Artificial intelligence approaches for spatial prediction of landslides in mountainous regions of western India

PANKAJ PRASAD (✉ ppankaj.earthscience@gmail.com)

National Institute of Oceanography CSIR <https://orcid.org/0000-0002-3118-2201>

Victor Joseph Loveson

National Institute of Oceanography CSIR

Sumit Das

University of Pune: Savitribai Phule Pune University

Priyankar Chandra

Banaras Hindu University

Research Article

Keywords: Landslide susceptibility, Boruta, GIS, Machine learning, Western India

Posted Date: August 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-251440/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Artificial intelligence approaches for spatial prediction of landslides in mountainous regions of western India

Pankaj Prasad^{1*}, Victor Joseph Loveson¹, Sumit Das², Priyankar Chandra³

¹Geological Oceanography Division, CSIR- National Institute of Oceanography, Dona Paula-403004, Goa, India

²Department of Geography, Savitribai Phule Pune University, Pune 411007, Maharashtra, India

³Department of Geography, Institute of Science, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, India

***Corresponding Author**

Pankaj Prasad

Email: ppankaj.earthscience@gmail.com; ppankaj@nio.org

Mobile No.: +919883329432

Acknowledgement

We acknowledge the support from the University Grant Commission (3160 NET-June 2015) to the first author. The authors are thankful to the Director CSIR-NIO for constant encouragement. The authors are also grateful to the reviewer for critical comments and constructive suggestions that improved the manuscript significantly. Field support from the survey team members is thankfully acknowledged. The NIO contribution number is.....

Abstract

The prediction of landslide is a complex task but preparing the landslide susceptibility map through artificial intelligence approaches can reduce life loss and damages resulting from landslides. The purpose of this study is to evaluate and compare the landslide susceptibility mapping (LSM) using six machine learning models, including random forest (RF), deep boost (DB), stochastic gradient boosting (SGB), rotation forest (RoF), boosted regression tree (BRT), and logit boost (LB) in the mountainous regions of western India. The landslide inventory map consisting of 184 landslide locations has been divided into two groups for training (70% dataset) and validation (30% dataset) purposes. Fourteen landslide triggering factors including slope, topographical roughness index, road density, topographical wetness index, elevation, slope length, drainage density, stream power index, geomorphology, rainfall, soil, lithology, lineament density, and normalized difference vegetation index have been considered using the boruta approach for the LSM. The results reveal that the RF model has the highest precision in terms of AUC (0.88; 0.89), kappa (0.62; 0.50), accuracy (0.81; 0.77), and specificity (0.86; 0.86) both in the study region and secondary region, respectively. Hence, it can be concluded that the RF is an effective and promising technique as compared to DB, SGB, RoF, BRT, and LB for landslide susceptibility assessment in the research area as well as in regions having similar geo-environmental configuration.

Keywords: Landslide susceptibility; Boruta; GIS; Machine learning; Western India

1. Introduction

Globally, the frequency of natural disasters has been dramatically increased over the last few decades (UNISDR, 2015). Each year, a loss of billions of dollars and thousands of lives due to natural disasters, such as earthquakes, tsunamis, cyclones, floods, and landslides are recorded (Yesilnacar and Topal, 2005; Yilmaz, 2009). Almost there is no region in this earth which is free from the impact of such devastating natural hazards (Suzen and Doyuran, 2003). The occurrence of landslides constitutes about 17% of the natural disasters worldwide as reported by the Center for Research on the Epidemiology of Disasters (Pourghasemi et al., 2012; Nohani et al., 2019). Yilmaz (2009) and Pourghasemi et al. (2012) believe that landslide events will tend to rise in the next decades owing to deforestation, land-use change, and changing climatic condition. Many scientists have assessed landslide hazards and delineated the high-risk zones in past few decades.

In general, landslide mapping techniques can be divided into two major methods, i.e. direct and indirect. In case of direct mapping, geomorphologists determine the degree of susceptibility based on the experience and knowledge of terrain characteristics, hillslopes etc. (Van et al., 2003). However, indirect mapping utilizes different statistical or high-performance deterministic machine learning or hybrid models to predict the landslide probable areas on the basis of previous landslide distribution. In the preceding two decades, multitudes of LSM studies have been undertaken with increasing popularity of Geographic Information System (Aleotti and Chowdury, 1999; Guzzetti et al., 2006).

The pioneering works of the landslide mapping were done in 1970s (Carrara et al., 1977; Brabb et al., 1978). Since then, various approaches and methods have been put forwarded and employed for LSM, including heuristic terrain and susceptibility zoning, geomorphological mapping, statistical modelling, physically-based numerical modelling etc. (Montgomery and Dietrich, 1994; Hansen et al., 1995; Aleotti and Chowdhury, 1999; Guzzetti et al., 1999; Lee et al., 2001; Van et al., 2008; Reichenbach et al., 2018). However, the above approaches are no longer capable because they were restricted to large amounts of field data in a smaller area. From the beginning of the current century, statistical modelling with high-resolution GIS data has been in use in different geographical regions for LSM, including frequency ratio (Yilmaz, 2009; Nohani et al., 2019), logistic regression (Pourghasemi et al., 2013; Zhao and Chen, 2020), weight of evidence (Lee et al., 2004a; Regmi et al., 2010). The statistical methods are less accurate compared to the machine learning methods because it can overcome the statistical limitations and predict non-linear structures (Tehrany et al., 2013; Costache and Bui, 2019; Prasad et al., 2020). Hence, in recent years, several machine learning models have been developed and applied successfully for LSM, including random forest (Kim et al., 2018; Arabameri et al., 2019), rotation forest (Hong et al., 2018; Pham et al., 2020a), support vector machine (Chen et al., 2017; Phong et al., 2019), boosted regression tree (Youssef et al., 2016; Park and Kim, 2019), naive bayes (Bui et al., 2012; Pham et al., 2020a), adaboost (Hong et al., 2018; Wu et al., 2020), and artificial neural network (Lee et al., 2004b; Yilmaz, 2009). Such methods have also been employed in various spatial prediction mapping, viz. flood susceptibility (Khosravi et al., 2018; Pham et al., 2020b; Shahabi et al., 2020), land subsidence (Bui et al., 2018; Oh et al., 2019), gully erosion (Amiri et al., 2019; Pourghasemi et al., 2020; Chen et al., 2021b), piping erosion (Hosseinalizadeh et al., 2019; Chen et al., 2021),

land-use change (Aburas et al., 2019; Jamali. 2020), and groundwater potentiality (Rahmati et al., 2016; Naghibi et al., 2019). Besides statistical and machine learning models, novel models with high accuracy have been also used in landslide mapping (Bui et al., 2017; Nguyen et al., 2017; Abedini et al., 2019; Chen et al., 2020; Chen and Li., 2020; Pham et al., 2020a).

In the aforementioned studies, the performance of the applied models has been assessed with respect to particular study regions. It is imperative to show the model applicability in another region for general findings (Prasad et al., 2020). To the best of our knowledge, in landslide susceptibility mapping, the models have been validated by applying them in more than one geographic region. Thus, the present study aims to map the landslide-prone zones in the western India through different approaches of artificial intelligence and to validate the models in another region. In spite of being a high landslide prone area, the present area of investigation has not received much attention from the researchers. The current study is the first attempt to prepare the landslide susceptibility map hereof using machine learning approaches. The LSM is foremost mitigative step to reduce the havoc fatalities and infrastructural damages as a dire consequence of frequent severe landslides along the precarious Ghats escarpment triggered by torrential rainfall during the Indian Summer Monsoon. In addition to the comparison of different machine learning approaches, the scientific delineation of landslide-prone regions will be helpful for proper planning and management of the study area.

2. Study area and GIS database

2.1 Study area

The study area lying approximately between 17°50 to 18°19 North latitudes and 73°18 to 73°40 East longitudes, comprises an approximate area of 1180 km² (Fig. 1). The study region represents a hilly topography ranging from 4 m to 1404 m above the mean sea level. The average slope of the area is about 15°, though some areas are close to the vertical slope. The study area shows a general tilting facing towards the Arabian Sea.

Geologically, the area spans from upper Cretaceous to lower Eocene, which comes under the Sahyadri group of formation. From the geomorphological perspective, more than fifty percent of the area belongs to structural origin highly and moderately dissected plateau. The climatic condition of the study region is subtropical monsoon type. The minimum and maximum average

temperatures of the area are 25°C and 30°C, respectively. The Arabian branch of south-west monsoon brings huge amount of cloud that leads to intense orographic rainfall along the hilly slopes. The area receives about 3000 mm annual mean rainfall during the monsoon seasons (June-September) (Department of Agriculture, Maharashtra State). The principal river Savitri, descends from the Mahabaleshwar peak of Western Ghats and flows for about 110 km before it empties into the Arabian sea.

