



Landslide Susceptibility Mapping using Machine Learning Algorithm

Muhammad Afaq Hussain ¹, Zhanlong Chen ^{1*} , Run Wang ¹, Safeer Ullah Shah ²,
Muhammad Shoaib ³, Nafees Ali ⁴, Daozhu Xu ⁵, Chao Ma ⁵

¹ School of Geography Information Engineering, China University of Geosciences, Wuhan, China.

² Ministry of Climate Change, Islamabad, Pakistan.

³ State Key Laboratory of Hydraulic Engineering Simulation and Safety, School of Civil Engineering, Tianjin University, Tianjin, China.

⁴ Chinese Academy of Sciences, China.

⁵ State Key Laboratory of Geo-information Engineering, Xi'an Station of Surveying and Mapping, Xi'an, 710054, China.

Received 28 August 2021; Revised 01 January 2022; Accepted 11 January 2022; Published 01 February 2022

Abstract

Landslides are natural disasters that have resulted in the loss of economies and lives over the years. The landslides caused by the 2005 Muzaffarabad earthquake heavily impacted the area, and slopes in the region have become unstable. This research was carried out to find out which areas, as in Muzaffarabad district, are sensitive to landslides and to define the relationship between landslides and geo-environmental factors using three tree-based classifiers, namely, Extreme Gradient Boosting (XGBoost), Random Forest (RF), and k-Nearest Neighbors (KNN). These machine learning models are innovative and can assess environmental problems and hazards for any given area on a regional scale. The research consists of three steps: Firstly, for training and validation, 94 historical landslides were randomly split into a proportion of 7/3. Secondly, topographical and geological data as well as satellite imagery were gathered, analyzed, and built into a spatial database using GIS Environment. Nine layers of landslide-conditioning factors were developed, including Aspect, Elevation, Slope, NDVI, Curvature, SPI, TWI, Lithology, and Landcover. Finally, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) value were used to estimate the model's efficiency. The area under the curve values for the RF, XGBoost, and KNN models are 0.895 (89.5%), 0.893 (89.3%), and 0.790 (79.0%), respectively. Based on the three machine learning techniques, the innovative outputs show that the performance of the Random Forest model has a maximum AUC value of 0.895, and it is more efficient than the other tree-based classifiers. Elevation and Slope were determined as the most important factors affecting landslides in this research area. The landslide susceptibility maps were classified into four classes: low, moderate, high, and very high susceptibility. The result maps are useful for future generalized construction operations, such as selecting and conserving new urban and infrastructural areas.

Keywords: Landslide Susceptibility Modelling; Proportion; Efficient; Random Forest.

1. Introduction

Landslides are caused by many natural and human action-induced disruptions to the soil [1-3]. Landslides inflict significant economic loss every year, whether caused by natural or human activity [4-6]. The Himalayan Mountain Range is vulnerable to frequent and devastating landslides due to robust topography, poor lithology, climate change, soil, fragile sloping infrastructure, and active tectonic action. The probability of landslides happening in an area is

* Corresponding author: chenzl@cug.edu.cn



<http://dx.doi.org/10.28991/CEJ-2022-08-02-02>



© 2022 by the authors. Licensee C.E.J., Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

dependent on ground situations, i.e., the assessment of where landslides may happen [7]. Muzaffarabad is the capital of Azad Jammu and Kashmir, Pakistan, and lies in the Himalayan Mountains. In Muzaffarabad, the earthquake also brought about significant land losses, namely the landslide of Hattian Bala rock, causing the deaths of 595 people [8, 9]. Since the 2005 earthquake in the Kashmir region, various landslide identification, spatial distribution, duration, structural landslide tracking, and hazard analysis have been conducted [10, 11]. In North Pakistan, the Kashmir earthquake in 2005 led to thousands of landslides, killing over 26,000 individuals, with disastrous consequences for infrastructure, directly or indirectly [12]. Therefore, landslide susceptibility starts with the inventory of past landslides in the area and the collection of optimal causative factors. Then, the landslide and a practical approach for analyzing the association between landslide occurrences and influencing factors are needed, leading to the landslide susceptibility map [13].

