. Nhu V-H et al (2020) GIS-based gully erosion susceptibility mapping: a comparison of computational ensemble data mining models. *Appl Sci* 10:2039
 74. Ocakoglu F, Gokceoglu C, Ercanoglu M (2002) Dynamics of a complex mass movement triggered by heavy rainfall: a case study from NW Turkey. *Geomorphology* 42:329–341. [https://doi.org/10.1016/S0169-555x\(01\)00094-0](https://doi.org/10.1016/S0169-555x(01)00094-0)
 75. Oh HJ, Kadavi PR, Lee CW, Lee S (2018) Evaluation of landslide susceptibility mapping by evidential belief function, logistic regression and support vector machine models. *Geomat Nat Hazard Risk* 9:1053–1070. <https://doi.org/10.1080/19475705.2018.1481147>
 76. Oh HJ, Lee S (2017) Shallow landslide susceptibility modeling using the data mining models artificial neural network and boosted tree. *Appl Sci*. <https://doi.org/10.3390/app7101000>
 77. Ozdemir N (2005) One of the effective natural disaster in Sinop: the landslide. *J Ziya Gokalp Fac Educ* 5:67–106
 78. Pardeshi SD, Autade SE, Pardeshi SS (2013) Landslide hazard assessment: recent trends and techniques. Springerplus. <https://doi.org/10.1186/2193-1801-2-523>
 79. Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl Sci*. <https://doi.org/10.3390/app9050942>
 80. Pellicani R, Van Westen CJ, Spilotro G (2014) Assessing landslide exposure in areas with limited landslide information. *Landslides* 11:463–480. <https://doi.org/10.1007/s10346-013-0386-4>
 81. Pham BT et al (2020) Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int*. <https://doi.org/10.1080/10106049.2020.1737972>
 82. Pham BT, Bui DT, Prakash I, Dholakia MB (2017) Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* 149:52–63. <https://doi.org/10.1016/j.catena.2016.09.007>
 83. Pham BT, Jaafari A, Prakash I, Bui DT (2019) A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull Eng Geol Environ* 78:2865–2886. <https://doi.org/10.1007/s10064-018-1281-y>
 84. Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ Model Softw* 84:240–250. <https://doi.org/10.1016/j.envsoft.2016.07.005>
 85. Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: hybrid machine learning approaches. *CATENA* 175:203–218. <https://doi.org/10.1016/j.catena.2018.12.018>
 86. Potharaju SP, Sreedevi M (2018) A novel cluster of quarter feature selection based on symmetrical uncertainty. *Gazi Univ J Sci* 31:456–470
 87. Pradhan AMS, Kang HS, Kim YT (2017) Hybrid Landslide Warning Model for Rainfall Triggered Shallow Landslides in Korean Mountain. In: *Advancing culture of living with landslides*, vol 3: advances in landslide technology, pp 193–200. https://doi.org/10.1007/978-3-319-53487-9_22
 88. Pradhan B, Seeni MI, Kalantar B (2017) Performance evaluation and sensitivity analysis of expert-based, statistical, machine learning, and hybrid models for producing landslide susceptibility maps. In: Pradhan B (ed) *Laser scanning applications in landslide assessment*. Springer, Cham, pp 193–232. https://doi.org/10.1007/978-3-319-55342-9_11
 89. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
 90. Sahin EK, Colkesen I, Kavzoglu T (2020) A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto Int* 35:341–363. <https://doi.org/10.1080/10106049.2018.1516248>
 91. Sevgen E, Kocaman S, Nefeslioglu HA, Gokceoglu C (2019) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random Forest. *Sensors*. <https://doi.org/10.3390/s19183940>
 92. Shahabi H, Khezri S, Bin Ahmad B, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. *CATENA* 115:55–70. <https://doi.org/10.1016/j.catena.2013.11.014>
 93. Shirzadi A et al (2019) Uncertainties of prediction accuracy in shallow landslide modeling: sample size and raster resolution. *CATENA* 178:172–188. <https://doi.org/10.1016/j.catena.2019.03.017>
 94. Shirzadi A, Soliamani K, Habibnejhad M, Kaviani A, Chapi K, Shahabi H, Chen W, Khosravi K, Pham T, Pradhan B et al (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* 18(11):3777
 95. Song YX, Niu RQ, Xu SL, Ye RQ, Peng L, Guo T, Li SY, Chen T (2019) Landslide susceptibility mapping based on weighted gradient boosting decision tree in Wanzhou section of the three gorges reservoir area (China). *ISPRS Int Geo Inf*. <https://doi.org/10.3390/ijgi8010004>
 96. Strickland J (2016) *Data analytics using open-source tools*. Lulu.com
 97. Thongkam J, Xu GD, Zhang YC (2008) AdaBoost algorithm with random forests for predicting breast cancer survivability. In: 2008 IEEE international joint conference on neural networks, vol 1–8, pp 3062–3069. <https://doi.org/10.1109/ijcnn.2008.4634231>
 98. Tien Bui D, Ho T-C, Pradhan B, Pham B-T, Nhu V-H, Revhaug I (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ Earth Sci* 75:1101. <https://doi.org/10.1007/s12665-016-5919-4>
 99. Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and Naive Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *CATENA* 145:164–179. <https://doi.org/10.1016/j.catena.2016.06.004>
 100. Turan ID, Ozkan B, Turkes M, Dengiz O (2020) Landslide susceptibility mapping for the Black Sea Region with spatial fuzzy multi-criteria decision analysis under semi-humid and humid terrestrial ecosystems. *Theor Appl Climatol* 140:1233–1246. <https://doi.org/10.1007/s00704-020-03126-2>
 101. Vilhelmsen TN, Maher K, Da Silva C, Hermans T, Grujic O, Park J, Yang G (2018) Quantifying uncertainty in subsurface systems. *Geophys Monogr Ser* 236:217–262. <https://doi.org/10.1002/9781119325888>

102. Wilcoxon F (1965) Individual comparisons by ranking methods. *Biom Bull* 1(6):80–83
103. Wilson JP, Gallant JC (2000) *Terrain analysis: principles and applications*. Wiley, Chichester
104. Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *CATENA* 85:274–287. <https://doi.org/10.1016/j.catena.2011.01.014>
105. Yao X, Zhang Y, Zhou N, Guo C, Yu K, Li LJ (2014) Project planning and project success: the 25% solution. In: *Application of two-class SVM applied in landslide susceptibility mapping*. Taylor & Francis Group, England, UK, p 203
106. Zhang YH, Ge TT, Tian W, Liou YA (2019) Debris flow susceptibility mapping using machine-learning techniques in shigatse Area. *China Remote Sens*. <https://doi.org/10.3390/rs11232801>
107. Zhou C, Yin KL, Cao Y, Ahmed B, Li YY, Catani F, Pourghasemi HR (2018) Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. *Comput Geosci* 112:23–37. <https://doi.org/10.1016/j.cageo.2017.11.019>

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