2.2 GIS database

2.2.1 Landslide inventory map

Inventory data is crucial for any natural hazard mapping especially to understand the landslide distribution and effective variables (Kadavi et al., 2018; Panahi et al., 2020). In the present study, the landslide locations were collected from the Public Works Department (PWD) office, Google Earth historical images and intensive field survey (Fig. 2). A total of 184 landslide locations were taken into account for this study. The dataset was divided into two groups 70% (129 locations) as training data for model fitting and the rest 30% (55 locations) as testing data for model predictive capability assessment (Fig. 1). Simultaneously, an equal number of non-landslide localities were recognized from the topographical map, field survey, and satellite images for the purpose of model establishment. Likewise, the inventory map of another region was generated (Fig. 1).

2.2.1 Landslide triggering factors

The important step for the LSM is to prepare a spatial database of landslide conditioning factors (Reichaenbach et al., 2018; Nohani et al., 2019). The predisposing variables of the landslide were selected after extensive literature survey and field investigation (Hong et al., 2018; Wang et al., 2019). These factors are: slope, elevation, slope length (LS), topographical wetness index (TRI), stream power index (SPI), topographical wetness index (TWI), drainage density (DD), road density (RD), lineament density (LD), normalized difference vegetation index (NDVI), rainfall, lithology, geomorphology, and soil (Fig. 3a-3n). The terrain variables (elevation, slope, LS, SPI, TWI, and TRI) were extracted from the SRTM digital elevation model (DEM) with 30 m spatial resolution. Each conditioning variable was transformed into a spatial database of 30× 30 m size

and the grid of the study region was generated by 1337 columns and 1740 rows (1312960 pixels; 1180 km²).

The slope is one of the most important triggering variables since it directly influences the landslide occurrences (Yalchin, 2008). The slope map of the study region was categorized into five classes: 0-7.19, 7.19-15.25, 15.25-24.18, 24.18-35.98, and 35.98-73.40 degrees. The altitude of the area is classified into five classes. LS represents slope steepness and slope length of the topography, which affects the soil loss potential (Naghibi et al., 2019). The LS values are extracted by the given mathematical formula (Moore and Burch, 1986).

$$LS = (A \times cellsize/22.13)^{0.4} \times (\sin\theta/0.0896)^{1.3} \quad (1)$$

here A and θ indicate the flow accumulation and slope values in degree, respectively.

The ratio between the land surface and flat surface in a specific area is defined as TRI (Stambaugh and Guyette, 2008). TRI map are derived from the following formula

$$TRI = Abs(max^2 - min^2) \quad (2)$$

where maximum and minimum altitude are written as max and min, respectively.

SPI and TWI are vital hydrological factors for LSM. The erosive power and discharge in a specific area are expressed by SPI (Moore and Grayson, 1991; Chapi et al., 2017). TWI is defined as the ratio between the catchment area and slope (Gallant, 2000; Chapi et al., 2017). Both variables were calculated from the catchment area (A) and slope (β) by the given formulae (Moore et al., 1991).

$$SPI = A \times \tan\beta \quad (3)$$

$$TWI = \ln(A/\tan\beta) \quad (4)$$

The stability of slope is controlled by the saturation level of material, which is indirectly related to the drainage density (Yalchin, 2008). The construction of roads beside high slope area reduces burden on both the heel of slope and topography that causes landslide (Yalchin, 2008). Lineament is the linear or curvilinear feature on the earth surface (Das et al., 2018). It represents weak zones where the slope instability leads to the landslide. Lineament map of the study area was prepared from the shaded relief maps. The density maps of drainage, road, and lineament were prepared in Arc GIS 10 environment using spatial analyst tools.

To prepare the NDVI thematic layer, band 4 (red band) and band 5 (infrared) of Landsat 8 OLI (<https://earthexplorer.usgs.gov/>) image (date: 29.10.2018) were used. The NDVI value was extracted as follows:

$$NDVI = (NIR - R)/(NIR + R) \quad (5)$$

where NIR and R denote the infrared and red bands of the electromagnetic spectra, respectively.

The lithology map obtained from the Geological Society of India (GSI) was reclassified into four broad categories: Karla, Indrayani, Diveghat, and Purandgarh formation on 1:250000 scale. The geomorphology map of the study area was reclassified into five major groups namely, structural origin moderately dissected plateau (SoMDp), structural origin highly dissected plateau (SoHDp), structural origin low dissected plateau (SoLDp), denudational origin moderately dissected lower plateau (DoMDLp), denudational origin-pediment pediplain complex (DoPPc) on 1:50000 based on Landsat 8 image, NRSC base map, topographical maps and field survey scale. The thematic map of soil was modified using the base map of the National Bureau of Soil Survey and Land Use planning (NBSSLUP) on 1: 500,000 scale. The soil map was categorized into three broad classes: Typic Ustorthents, Lithic Ustorthents, and Udic Rhodustalfs . Rainfall has been the most influential factor in landslide occurrences especially in the Indian sub-continent due to arrival of the southwest monsoon. The rainfall data of the last five years (2013-2018) was collected from the rain gauge stations of the department of agriculture, Maharashtra state. Based on this data, rainfall map was classified into four classes using the inverse distance weighted method. Among the classification methods, natural break method was extensively applied for classification of landslide variables (Phong et al., 2019; Chen and Chen, 2021). In the present study, natural break method was used for the classification of the causative factors of landslide in ArcGIS platform.

3 Methodology

The methodology flow-chart of the present study is presented in Figure 4.

3.1 Boruta and multicollinearity feature selection methods

Variable selection is the essential function in natural hazard mapping. The factor determination method improves the model performance by increasing the data quality and reducing noise,

overfitting, dimensionality of feature space and preventing redundancy problems (Bui et al., 2019). There are two popular features choosing methods for modeling including wrapper and filter accessible. In the present study, the boruta algorithm was used for feature selection based on wrapper method. The classification in boruta approach is implemented by voting of multiple unbiased weak classifier decision trees. The scale of importance of the variable is obtained as the loss of classification accuracy due to the random alternate of factor values between aims. The algorithm works as follows: (i) At the first step, it adds randomness to the dataframe by shuffling each feature which are called shadow feature (ii) Then, it trains the RF classifier to measure the Z score values and checks whether the attribute has a higher Z score than the maximum Z score among shadow attributes (MZSA). (iii) Finally, the features having value less than MZSA are rejected for being irrelevant and vice versa (Kursa and Rudnicki, 2010; Amiri et al., 2019; Prasad et al., 2021). Shadow minimum, mean, and maximum values represent the Z score of a shadow attribute. On the other hand, tentative variables lying close to the maximum shadow (maximum Z score), are confirmed or rejected based on the mean shadow (mean Z score).

In addition, multicollinearity analysis was performed to verify the results of boruta method. Multicollinearity is the popular statistical method to find out whether any positive correlation between the conditioning factors exists or not. Researchers like Chen et al. (2018a), Pourghasemi et al. (2020) showed that the influencing variables containing the variance inflation factors (VIF) value more than 5 and tolerance (TOL) value less than 0.2 were subjected to multicollinearity while in the studies of Nguyen et al. (2017), Wang et al. (2019) the VIF value >10 or TOL value <0.1 of the controlling factors were rejected.

3.2 Landslide events and variable inter-relationship by Weight of evidence (WoE) method

The WoE is a quantitative approach used to forecast the occurrence of hazard events (Armas, 2012). Based on Bayes rule, the approach was primarily propounded for mineral potential mapping (Bonham-carter, 1994; Chen et al., 2018a). With the march of time, WoE approach was popularized in other spatial prediction mapping (Rahmati et al., 2015; Aghdam et al., 2017; Chen et al., 2018a; Arabameri et al., 2018). In this study, WoE method was used to determine the correlation between the landslide events and effective parameters. The primary purpose of this model is to measure the positive (W_i^+) and negative weights (W_i^-).

$$W_i^+ = \ln \left[\frac{P(B|A)}{P(B|\bar{A})} \right] \quad (6)$$

$$W_i^- = \ln \left[\frac{P(\bar{B}|A)}{P(\bar{B}|\bar{A})} \right] \quad (7)$$

where P and ln stand for the probability and natural logarithm, respectively. B and $\bar{B}$ indicate the presence and absence of desired sub-class of landslide predisposing factors, respectively whereas A and $\bar{A}$ are the presence and absence of landslide events, respectively.

3.3 Application of machine learning models

Random forest introduced by Breiman (2001), is a non-parametric method for classification and regression analysis. The RF algorithm creates many decision trees to enhance the performance of the model (Breimen, 2001; Rahmati et al., 2016). Each decision tree is originated by bootstrap samples and leaves around one-third of samples for validation purpose using the out of bag (OOB) error. The advantages of the model are: i) large dataset handling, ii) no overfitting, iii) OOB error estimation, iv) low bias and low variance, v) no requirement of prior data transformation and rescaling, vi) higher prediction performance. To tune the model, the algorithm requires two important criteria: number of trees (n_{tree}) and number of features (m_{try}).