Significant advances in computation power, remote sensing, and GIS have aided in creating Landslide Susceptibility Mapping (LSM) over the past two decades. For LSM, a diversity of methods have been proposed, including knowledge-based approaches such as:

- Analytical hierarchy [14],
- Empirical approaches [15],
- Frequency ratio [16],
- Entropy index, certitude factor [17,18],
- Logistic regression analysis [19],
- Conditional analysis method [20],
- Machine-learning methods [21],
- Support vector machine [22],
- Adaptive neuro-fuzzy inference system [23],
- Decision trees [24-26],
- Support vector regression (SVR) [27].

Landslide susceptibility maps were produced in previous studies by using the following approaches in the research area such as:

- GIS-based LSM [28]
- Data-Driven approach [29]
- Weighted overlay method [10]

It is worth noting that if the size of the evaluation area is relatively small, we can use such models [30, 31] to determine the potential hazard for different scenarios. Also, to speed up the simulations, these models can be implemented on GPUs [32]. Nonetheless, for a regional assessment, these models still do not have the computational power. It is possible when the multivariate regression models, the discriminant analysis, and the machine learning models emerge and can provide a useful and reliable solution. Different models generate different evaluation results because each model has its own implementation method, benefits, and limitations. Investigators frequently use comparative techniques to fill in the gaps [18, 33, 34].

A revision of the literature reveals that each machine learning model has its weaknesses and strengths and that the characteristics of various research areas influence its behavior patterns. As a result, machine learning accesses for LSM comparison are highly desirable. We need to explore the excellent deep learning techniques and extend tree-based ensemble classifier applications on a regional scale in LSM. Therefore, we investigate Extreme Gradient Boosting, k-Nearest Neighbors, and Random Forest by tuning the hyper-parameters and verifying and comparing these models for LSM with a case study in the Muzaffarabad area. Ensembles are well-known machine learning algorithms that combine many base learners to provide more accurate prediction outputs [35]. The ensemble approaches can be classified into two categories: boosting and bagging [36].

On the other hand, ensemble learning approaches are built on sequential or parallel base learners, who can be homogenous in boosting and bagging. In recent years, many ensemble learning approaches have been used in LSM to overcome the limits of a single machine learning algorithm [37, 38]. To increase the performance of landslide susceptibility modelling, most of them mix homogenous ensemble frameworks of boosting and bagging with tree-based classifiers. Nine landslide conditioning factors were chosen after the examination of the geo-environmental setting of the research area. A series of statistical measures were used to verify the model's efficiency, including Recall value, Accuracy, the receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC) value.

In this article, two main concise contributions are as follows: the first, the study area, Muzaffarabad, being the capital of Azad Jammu and Kashmir (Pakistan), and the densely populated area is exposed to the threat of landslides, especially after the 2005 earthquake. Therefore, the need to make an accurate Landslide Susceptibility Map for future planning using ML algorithms is imperative to minimize this impending threat subsequently. Secondly, to the best of our knowledge, there is no publication to investigate and compare the tree-based classifiers by tuning the hyper-parameters in the research area for the LSM. The LSM can be used for land-use planning to mitigate risks associated with landslides and is valuable for future general planned construction operations, such as selecting and environmental conservation of the new urban and infrastructural areas.

2. Research Methodology

2.1. Study Area

The Study area covers 77.45 km² in Muzaffarabad District, northern Pakistan (Figure 1). Muzaffarabad is situated at the convergence of the Jhelum and Neelum River. The 94 landslides were mapped using visual interpretation of optical images and field investigations. In the field, some active landslides were captured, shown in Figure 2. The study area's key geographical features are valleys, plains, hills, lakes, and rivers. The weather of the region is damp to the subtropical highlands. The Elevation ranges between 615 and 2014 meters above sea level. The maximum and minimum average temperatures are around 23° and 35°C in the study area for July and 3° and 16° C for January. The average rainfall in the region is 1,511 mm/year [39]. The region consists of active tectonics, steep slopes, and delicate geological situations. Pre-Cambrian, Cambrian, Paleocene to Eocene, Early Miocene, and Quaternary rocks are exposed from a stratigraphic perspective in the investigated area. (Figure 3). A Digital Elevation Model with a resolution of 12.5×12.5 meters produced is used to create variables (source: <http://www.gscloud.cn>). The USGS (<https://earthexplorer.usgs.gov/>) given a Landsat-8 satellite image acquired in December 2020. All the data presented is analyzed for output of QGIS 2.18.16, ArcGIS 10.6, and R-Studio-related variables

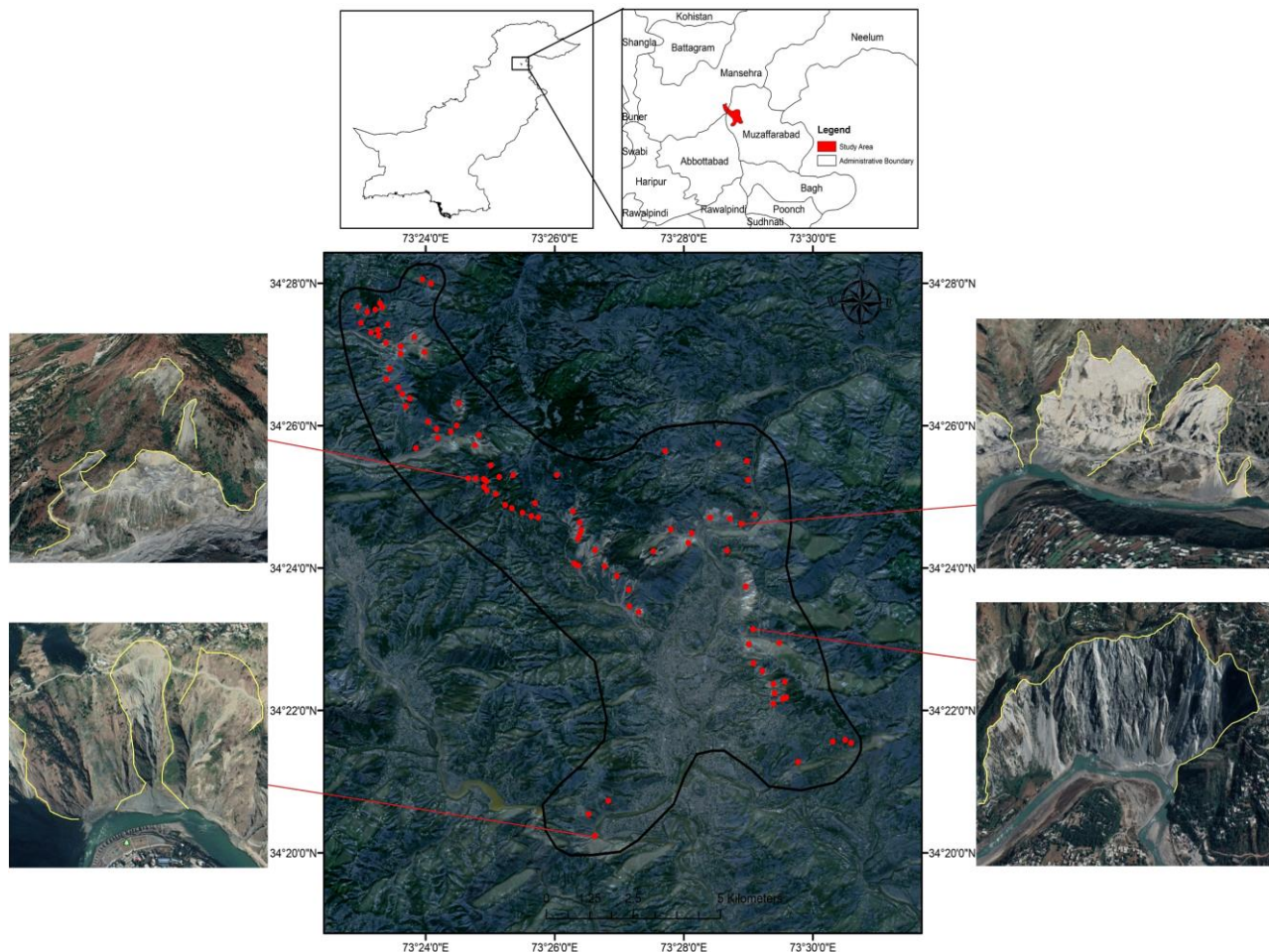


Figure 1. Location map of the study area



Figure 2. Some active landslides in the field

Legend	Formation	Lithology	Symbol
A	Muzaffarabad formation	Stromatolites and cherty dolomites, shales	PcM
B	Paleocene to Eocene Rocks	Bauxite, quartzite, limonite, carbonates, clay, sandstone, and dolomitic limestone, silt, coal	Tp
C	Murree formation	Siltstones, sandstones, clay stones, and reddish-brown shales	Tm
D	Alluvium deposits	Clay, silt, and sand	Qa
E	Hazara Formation	Granite, slate, gypsum, and graphite deposits	PcH