Deep boost is a new ensemble learning algorithm formulated by Cortes, Mohri, and Syed (2014). The method is updated and some new solutions that are left about the theory underlying AdaBoost. It consists of deep decision trees or members of complex families. The method is a capacity-conscious criterion to achieve high accuracy without overfitting the data. The advantage of the algorithm is to minimize the corresponding learning bound and the mixture weight assigned to each sub-family (Cortes et al., 2014; Chen et al., 2021). The algorithm can extend by examining non-differentiable convex surrogate losses. As far as we are concerned, the DB model has been used for the first time in natural hazard mapping.

Stochastic gradient boosting is an ensemble machine learning technique that merges both the advantages of bagging and boosting methods and is used for classification and regression purpose (Friedman, 2001, 2002; Zhou et al., 2016). SGB is the modification of adaboost and adabative bagging method. SGB has many advantages such as i) it uses random sub-sample

(selected without replacement) of the dataset at each stage of boosting process instead of the whole dataset, ii) at each step, small trees are generated rather than large entire classification trees, iii) since small trees are developed, SGB method is highly resistant to overfitting iv) the steepest gradient algorithm of the model is resistant to outliers (Friedman, 2002; Lawrence et al., 2004; Moisen et al., 2006).

Rotation forest is an ensemble learning method devised by Rodriguez et al. (2006) to enhance the predictive performance of individual classifiers. In this method, every single classifier is merged with one another and each classifier may be trained using different i) training datasets; ii) subsets of features; iii) parameters of the classifier; or iv) classifier models (Polikar et al., 2010; Nguyen et al., 2017). The model is a combination of the random subspace and bagging method with principal component analysis (PCA) for the ensemble classifier. By applying PCA technique, the training dataset was randomly classified into K subsets (Rodriguez et al., 2006; Nguyen et al., 2017; Naghibi et al., 2019).

Boosted regression tree utilizes both machine learning and statistical techniques for classification and regression problems (Youssef et al., 2016; Naghibi et al., 2016). BRT algorithm uses boosting technique to combine a large number of simple tree for maximizing the predictive capacity (Elith et al., 2008; Naghibi et al., 2019). The advantages of this model are: i) it replaces the missing data using surrogates, ii) it can deal with any type of predictive variable (numeric, categorical, binary), iii) the predictive value get affected by monotone transformation, and different scales of measurement. To run the BRT model, three parameters are required to optimize it, namely interaction depth, shrinkage or learning rate, and number of trees. Interaction depth represents the number of nodes and the shrinkage indicates the contribution of trees.

Logit Boost developed by Friedman and Tibshirani (2000), uses boosting method to reduce the bias and variance (Oh et al., 2019). The LB method is a minor variation of popular boosting technique (Adaboost). In this method, the additive logistic regression function is used to classify the dataset (Tehrany et al., 2018). The main advantages of the LB technique are: i) it can handle noisy data, ii) it has the ability to reduce the training errors, iii) it is used for the multiclass tasks (Oh et al., 2019). The vector values of the model define the input variables and two output groups (landslide and non-landslide events).

3.4 Validation of the models

Validation is an imperative stage in machine learning modelling for the evaluation of the study, without which the models have not any scientific significance (Rahmati et al., 2016; Naghibi et al., 2016; Prasad et al., 2020). In the validation dataset, dependent variables are not considered for verifying the model efficiency to predict the location of landslides. In the present study, the models were validated based on a set of quantitative parameters such as accuracy, kappa index, sensitivity, specificity, and area under the receiver operating characteristics curve (AUROC). These quantitative criteria were assessed using the possible consequences (PC), namely true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The detailed information of these quantitative parameters is available in Chapi et al. (2017).

The ROC curve is a scientific method to examine the overall performance of the models (Rahmati et al., 2016; Chapi et al., 2017; Bui et al., 2019). The AUROC is a quantitative index for model capability to predict correctly the occurrence or non-occurrence of specific events. AUC value lying in between 0 to 1, was classified as poor (0.5–0.6), average (0.6–0.7), good (0.7–0.8), very good (0.8–0.9), and excellent (0.9–1) (Chen et al., 2018b). The higher AUC value, more will be the model perfection (Chapi et al., 2017; Naghibi et al., 2017; Das 2019, 2020). AUROC value is calculated as follows:

$$AUROC = \frac{\sum TP + \sum TN}{P + N} \quad (8)$$

here the P and N refer to the total numbers of landslide occurrences and non-landslide occurrences in the study area, respectively.

4. Results

4.1 Feature selection and important landslide triggering variables

The selection of significant conditioning factors in machine learning techniques is of vital importance to improve model accuracy by eliminating impertinent factors (Khosravi et al., 2018; Bui et al., 2019). In the current research, the boruta technique was used for the relevant variable selection. From the nineteen landslide predisposing factors, boruta evaluation elected fourteen suitable variables (slope, TRI, RD, TWI, elevation, LS, DD, SPI, geomorphology, rainfall, soil,

lithology, LD, and NDVI) whereas five factors including standard curvature, plan curvature, profile curvature, land-use, and aspect were removed (Fig 5a). These rejected variables were also excluded in the study of Chen et al., 2018b. The boruta approach is capable of determining the mean, median, minimum, and maximum importance of the factors displayed in Table 1. In congruence with the result of boruta algorithm, the result of multicollinearity analysis presented in Table 2, showed that the standard, plan and profile curvature were disqualified since their VIF value exceeded 10 and TOL fell below 0.1.

After choosing the significant factors that are crucial for the modelling, the variable importance function of RF method was used to measure the feature importance for LSM in the study area. The measured importance values of the landslide conditioning variables were 100 (slope), 94.49 (TRI), 65.99 (elevation), 56.98 (TWI), 56.34 (LS), 46.26 (DD), 44.56 (RD), 41.99 (rainfall), 34.84 (SPI), 33.47 (NDVI), 27.54 (LD), 22.04 (geomorphology), 8.44 (soil), and 0 (lithology). These results of the variable importance analysis revealed that the slope, TRI, elevation, and TWI had more influence on landslides, followed by LS, DD, RD, rainfall, SPI, NDVI, LD, geomorphology, soil, and lithology (Fig 5b). From the literature, it was noticed that the contribution of the conditioning factors in a particular region varies with the use of different measuring methods (Chen et al., 2018b). In the case of present study, variable importance results of both boruta and RF methods were almost same that confirmed the predictive ability of the algorithms.

4.2 Spatial interaction of landslides and conditioning factors using WoE model

The spatial interconnection between landslide events and affecting variables applying WoE method is presented in Table 3. In this table, C/SC value is the final weight used for the above relationship. The C/SC value of the present study was lying between -9.60 and 9.11. The higher positive value shows strong correlation between landslide events and conditioning factors, whereas negative weight indicates poor relation. Regarding slope, increasing slope angle is directly related to the landslide events in a certain range (Bui et al., 2017; Chen and Chen, 2021). This fact was reflected from the current research where most of the landslide happened in the higher slope angle classes viz. 4°-34° and 35°-73° with 8.87 and 7.69 C/SC values, respectively whereas lower slope angle held negative weights. The elevation is another significant factor of landslide events as shown in the study of Wang et al. (2019) and Arabameri et al., (2019). The analysis of WoE for

landslide events and elevation expressed that the moderate elevation (430-713 m) recorded the maximum C/SC value (9.11). On the other side, lower (4-154 m) and higher (920-1404 m) elevation classes had negative weights. The increasing slope length is favorable for water infiltration into sub-soil that enhances the possibility of slope failure (Bui et al., 2017). The LS value was highest (7.61) in the class of 0.42-2.10, indicating strong relation with landslide occurrences. The weight of the C/SC went upward with increasing topographic roughness. The C/SC values of sub classes 0.13-0.58, 0.58-0.85, and 0.85-1.23 of TRI were -9.60, 7.41, and 7.98, respectively. The zero value of all the classes of SPI made it clear that SPI had no relation with landslide events. The first class (2.23-5.68) of TWI contented the highest C/SC value (6.50) implying the presence of high soil moisture content rendered landslide probability high. The maximum weight (7.81) for DD was found in the second class (0.51-0.75), followed by 0.13-0.51 (-1.08), 0.75-0.96 (-2.73), and 0.96-1.23 (-4.08). The region having positive DD value denoted greater chances of landslide occurrence. With respect to RD, the C/SC value (4.93) was observed as the highest in class 1.01-4.11 where more than forty five percent of landslides occurred. In the case of LD, results showed that the class of 0-0.28 was the greatest weight of 2.50. A high vegetation density area minimizes the chance of landslide by reducing soil erosion and its transmission. As per NDVI results the 0.21-0.35 class contained the highest C/SC value (2.57) suggesting high susceptibility to landslides followed by -0.03-0.21 (1.32), 0.35-0.41 (-0.06), and 0.41-0.58 (-2.12). The longer period of monsoon rainfall in the study area had direct impact on landslide occurrence because of more runoff caused soil loosening and sediment flow. The maximal rainfall class (619-656 mm) in the area under study indicates that around fifty percent of landslides took place in this group. Regarding lithology, the highest weights were obtained from the bdd (3.31) and bdp (2.47) groups. Since, the rock of these groups belongs to 350-450 m elevation (landslide occurrence zone), it can be inferred that they are more weathered. For geomorphology, structural origin highly dissected plateau has strongly correlated with the landslide events of C/SC value 6.97. This sub class also obtained maximum score by AHP method in the research of Bera et al., 2018. In the case of soil, Udic Rhodustalfs soil was only positive weight (6.21) among the soil groups.