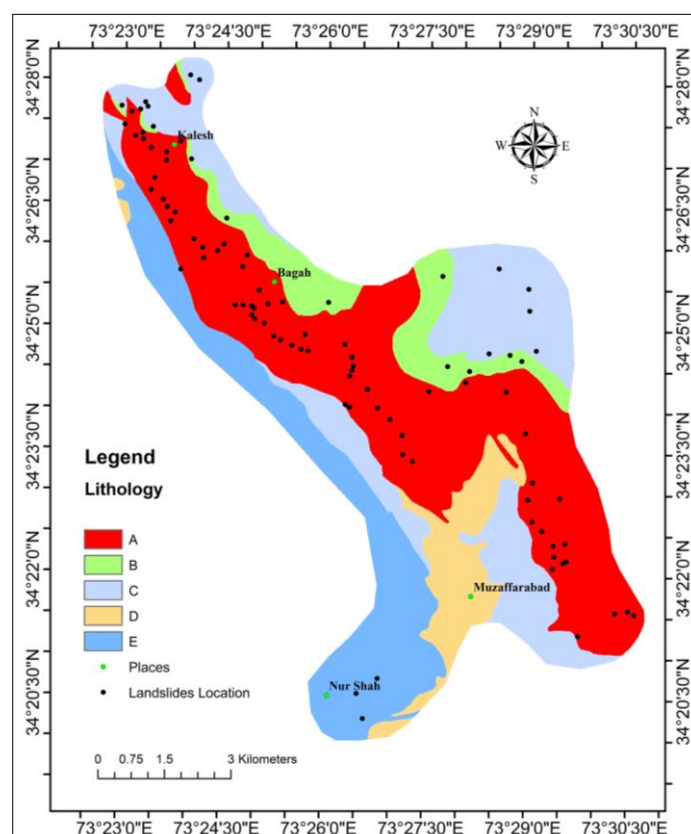


Figure 3. Stratigraphic map of the investigated area

2.2. Landslide Susceptibility Mapping

The workflow is shown in Figure 4.