4.3. Model construction

In machine learning models, researchers applied 10 cross-validation method in order to enhance the precision of classification and to reduce the noise (Wang et al., 2020). Therefore, the present work adopted this method with five repetitions using 70% of the landslide and non-landslide points. These points along with their extracted fourteen conditioning factor values comprised the training dataset whereas the remaining points were used for validation purpose. The exact pixel values of the derived parameters influenced the validation result due to the higher range of the values. To solve this problem, values were transformed into 0 to 1 range with the help of min-max normalization. After the model establishment with training dataset, testing dataset was utilized to verify the performance and predictive capability of each model. This complete procedure was done using the R and Arc GIS software.

4.4 Performance and comparison of the models

In this study, AUC and selected statistical measurements were used for evaluation of the models. These criteria considered both training and testing results for model comparison and application. The training dataset defines the model fitness whereas the testing dataset represents the predictive capability of the model (Bui et al., 2018; Prasad et al., 2021). In the case of training dataset, among the six models, the RF performed excellently with the highest value of 1 quantified by accuracy, Kappa, sensitivity, specificity, and AUC in both the primary and secondary study area (Table 4). Landis and Koch, (1977) determined that the kappa value > 0.8 means the perfect match of measured value with observed value. The value of 1 for sensitivity and specificity represent that all landslide and non-landslide points were correctly classified. The AUC values of the remaining models for primary and secondary regions were 0.89, 0.99 (DB); 0.87, 1 (SBM); 0.95, 0.97 (RoF); 0.82, 0.89 (BRT); and 0.86, 1 (LB), respectively. The highest AUC value of RF method in both the region manifested the better goodness of fit compared to the rest.

The established models were corroborated with the testing dataset to check the predictive ability of the models. The results of validation dataset revealed that RF method held the highest as measured by accuracy (0.81), kappa (0.62), specificity (0.86), and AUC (0.88) compared to other models (Fig. 6a, Table 5). It is also observed that the RF model appeared to be ideal in the secondary region based on accuracy (0.77), kappa (0.50), specificity (0.86), and AUC (0.89) (Fig. 6b, Table 5). For primary and secondary regions, AUC values of the other models were 0.87; 0.82

(DB), 0.84; 0.85 (SBM), 0.83; 0.82 (RoF), 0.86; 0.81 (BRT), and 0.77; 0.76 (LB) respectively. The comparative results from Tables 4 and 5 exhibited suitability of the models in the following descending manner: RF DB, SBM, RoF, BRT, and LB. The model's AUC values which were more than 0.8 (except LB) confirmed the high precision landslide susceptibility mapping in the areas of investigation. In a nutshell, RF model outperformed other models in both the regions.

4.5 Landslide susceptibility map construction

The preparation of the landslide susceptibility maps is essential to know the possibility of future landslide occurrences in an area. After the training processes, each pixel of the study area was computed for the prediction of landslide. These values ranging from 0 to 1 indicate low to high chances of landslide occurrences. In general, the classification of the derived probability values includes automatic and user-defined methods (Chen et al., 2021a). However, there are several standard techniques in GIS to reclassify the probability maps, namely natural breaks, min-max normalizations, standard deviations, equal intervals, geometrical, and quantiles. A specific method is selected based on the research objective and distribution of data (Chapi et al., 2017). It was learned from the previous studies that many researchers had employed the natural break method for generating the susceptibility maps (Hong et al., 2018; Park and Kim, 2019; Panahi et al., 2020; Chen et al., 2021a). This classification method has advantage of no class bias as well as minimum intra-class and maximum inter-class deviations (Arabameri et al., 2020). In the present work, the distribution of probability values of the models abruptly changed. For this reason, the landslide susceptibility maps were classified using the natural break method in ArcGIS environment into four groups, including low, moderate, high, and very high (Fig. 7). The low and moderate prone zones of the entire study area account for 22.74%, 40.06 (RF); 30.08%, 40.94% (DB); 38.99%, 30.37% (SBM); 54.59%, 28.75% (RoF); 57.13%, 15.93% (BRT); and 9.02%, 28.8 (LB), respectively. The high susceptible classes of RF, DB, SBM, RoF, BRT, and LB models constitute 26.06%, 20.13%, 19.37%, 13.02%, 13.51%, and 44.3% of the study region, respectively. The share of 10.58% (RF), 8.83% (DB), 11.25% (SBM), 3.61% (RoF), 13.42% (BRT), and 17.85% (LB) belong to very high vulnerable zones (Fig. 8). The lowest minimal variations of high and very high susceptibility classes indicate the high model precision (Naghbi et al., 2017; Prasad et al., 2020). In the present study, minuscule difference between RF, DB, and SBM, models with regard to high and very high landslide prone zones ensured the models high accuracy (Fig. 8). The low

susceptibility classes of the RoF and BRT models occupied much larger area than the moderate susceptibility classes in compared to other models (except LB model). This is mainly on account of the fact that the degree of prediction varies model to model for their different algorithm. Regarding LB model, susceptible classes did not correspond with the classes of other models because of less precision in prediction. The study pointed out that the north-eastern, eastern, and south-eastern parts of the research area are more vulnerable to landslides owing to the steepness, roughness, and wetness of these mountainous regions.

5. Discussion

5.1 Variable Contribution Analysis

Although many studies have been accomplished worldwide on LSM, still there is no specific set of variables (Van et al., 2006; Chen and Li, 2020). The influencing factors vary from one study region to another due to heterogeneity of the earth. In the current study, a total of nineteen factors were chosen consulting recent studies carried out in the different regions with more or less same elevation as of the study area under consideration and subsequently fourteen relevant factors were selected based on boruta algorithm for LSM (Hong et al., 2018; Kadavi et al., 2018; Nohani et al., 2019; Arabameri et al., 2019; Wang et al., 2019). It is worth mentioning that the result of boruta approach almost matched with the result of multicollinearity which reflects the robustness of boruta algorithm in feature selection. It can be seen that the topographical factors such as slope, TRI, and elevation exerted more influence on landslide occurrences compared to the categorical factors (lithology, soil, and geomorphology) in the study area. Researcher inferred that topographical variables are the most responsible for landslide incidents (Nakileza and Nedala, 2020; Arabameri et al., 2020). This is because of the topographical factors mainly slope angle which plays a key role in slope stability and controls the shear and normal strength on the plane of weakness. The study area is basically more or less homogeneous bedrock region (Sahyadri group) with a thin veneer of top soil. Therefore, the factors lithology and soil did not have much influence on landslide occurrence. Most of the slides in this area that occurred within accessible locations are rock falls or debris flow. On the other side, WoE method expressed the influence variability of sub classes of each landslide governing factor. Based on the result, the highest C/SC values noted from the slope and elevation sub classes revealed that the landslide occurrences mostly took place

into moderate elevation (430-713 m) with higher slope angle (24-73°) probably due to more weathering at this height of the study area.

5.2 Previous works and future work proposal

Mathematically, landslide susceptibility is the likelihood of the occurrence of slope failures in location while the geo-environmental condition is fixed (Guzzeti et al., 2005). LSM is generally done to predict the potential geographic locations of future landslides (Guzzeti et al. 1999; 2006). In recent years, machine learning approaches have become more popular and outstripped other conventional methods in landslide susceptibility and hazard mapping (Bui et al., 2019; Arabameri et al., 2020). For the mapping of potential landslide locations, a large number of machine learning methods have been developed, artificial neural network (Lee et al., 2001), support vector machine (Dou et al., 2015), random forest (Miner et al., 2010), neuro-fuzzy (Pradhan, 2013), naïve bayes (Bui et al., 2012), quadratic discriminant (Rossi et al., 2010), adaptive neuro-fuzzy inference system-satin bowerbird optimizer (Chen et al., 2021a) just to name a few. However, there is no universal model for spatial mapping of landslide but the applied model can be appropriate for similar geo-environmental settings (Chen et al., 2020; Prasad et al., 2020). Therefore, the present study aims to select suitable method for LSM in the mountainous regions of western India. In this research, six machine learning approaches, including RF, DB, SBM, RoF, BRT, and LB were compared in the study area as well as in the secondary region to choose the significant model. Among these models, RF model showed the highest precision, consistent performance, and minimal difference between training and validation results, followed by DB, SBM, RoF, BRT, and LB for LSM in both the regions. The robustness of the RF method is also manifested in other spatial prediction mappings with very high precision (Catani et al., 2013; Naghibi et al., 2019; Costache et al., 2019; Amiri et al., 2019; Prasad et al., 2020). A study by Chen et al. (2018b) in Chongren County, China, located at about an elevation (1-1230 m) as of the current study area, RF model proved to be the most suitable with respect to the performance of other applied models in landslide prediction. Arabameri et al. (2018) found that the RF model was the best fit in the Gallicash River Watershed, Iran. Bera et al. (2018) worked on LSM in mountainous region of western India. In this study, accuracy of fuzzy gamma operator (81%) and analytical hierarchy process (AHP) (69%) models were lesser than the present research. In the above research works, the researchers employed several methods including novel model only to one respective study

area, which restricted the applicability of the models. To widen applicability of the models, it is important to apply the selected models in alike geographic environment (Prasad et al., 2020). In the present study, the outcomes of the models were almost same for both the primary and secondary regions. Therefore, the comparability of the models remained consistent with the either regions under investigation. From the results, it can be stated that the RF model is appropriate for LSM in the western India and also in regions with similar geo-environmental conditions. Likely, in future, more researches are required to conduct in different geographical areas around the world for the assessment of applicability of machine learning models. Furthermore, it can be suggested that the integration of SAR and optical datasets may improve the precision of the models and help to identify the landslide points in a larger area.