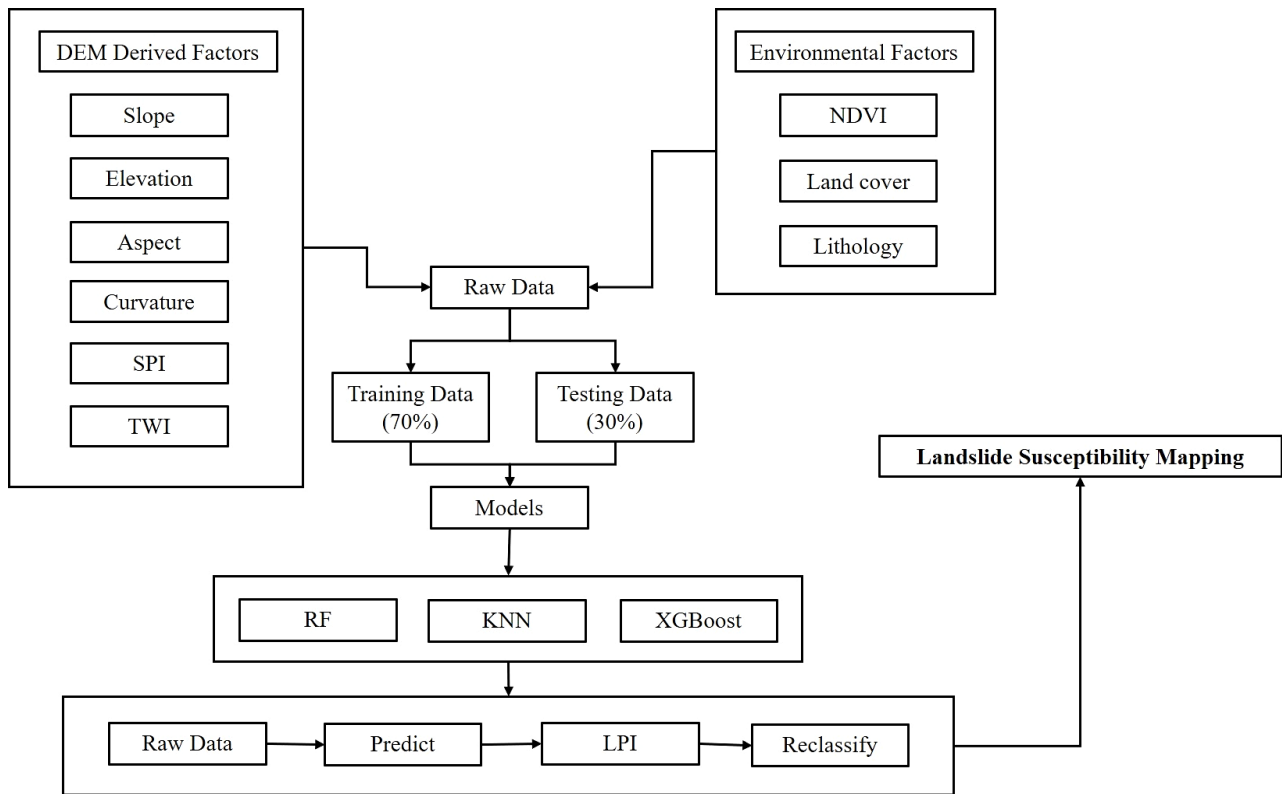


Figure 4. The workflow of the study

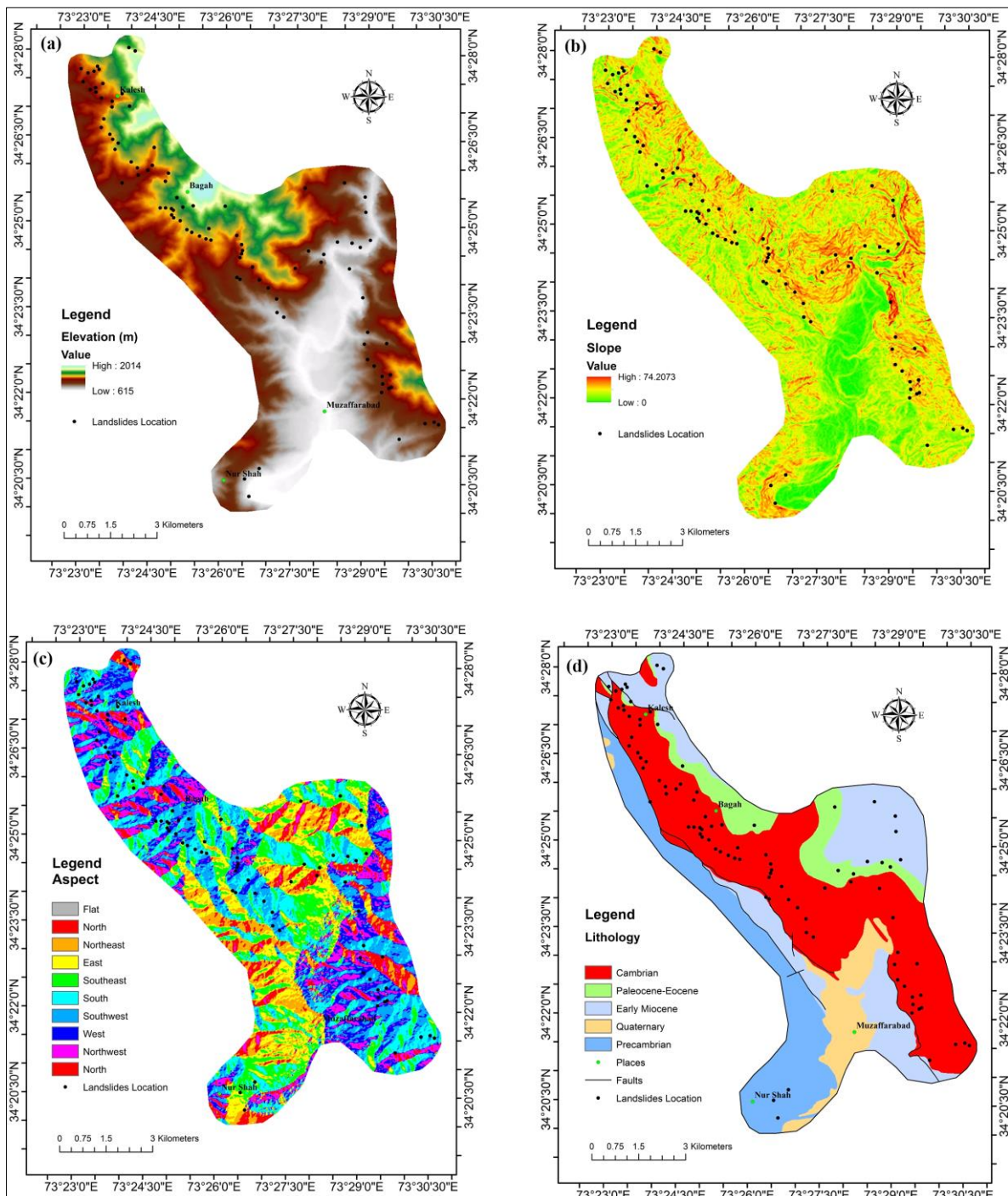
Nine evaluation parameters are used as landslide predisposing factors, whereas the landslide is the dependent variable. A preliminary decision table is created using the circumstances attribute and intention characteristics; the input characteristic settings are then used to form the three models. Eventually, landslide susceptibility mapping is performed the following fundamental principle of landslide sensitivity mapping: The following measures are as follows:

- **Model unit division:** As a model unit in this analysis, the grid unit was used. Remote sensing data and DEM data spatial resolution correspond to 12.5 m, and evaluation factors have all been re-evaluated at 12.5 m.
- **Initial decision table:** A two-dimensional table was created by a condition attribute that corresponds to 9 evaluation factors and a landslide decision attribute (1 represents landslides, 0 represents non-landslides), each line of which defines an object, and each column is the object attribute.
- **Two sections have been changed:** training (70%) and testing the two-dimensional table (30%). The Training data apply to develop the models and to forecast the studies.
- **The evolution of landslide sensitivity:** The three models above were used to calculate all model units in the research field. To generated landslide prediction maps (LPI), each model unit's prediction values per group have been developed.
- **Division of LPI maps:** The results acquired using the three algorithms were importing into GIS, and the Janks natural breakpoint approach [18] was applied to division landslide susceptibility into four grades [40]: low, moderate, high, and very high.
- **Findings are analyzed:** The three models have been extensively tested using the ROC curve and the area under the ROC curve.

The landslide susceptibility assessment has described nine variables with highly explained variances as input factors. Table 1 displays the input variable names, descriptions, and classification, whereas Figures 5 to 7 show the nine landslide factors.

Table 1. The factors, description/Extraction, and categorization of input parameters

Factors	Description/Extraction	Category
Slope	DEM	Topography
Elevation	DEM	Topography
Aspect	DEM	Topography
Lithology	Different types of rocks	Geology
Land cover	Land cover classes	Environmental factor
NDVI	Normalized Different Vegetation Index	Environmental factor
TWI	DEM	Topography
SPI	DEM	Hydrology
Curvature	DEM	Topography

**Figure 5. Landslide Variables used in the investigated area; (a) Elevation (b) Slope (c) Aspect (d) Lithology**

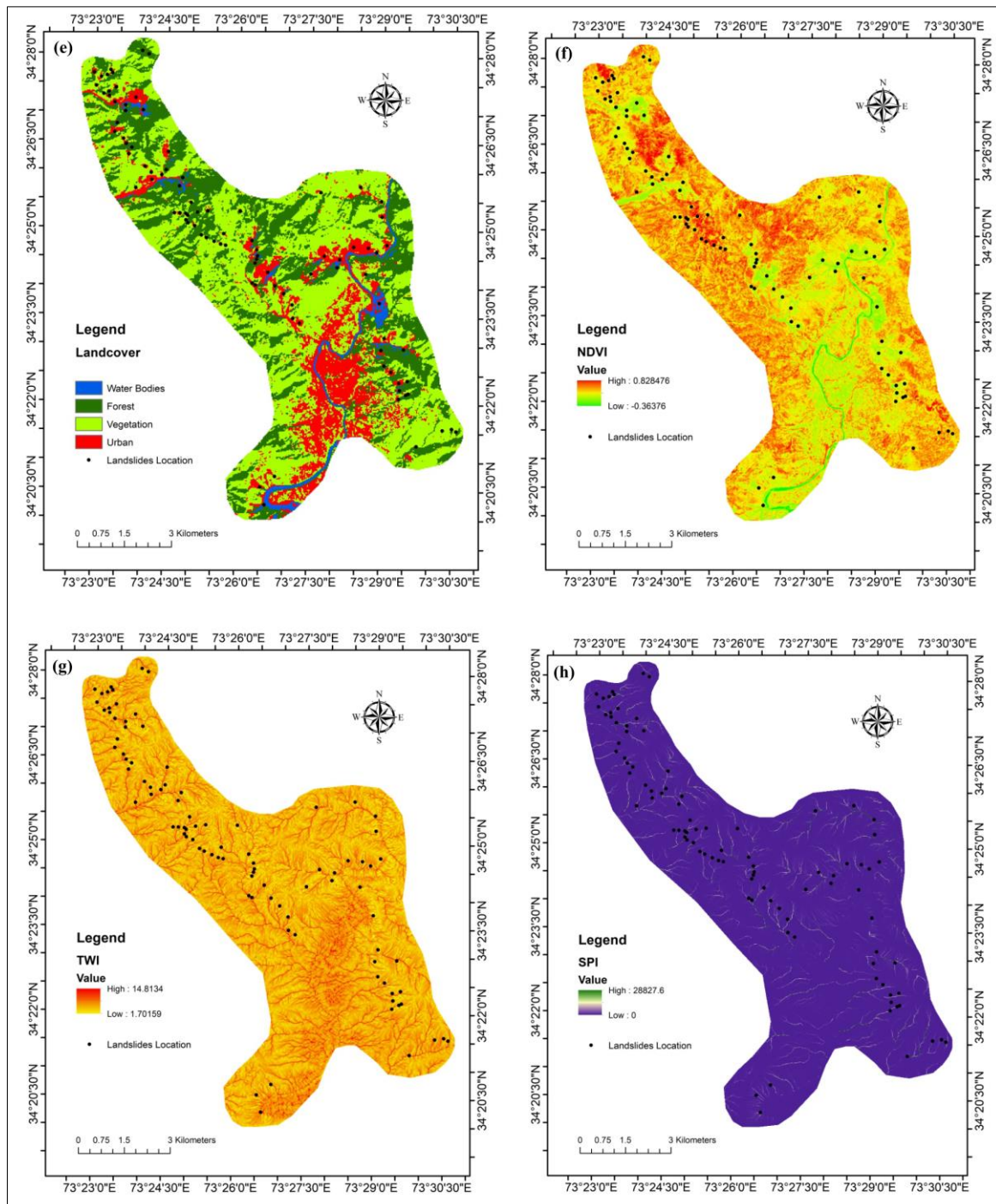


Figure 6. Landslide Variables used in the investigated area; (e) Landcover (f) NDVI (g) TWI (h) SPI

2.3. Random Forest

RF is a commonly used algorithm that is based on the bagging ensemble approach. The RF was selected for an ML algorithm for various classification and regression tasks [41]. The RF model is very flexible and fast. The algorithm is one of the entire tree learning methods based on CART (Tree of Regression and Classification). The number of bootstrap samples is used to balance the number of trees picked. An RF algorithm's key advantage is that it needs minor hyperparameter tuning in for accurate results than the other boosting algorithms [42]. Table 2 displays the RF hyperparameter ranges and maximized values discovered by grid search. In RF models analysis, some hyper-parameters must be tuned, such as:

- The number of trees to be integrated
- The maximum depth of the trees
- The maximum number of features considered at each split