5.3 Advantages and disadvantages of the models

The machine learning method is applied in natural hazard and resource mapping because of its unique advantageous capability of automatically performing the task of searching several datasets aimed at obtaining necessary information (Chen et al., 2020). It is also noted that the predictive accuracy of the model, to some extent, depends on the quality of data and the establishment of the model from training dataset (Nohani et al., 2019). In present study, an attempt was made to find out the probable reasons behind the better performance of RF method and comparatively poor achievement of LB model. RF method suitably fitted resulting in reduction of the error for validation dataset in the primary and secondary regions. The reason for better performance of the RF method may be due to composition of multiple decision trees and interconnection capability of conditioning variables and non-linearity (Catani et al., 2013; Prasad et al., 2020). Besides, RF method escapes the overfitting of dataset which is common problem in artificial intelligence approaches. In the case of LB method, the performance is relatively less in either region. The underperformance of LB model may be because of i) steady property indicated by logistic loss, ii) any change of output value does not affect the loss value, iii) oversimplification in model establishment can reduce the validation accuracy, iv) problem in node split gain and value fitting (Sun et al., 2014).

6. Conclusions

The Landslide susceptibility mapping through high performing artificial intelligence models is a indispensable step for downscaling future loss of properties and human lives. Therefore, in this study, six machine learning approaches (RF, DB, SGB, RoF, BRT, and LB) were employed for effective LSM. To confirm the model applicability, the selected methods were tested in another region as well. The major findings of the study are as follows:

i) Among the nineteen landslide causing factors, fourteen (slope, elevation, LS, SPI, TRI, TWI, RD, LD, DD, NDVI, rainfall, lithology, soil, and geomorphology) were selected and remaining five (aspect, land-use, plan curvature, profile curvature and standard curvature) were discarded with the help of boruta approach for LSM.

ii) From the attribute evaluation of WoE method, it was found that the landslide events mostly occurred in the moderately elevated area within higher slope angle where the topographical roughness is maximum. As per the variable importance of RF method, topographical factors were most important compared to the other factors for the LSM.

iii) Out of six methods, RF model is the most desirable for its superlative performance in model fitness and validation in both the primary and secondary regions. Besides, novel application of the second most suitable model DB in LSM has been successful which widens its scope in natural hazard mapping.

iv) According to landslide susceptibility classes of RF method, about 11% and 26% of the study area come under the very high and high landslide vulnerable zones, respectively.

The outcomes of this study can be useful to the planners for planning and management of future landslides events. To meet the purpose, RF method can be opted for preparing LSM in alike geographic environment.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

539 The authors confirm that the data supporting the findings of this study are available within the
540 article.

541 **References**

542 Abedini, M., Ghasemian, B., Shirzadi, A., Shahabi, H., Chapi, K., Pham, B.T., Bin Ahmad, B.,
543 Tien Bui, D., 2019. A novel hybrid approach of bayesian logistic regression and its ensembles
544 for landslide susceptibility assessment. *Geocarto International* 34, 1427–1457.
545 <https://doi.org/10.1080/10106049.2018.1499820>

546 Aburas, M.M., Ahamad, M.S.S. and Omar, N.Q., 2019. Spatio-temporal simulation and prediction
547 of land-use change using conventional and machine learning models: a review.
548 *Environmental Monitoring and Assessment*, 191(4), 205. [https://doi.org/10.1007/s10661-](https://doi.org/10.1007/s10661-019-7330-6)
549 [019-7330-6](https://doi.org/10.1007/s10661-019-7330-6)

550 Aghdam, I.N., Pradhan, B. and Panahi, M., 2017. Landslide susceptibility assessment using a novel
551 hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy
552 inference system (ANFIS) at southern Zagros Mountains in Iran. *Environmental Earth*
553 *Sciences*, 76(6), 237. DOI: [10.1007/s12665-017-6558-0](https://doi.org/10.1007/s12665-017-6558-0)

554 Aleotti, P. and Chowdhury, R., 1999. Landslide hazard assessment: summary review and new
555 perspectives. *Bulletin of Engineering Geology and the Environment*, 58(1), 21-44.
556 <https://doi.org/10.1007/s100640050066>

557 Amiri, M., Pourghasemi, H.R., Ghanbarian, G.A. and Afzali, S.F., 2019. Assessment of the
558 importance of gully erosion effective factors using Boruta algorithm and its spatial modeling
559 and mapping using three machine learning algorithms. *Geoderma*, 340, 55-69.
560 <https://doi.org/10.1016/j.geoderma.2018.12.042>

561 Arabameri, A., Rezaei, K., Pourghasemi, H.R., Lee, S. and Yamani, M., 2018. GIS-based gully
562 erosion susceptibility mapping: a comparison among three data-driven models and AHP
563 knowledge-based technique. *Environmental Earth Sciences*, 77(17), 628.
564 <https://doi.org/10.1007/s12665-018-7808-5>

565 Arabameri, A., Pradhan, B., Rezaei, K. and Lee, C.W., 2019. Assessment of landslide
566 susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and
567 multiresolution DEMs. *Remote Sensing*, 11(9), 999. <https://doi.org/10.3390/rs11090999>

568 Arabameri, A., Saha, S., Roy, J., Chen, W., Blaschke, T. and Tien Bui, D., 2020. Landslide
569 susceptibility evaluation and management using different machine learning methods in the
570 Gallicash River Watershed, Iran. *Remote Sensing*, 12(3), 475. [doi:10.3390/rs12030475](https://doi.org/10.3390/rs12030475)

571 Armaş, I., 2012. Weights of evidence method for landslide susceptibility mapping. Prahova
572 Subcarpathians, Romania. *Natural Hazards*, 60(3), 937-950. [https://doi.org/10.1007/s11069-](https://doi.org/10.1007/s11069-011-9879-4)
573 [011-9879-4](https://doi.org/10.1007/s11069-011-9879-4)

574 Bera, S., Guru, B. and Ramesh, V., 2019. Evaluation of landslide susceptibility models: A
575 comparative study on the part of Western Ghat Region, India. *Remote Sensing Applications:*
576 *Society and Environment*, 13, 39-52. <https://doi.org/10.1016/j.rsase.2018.10.010>

577 Bonham-Carter, G.F., 1994. *Geographic Information Systems for Geoscientists: Modelling with*
578 *GIS*, Elsevier, UK.

579 Brabb, E.E., Pampeyan, E.H., and Bonilla, M.G., 1978. Landslide susceptibility in San Mateo
580 County, California. US Geological Survey Miscellaneous Field Studies Map, MF-360, Map
581 at 1: 62,500 scale.

582 Breiman, L. 2001. "Random Forests." *Machine Learning*. 45 (1), 5–32.

583 Budimir, M.E.A., Atkinson, P.M., Lewis, and H.G., 2015. A systematic review of landslide
584 probability mapping using logistic regression. *Landslides*, 12, 419–436
585 <https://doi.org/10.1007/s10346-014-0550-5>

586 Bui, D.T., Pradhan, B., Lofman, O. and Revhaug, I., 2012. Landslide susceptibility assessment in
587 vietnam using support vector machines, decision tree, and Naive Bayes Models.
588 *Mathematical Problems in Engineering*, <https://doi.org/10.1155/2012/974638>

589 Bui, D.T., Nguyen, Q.P., Hoang, N.D. and Klempe, H., 2017. A novel fuzzy K-nearest neighbor
590 inference model with differential evolution for spatial prediction of rainfall-induced shallow
591 landslides in a tropical hilly area using GIS. *Landslides*, 14(1), 1-17. [DOI 10.1007/s10346-](https://doi.org/10.1007/s10346-016-0708-4)
592 [016-0708-4](https://doi.org/10.1007/s10346-016-0708-4)

593 Bui, D.T., Shahabi, H., Shirzadi, A., Chapi, K., Pradhan, B., Chen, W., Khosravi, K., Panahi, M.,
594 Bin Ahmad, B. and Saro, L., 2018. Land subsidence susceptibility mapping in south korea
595 using machine learning algorithms. *Sensors*, 18(8), 2464. <https://doi.org/10.3390/s18082464>

596 Bui, D.T., Shirzadi, A., Shahabi, H., Geertsema, M., Omidvar, E., Clague, J.J., Thai Pham, B.,
597 Dou, J., Talebpour Asl, D., Bin Ahmad, B. and Lee, S., 2019. New ensemble models for
598 shallow landslide susceptibility modeling in a semi-arid watershed. *Forests*, 10(9), 743.
599 <https://doi.org/10.3390/f10090743>

600 Carrara, A., Carratelli, E.P., and Merenda, L., 1977. Computer-based data bank and statistical
601 analysis of slope instability phenomena. *Zeitschrift für Geomorphologie*, 21, 187–222.

602 Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V., 2013. Landslide susceptibility estimation
603 by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System*
604 *Sciences*, 13(11), 2815-2831. <https://doi.org/10.5194/nhess-13-2815-2013>, 2013

605 Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T. and Khosravi, K., 2017.
606 A novel hybrid artificial intelligence approach for flood susceptibility assessment.
607 Environmental Modelling & Software, 95, 229-245. <https://doi.org/10.1016/j.envsoft.2017.0>
608 [6.012](https://doi.org/10.1016/j.envsoft.2017.0)

609 Chen, W., Pourghasemi, H.R., Kornejady, A. and Zhang, N., 2017. Landslide spatial modeling:
610 Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques.
611 Geoderma, 305, 314-327. <https://doi.org/10.1016/j.geoderma.2017.06.020>

612 Chen, W., Chen, W., Li, H., Hou, E., Wang, S., Wang, G., and Peng, T., 2018a. GIS-based
613 groundwater potential analysis using novel ensemble weights-of-evidence with logistic
614 regression and functional tree models. Science of the Total Environment, 634, 853–867.
615 <https://doi.org/10.1016/j.scitotenv.2018.04.055>

616 Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A.X., Pei, X. and Duan, Z.,
617 2018b. Landslide susceptibility modelling using GIS-based machine learning techniques for
618 Chongren County, Jiangxi Province, China. Science of the total environment, 626, 1121-
619 1135.

620 Chen, W., Chen, Y., Tsangaratos, P., Ilia, I. and Wang, X., 2020. Combining evolutionary
621 algorithms and machine learning models in landslide susceptibility assessments. Remote
622 Sensing, 12(23), 3854. <https://doi.org/10.3390/rs12233854>

623 Chen, W. and Li, Y., 2020. GIS-based evaluation of landslide susceptibility using hybrid
624 computational intelligence models. Catena, 195, 104777.
625 <https://doi.org/10.1016/j.catena.2020.104777>

626 Chen, X. and Chen, W., 2021. GIS-based landslide susceptibility assessment using optimized
627 hybrid machine learning methods. Catena, 196, 104833.
628 <https://doi.org/10.1016/j.catena.2020.104833>

629 Chen, Y., Chen, W., Janizadeh, S., Bhunia, G.S., Bera, A., Pham, Q.B., Linh, N.T.T., Balogun,
630 A.L. and Wang, X., 2021. Deep learning and boosting framework for piping erosion
631 susceptibility modeling: spatial evaluation of agricultural areas in the semi-arid region.
632 Geocarto International, 1-27. <https://doi.org/10.1080/10106049.2021.1892212>

633 Chen, W., Chen, X., Peng, J., Panahi, M. and Lee, S., 2021a. Landslide susceptibility modeling
634 based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer.
635 Geoscience Frontiers, 12(1), 93-107. <https://doi.org/10.1016/j.gsf.2020.07.012>

636 Chen, W., Lei, X., Chakraborty, R., Pal, S.C., Sahana, M. and Janizadeh, S., 2021b. Evaluation
637 of different boosting ensemble machine learning models and novel deep learning and
638 boosting framework for head-cut gully erosion susceptibility. Journal of Environmental
639 Management, 284, 112015. <https://doi.org/10.1016/j.jenvman.2021.112015>

640 Cortes, C., Mohri, M. and Syed, U., 2014. Deep boosting. Proceedings of the 31 st International
641 Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

642 Costache, R. and Bui, D.T., 2019. Spatial prediction of flood potential using new ensembles of
643 bivariate statistics and artificial intelligence: A case study at the Putna river catchment of
644 Romania. Science of the Total Environment, 691, 1098-1118.
645 <https://doi.org/10.1016/j.scitotenv.2019.07.197>

646 Das, S., Pardeshi, S.D., Kulkarni, P.P., and Doke, A., 2018. Extraction of lineaments from different
647 azimuth angles using geospatial techniques: a case study of Pravara basin, Maharashtra, India.
648 Arabian Journal of Geosciences, 11 (8), 160.

649 Das, S., 2019. Geospatial mapping of flood susceptibility and hydro-geomorphic response to the
650 floods in Ulhas basin, India. Remote Sensing Applications Society and Environment, 14, 60-
651 74.

652 Das, S., 2020. Flood susceptibility mapping of the Western Ghat coastal belt using multi-source
653 geospatial data and analytical hierarchy process (AHP). Remote Sensing Applications
654 Society and Environment 20, 100379.

655 Dou, J., Paudel, U., Oguchi, T., Uchiyama, S., and Hayakawa, Y.S., 2015. Shallow and
656 DeepSeated Landslide Differentiation using support Vector Machines: a Case Study of the
657 Chuetsu Area, Japan. Terrestrial Atmospheric and Oceanic Sciences, 26, 227.
658 [https://doi.org/10.3319/TAO.2014.12.02.07\(EOSI\)](https://doi.org/10.3319/TAO.2014.12.02.07(EOSI))

659 Elith, J., Leathwick, J.R. and Hastie, T., 2008. A working guide to boosted regression trees. Journal
660 of Animal Ecology, 77(4), 802-813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

661 Friedman, J.H., and Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting
662 (with discussion and a rejoinder by the authors). The Annals of Statistics, 28(2), 337-407.

663 Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of
664 Statistics, 1189-1232.

665 Friedman, J.H., 2002. Stochastic gradient boosting. Computational statistics & data analysis,
666 38(4), 367-378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)

667 Gallant, J.C., 2000. Terrain analysis: principles and applications. John Wiley & Sons.

668 Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P., 1999. Landslide hazard evaluation: a
669 review of current techniques and their application in a multi-scale study, Central Italy.
670 Geomorphology, 31, 181-216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)

671 Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F., 2005. Probabilistic
672 landslide hazard assessment at the basin scale. Geomorphology, 72, 272-299.
673 <http://dx.doi.org/10.1016/j.geomorph.2005.06.002>

674 Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M., 2006. Estimating the
675 quality of landslide susceptibility models. *Geomorphology*, 81, 166-184.

676 Hansen, A., 1984. Landslide hazard analysis. In: Brunsden, D., Prior, D.B. (Eds.), *Slope*
677 *Instability*. Wiley & Sons, New York, 523–602

678 Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W. and
679 Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with
680 AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*,
681 163, 399-413. <https://doi.org/10.1016/j.catena.2018.01.005>

682 Hosseinalizadeh, M., Kariminejad, N., Rahmati, O., Keesstra, S., Alinejad, M. and Behbahani,
683 A.M., 2019. How can statistical and artificial intelligence approaches predict piping erosion
684 susceptibility?. *Science of the Total Environment*, 646, 1554-1566.
685 <https://doi.org/10.1016/j.scitotenv.2018.07.396>

686 Jamali, A., 2020. Land use land cover modeling using optimized machine learning classifiers: a
687 case study of Shiraz, Iran. *Modeling Earth Systems and Environment*, 1-12.
688 <https://doi.org/10.1007/s40808-020-00859-x>

689 Kadavi, P.R., Lee, C.W. and Lee, S., 2018. Application of ensemble-based machine learning
690 models to landslide susceptibility mapping. *Remote Sensing*, 10(8),
691 1252. <https://doi.org/10.3390/rs10081252>

692 Khosravi, K., Pham, B.T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I. and Bui,
693 D.T., 2018. A comparative assessment of decision trees algorithms for flash flood
694 susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*,
695 627, 744-755. <https://doi.org/10.1016/j.scitotenv.2018.01.266>

696 Kim, J.C., Lee, S., Jung, H.S., and Lee, S., 2018. Landslide susceptibility mapping using random
697 forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto International*, 33(9),
698 1000-1015. <https://doi.org/10.1080/10106049.2017.1323964>

699 Kursu, M.B. and Rudnicki, W.R., 2010. Feature selection with the Boruta package. *Journal of*
700 *Statistical Software*, 36(11), 1-13.

701 Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data.
702 *biometrics* 159e174.

703 Lawrence, R., Bunn, A., Powell, S. and Zambon, M., 2004. Classification of remotely sensed
704 imagery using stochastic gradient boosting as a refinement of classification tree analysis.
705 *Remote Sensing of Environment*, 90(3), 331-336. <https://doi.org/10.1016/j.rse.2004.01.007>

706 Lee, S., Ryu, J., Min, K., and Won, J., 2001. Development of two artificial neural network methods
707 for landslide susceptibility analysis. In: *IGARSS 2001. Scanning the Present and Resolving*
708 *the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing*

709 Symposium,(Cat.No.01CH37217).5.2364–2366.
710 <https://doi.org/10.1109/IGARSS.2001.978003>

711 Lee, S. and Choi, J., 2004a. Landslide susceptibility mapping using GIS and the weight-of-
712 evidence model. *International Journal of Geographical Information Science*, 18(8), 789-814.
713 <https://doi.org/10.1080/13658810410001702003>

714 Lee, S., Ryu, J.H., Won, J.S. and Park, H.J., 2004b. Determination and application of the weights
715 for landslide susceptibility mapping using an artificial neural network. *Engineering Geology*,
716 71, 289-302. [https://doi.org/10.1016/S0013-7952\(03\)00142-X](https://doi.org/10.1016/S0013-7952(03)00142-X)

717 Miner, A., Vamplew, P., Windle, D.J., Flentje, P., and Warner, P., 2010. A Comparative Study of
718 Various Data Mining Techniques as Applied to the Modeling of Landslide Susceptibility on
719 the Bellarine Peninsula. *Geologically Active, Proceedings of the 11th IAEG Congress of the*
720 *International Association of Engineering Geology and the Environment*, Auckland, New
721 Zealand.

722 Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E. and Edwards Jr,
723 T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic
724 gradient boosting, generalized additive models, and tree-based methods. *Ecological*
725 *Modelling*, 199(2), 176-187. <https://doi.org/10.1016/j.ecolmodel.2006.05.021>

726 Montgomery, D.R., and Dietrich, W.E., 1994. A physically based model for the topographic
727 control of shallow landsliding. *Water Resources Research*, 30 (4), 1153–1171.
728 <http://dx.doi.org/10.1029/93wr02979>

729 Moore, I.D., and Burch, G.J., 1986. Physical basis of the length-slope factor in the universal soil
730 loss equation 1. *Soil Science Society of America Journal*, 50(5), 1294-1298.
731 [doi:10.2136/sssaj1986.036159950050000500042x](https://doi.org/10.2136/sssaj1986.036159950050000500042x)

732 Moore, I.D. and Grayson, R.B., 1991. Terrain-based catchment partitioning and runoff prediction
733 using vector elevation data. *Water Resources Research*, 27(6), 1177-1191.
734 <https://doi.org/10.1029/91WR00090>

735 Moore, I.D., Grayson, R.B., and Ladson, A.R., 1991. Digital terrain modelling: a review of
736 hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1),
737 3-30. <https://doi.org/10.1002/hyp.3360050103>

738 Naghibi, S.A., Pourghasemi, H.R., and Dixon, B., 2016. GIS-based groundwater potential
739 mapping using boosted regression tree, classification and regression tree, and random forest
740 machine learning models in Iran. *Environmental Monitoring and Assessment*, 188, 44.
741 <https://doi.org/10.1007/s10661-015-5049-6>

742 Naghibi, S.A., Ahmadi, K., and Daneshi, A., 2017. Application of support vector machine, random
743 forest, and genetic algorithm optimized random forest models in groundwater potential

744 mapping. *Water Resources Management*, 31(9), 2761-2775. [https://doi.org/10.1007/s11269-](https://doi.org/10.1007/s11269-017-1660-3)
745 [017-1660-3](https://doi.org/10.1007/s11269-017-1660-3)

746 Naghibi, S.A., Dolatkordestani, M., Rezaei, A., Amouzegari, P., Heravi, M.T., Kalantar, B. and
747 Pradhan, B., 2019. Application of rotation forest with decision trees as base classifier and a
748 novel ensemble model in spatial modeling of groundwater potential. *Environmental*
749 *Monitoring and Assessment*, 191(4), 248. <https://doi.org/10.1007/s10661-019-7362-y>

750 Nakileza, B.R. and Nedala, S., 2020. Topographic influence on landslides characteristics and
751 implication for risk management in upper Manafwa catchment, Mt Elgon Uganda.
752 *Geoenvironmental Disasters*, 7(1), 1-13. <https://doi.org/10.1186/s40677-020-00160-0>

753 Nguyen, Q.K., Tien Bui, D., Hoang, N.D., Trinh, P.T., Nguyen, V.H. and Yilmaz, I., 2017. A
754 novel hybrid approach based on instance based learning classifier and rotation forest
755 ensemble for spatial prediction of rainfall-induced shallow landslides using GIS.
756 *Sustainability*, 9(5), 813. <https://doi.org/10.3390/su9050813>

757 Nohani, E., Moharrami, M., Sharafi, S., Khosravi, K., Pradhan, B., Pham, B.T., Lee, S. and M
758 Melesse, A., 2019. Landslide susceptibility mapping using different GIS-based bivariate
759 models. *Water*, 11(7), 1402. <https://doi.org/10.3390/w11071402fgull>

760 Oh, H.J., Syifa, M., Lee, C.W. and Lee, S., 2019. Land subsidence susceptibility mapping using
761 bayesian, functional, and meta-ensemble machine learning models. *Applied Sciences*, 9(6),
762 1248. <https://doi.org/10.3390/app9061248>

763 Panahi, M., Gayen, A., Pourghasemi, H.R., Rezaie, F. and Lee, S., 2020. Spatial prediction of
764 landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-
765 fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Science of the Total*
766 *Environment*, 139937. <https://doi.org/10.1016/j.scitotenv.2020.139937>

767 Park, S. and Kim, J., 2019. Landslide susceptibility mapping based on random forest and boosted
768 regression tree models, and a comparison of their performance. *Applied Sciences*, 9(5), 942.
769 <https://doi.org/10.3390/app9050942>

770 Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., Le, T.M., Van Phong, T.,
771 Khoi, D.K., Shirzadi, A. and Bui, D.T., 2020a. A novel hybrid approach of landslide
772 susceptibility modelling using rotation forest ensemble and different base classifiers.
773 *Geocarto International*, 35(12), 1267-1292. <https://doi.org/10.1080/10106049.2018.1559885>

774 Pham, B.T., Avand, M., Janizadeh, S., Phong, T.V., Al-Ansari, N., Ho, L.S., Das, S., Le, H.V.,
775 Amini, A., Bozchaloei, S.K., Jafari, F., Prakash, I., 2020b. GIS based hybrid computational
776 approaches for flash flood susceptibility assessment. *Water*, 12 (3), 686.
777 <https://doi.org/10.3390/w12030683>

778 Phong, T.V., Phan, T.T., Prakash, I., Singh, S.K., Shirzadi, A., Chapi, K., Ly, H.B., Ho, L.S.,
779 Quoc, N.K. and Pham, B.T., 2019. Landslide susceptibility modeling using different artificial

780 intelligence methods: A case study at Muong Lay district, Vietnam. Geocarto International,
781 1-24. <https://doi.org/10.1080/10106049.2019.1665715>

782 Polikar, R., DePasquale, J., Mohammed, H.S., Brown, G. and Kuncheva, L.I., 2010. Learn++. MF:
783 A random subspace approach for the missing feature problem. Pattern Recognition, 43(11),
784 3817-3832. <https://doi.org/10.1016/j.patcog.2010.05.028>

785 Pourghasemi, H.R., Mohammady, M. and Pradhan, B., 2012. Landslide susceptibility mapping
786 using index of entropy and conditional probability models in GIS: Safarood Basin, Iran.
787 Catena, 97, 71-84. <https://doi.org/10.1016/j.catena.2012.05.005>

788 Pourghasemi, H.R., Moradi, H.R., and Aghda, S.F., 2013. Landslide susceptibility mapping by
789 binary logistic regression, analytical hierarchy process, and statistical index models and
790 assessment of their performances. Natural Hazards, 69(1), 749-779. <https://doi.org/10.1007/s11069-013-0728-5>

792 Pourghasemi, H.R., Sadhasivam, N., Kariminejad, N. and Collins, A.L., 2020. Gully erosion
793 spatial modelling: Role of machine learning algorithms in selection of the best controlling
794 factors and modelling process. Geoscience Frontiers,
795 <https://doi.org/10.1016/j.gsf.2020.03.005>

796 Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector
797 machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computer &
798 Geoscience, 51, 350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>

799 Prasad, P., Loveson, V.J., Kotha, M., and Yadav, R., 2020. Application of machine learning
800 techniques in groundwater potential mapping along the west coast of India. GIScience &
801 Remote Sensing, 57(6), 735-752. <https://doi.org/10.1080/15481603.2020.1794104>

802 Prasad, P., Loveson, V.J., Das, B. and Kotha, M., 2021. Novel ensemble machine learning models
803 in flood susceptibility mapping. Geocarto International, 1-23.
804 <https://doi.org/10.1080/10106049.2021.1892209>

805 Rahmati, O., Pourghasemi, H.R. and Zeinivand, H., 2015. Flood susceptibility mapping using
806 frequency ratio and weights-of-evidence models in the Golestan Province, Iran. Geocarto
807 International, 31(1), 42-70. <https://doi.org/10.1080/10106049.2015.1041559>

808 Rahmati, O., Pourghasemi, H.R., and Melesse, A.M., 2016. Application of GIS-based data driven
809 random forest and maximum entropy models for groundwater potential mapping: a case study
810 at Mehran Region, Iran. Catena, 137, 360-372. <https://doi.org/10.1016/j.catena.2015.10.010>

811 Regmi, N.R., Giardino, J.R. and Vitek, J.D., 2010. Modeling susceptibility to landslides using the
812 weight of evidence approach: Western Colorado, USA. Geomorphology, 115(1-2), 172-187.
813 <https://doi.org/10.1016/j.geomorph.2009.10.002>

814 Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of
815 statistically-based landslide susceptibility models. *Earth Science Reviews*, 180, 60-91.
816 <https://doi.org/10.1016/j.earscirev.2018.03.001>

817 Rodriguez, J.J., Kuncheva, L.I. and Alonso, C.J., 2006. Rotation forest: A new classifier ensemble
818 method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10), 1619-1630.
819 [doi: 10.1109/TPAMI.2006.211](https://doi.org/10.1109/TPAMI.2006.211)

820 Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., and Peruccacci, S., 2010. Optimal
821 landslide susceptibility zonation based on multiple forecasts. *Geomorphology*, 114, 129–142.
822 <https://doi.org/10.1016/j.geomorph.2009.06.020>

823 Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J.J., Geertsema, M.,
824 Khosravi, K., Amini, A., Bahrami, S. and Rahmati, O., 2020. Flood Detection and
825 Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning
826 Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor
827 Classifier. *Remote Sensing*, 12(2), 266. <https://doi.org/10.3390/rs12020266>

828 Stambaugh, M.C. and Guyette, R.P., 2008. Predicting spatio-temporal variability in fire return
829 intervals using a topographic roughness index. *Forest Ecology and Management*, 254(3), 463-
830 473. <https://doi.org/10.1016/j.foreco.2007.08.029>

831 Sun, P., Reid, M.D. and Zhou, J., 2014. An improved multiclass LogitBoost using adaptive-one-
832 vs-one. *Machine learning*, 97(3), 295-326. <https://doi.org/10.1007/s10994-014-5434-3>

833 Suzen, M.L., and Doyuran, V., 2003. A comparison of the GIS based landslide susceptibility
834 assessment methods: multivariate versus bivariate. *Environmental Geology*, 45, 665-679.

835 Tehrany, M.S., Pradhan, B. and Jebur, M.N., 2013. Spatial prediction of flood susceptible areas
836 using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical
837 models in GIS. *Journal of Hydrology*, 504, 69-79. [https://doi.org/10.1016/j.jhydrol.2013.](https://doi.org/10.1016/j.jhydrol.2013.09.034)
838 [09.034](https://doi.org/10.1016/j.jhydrol.2013.09.034)

839 Tehrany, M.S., Jones, S., Shabani, F., Martínez-Álvarez, F. and Bui, D.T., 2018. A novel ensemble
840 modeling approach for the spatial prediction of tropical forest fire susceptibility using
841 logitboost machine learning classifier and multi-source geospatial data. *Theoretical and*
842 *Applied Climatology*, 137, 637-653. <https://doi.org/10.1007/s10661-019-7362-y>

843 UNISDR, C., 2015. The human cost of natural disasters: A global perspective.

844 Van W.C.J., Rengers, N., and Soeters, R., 2003. Use of Geomorphological information in indirect
845 landslide assessment. *Natural Hazards*, 30, 399-419.

846 Van W.C.J., Van Asch, T.W., and Soeters, R., 2006. Landslide hazard and risk zonation-Why is it
847 still so difficult? *Bull. Eng. Geol. Environ.* 65, 167–184.

848 Van W.C.J., Castellanos, E., and Kuriakose, S.L., 2008. Spatial data for landslide susceptibility,
849 hazard, and vulnerability assessment: an overview. *Engineering Geology*, 102, 112–131.
850 <http://dx.doi.org/10.1016/j.enggeo.2008.03.010>

851 Wang, Y., Fang, Z. and Hong, H., 2019. Comparison of convolutional neural networks for
852 landslide susceptibility mapping in Yanshan County, China. *Science of the Total*
853 *Environment*, 666, 975-993. <https://doi.org/10.1016/j.scitotenv.2019.02.263>

854 Wang, Y., Sun, D., Wen, H., Zhang, H. and Zhang, F., 2020. Comparison of random forest model
855 and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County
856 (Chongqing, China). *International journal of environmental research and public health*,
857 17(12), 4206. <https://doi.org/10.3390/ijerph17124206>

858 Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H. and Hong, H., 2020. Application of alternating
859 decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping.
860 *Catena*, 187, 104396. <https://doi.org/10.1016/j.catena.2019.104396>

861 Yalcin, A., 2008. GIS-based landslide susceptibility mapping using analytical hierarchy process
862 and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations.
863 *Catena*, 72(1), 1-12.

864 Yesilnacar, E. and Topal, T., 2005. Landslide susceptibility mapping: a comparison of logistic
865 regression and neural networks methods in a medium scale study, Hendek region (Turkey).
866 *Engineering Geology*, 79, 251-266. <https://doi.org/10.1016/j.enggeo.2005.02.002>

867 Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression,
868 artificial neural networks and their comparison: a case study from Kat landslides (Tokat—
869 Turkey). *Computers & Geosciences*, 35(6), 1125-1138. [https://doi.org/10.1016/j.cageo.2008.](https://doi.org/10.1016/j.cageo.2008.08.007)
870 [08.007](https://doi.org/10.1016/j.cageo.2008.08.007)

871 Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., and Al-Katheeri, M.M., 2016. Landslide
872 susceptibility mapping using random forest, boosted regression tree, classification and
873 regression tree, and general linear models and comparison of their performance at Wadi
874 Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, 13(5), 839-856. [https://doi.org/](https://doi.org/10.1007/s10346-015-0614)
875 [10.1007/s10346-015-0614](https://doi.org/10.1007/s10346-015-0614)

876 Zhao, X. and Chen, W., 2020. Optimization of computational intelligence models for landslide
877 susceptibility evaluation. *Remote Sensing*, 12(14), 2180. <https://doi.org/10.3390/rs12142180>

878 Zhou, J., Shi, X.Z., Huang, R.D., Qiu, X.Y. and Chen, C., 2016. Feasibility of stochastic gradient
879 boosting approach for predicting rockburst damage in burst-prone mines. *Transactions of*
880 *Nonferrous Metals Society of China*, 26(7), 1938-1

List of caption for figures

- Figure 1** Location maps of the study area (a), and secondary region (b)
- Figure 2** Recent field photos of occurred landslides in the study area
- Figure 3** Thematic layers of the Landslide triggering factors (a) slope, (b) elevation, (c) LS, (d) TRI, (e) SPI, (f) TWI, (g) drainage density, (h) road density, (i) lineament density, (j) NDVI, (k) rainfall, (l) lithology, (m) geomorphology, and (n) soil. The current methodology of the research
- Figure 4** The current methodology of the research
- Figure 5** Feature selection by boruta approach (a), variables importance in landslide susceptibility mapping (b)
- Figure 6** ROC curve of the models in study area (a), secondary area (b)
- Figure 7** Landslide susceptibility mapping derived from (a) RF, (b) DB, (c) SGB, (d) RoF, (e) BRT, and (f) LB models
- Figure 8** Area of different landslide susceptibility classes of the models

Figures

Figure 1

Location maps of the study area (a), and secondary region (b)

Figure 2

Recent field photos of occurred landslides in the study area

Figure 3

Thematic layers of the Landslide triggering factors (a) slope, (b) elevation, (c) LS, (d) TRI, (e) SPI, (f) TWI, (g) drainage density, (h) road density, (i) lineament density, (j) NDVI, (k) rainfall, (l) lithology, (m) geomorphology, and (n) soil. The current methodology of the research

Figure 4

The current methodology of the research

Figure 5

Feature selection by boruta approach (a), variables importance in landslide susceptibility mapping (b)

Figure 6

ROC curve of the models in study area (a), secondary area (b)

Figure 7

Landslide susceptibility mapping derived from (a) RF, (b) DB, (c) SGB, (d) RoF, (e) BRT, and (f) LB models

Figure 8

Area of different landslide susceptibility classes of the models

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Database.docx](#)
- [InventoryofLandslide.pdf](#)