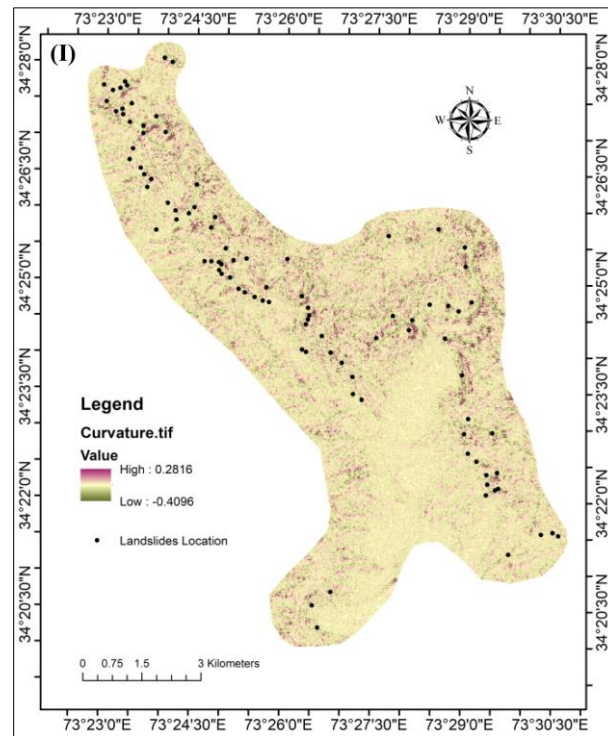


Figure 7. Landslide Variable used in the investigated area, (I) Curvature

Table 2. RF Parameters

Parameters	Values
ntree	500
mtry	10
node size	14

2.4. Extreme Gradient Boosting

XGBoost is a scalable endwise tree boosting system that's applied in a variety of machine learning problems. XGBoost follows the gradient boosting principle that incorporates all the poor Learner predictions [43]. Model XGBoost needs several model preview parameters to be chosen. Three main user-friendly parameters are required: nrounds (maximum number of iterations boosting), colsample_bytree (columns ratios sub-sample when each tree is constructed), and subsamples (the training instance subsample ratio); (Table 3).

Table 3. XGBoost Parameters

Parameters	Values
nrounds	200
colsample_bytree	0.75
Subsample	1
max_depth	6

2.5. k-Nearest Neighbor

k-Nearest Neighbors has been recently used in several fields (e.g., LSM) and is the easiest of all ML algorithms [44, 45]. The input in KNN consists essentially of the k most similar training illustration in the specified interval. The results are a possibility of becoming members; in classified snags, membership possibilities represent the ambiguity of a particular object allocated to a specific class [46]. According to [47], In KNN, an entity is identified by its Neighbor's majority vote; the entity is allocated to the most general class of its closest Neighbors (k is an optimistic number, characteristically minor). If k = 1, the entity is assigned to the next single Neighbor class.

KNN usually is stated as an instance-based, nonparametric, and "lazy" algorithm. Here, the terms 'lazy learning' mention models in which the acquired experience is deferred until an inquiry is posted on the process instead of 'eagerness learning.' The algorithm oversimplifies the training data previously having investigations [48].

The expression 'nonparametric' mentions that the algorithms are not dependent on the dispersion of the data and that the models respond to the data independently. For LSM, where landslide distribution usually doesn't suit smooth distributions, it is beneficial. Lazy learning, on the other hand, implies that the role is just approx. Domestically, and all estimation has been delayed. Consequently, KNN does not learn from training data a discriminatory feature but instead stores the training data set.

2.6. Model Assessment

The capacity of Landslide Susceptibility Models was tested using two statistical methods in this analysis. Specificity and Sensitivity are the percentages of the landslide pixels properly categorized and the proportion of the non-landslide pixels adequately classified in the following equations as Sensitivity and specificity [49];

$$\text{Sensitivity of LSM} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity of LSM} = \frac{TN}{TN + FP} \quad (2)$$

The number of correctly classified pixels is denoted by TP (true positive) and TN (true negative). FP and FN indicate the number of incorrectly classified pixels. The ROC curve is used here, the area under ROC, and the Recall value as appraisal metrics to determine LSM models' capability based on Sensitivity and specificity. The ROC curve, the output of diagnostic signals such as land change, is often measured [47, 48]. Their Sensitivity toward unique characteristics is used through the X-axis with different cut-off levels [24].

Moreover, the AUC reveals that a model can simulate landslides and non-landslides cells. AUC having value 1 illustrates an optimum model; an integer 0 corresponds to an informational model [50]. We also addressed the applicability of the precise value under unbalanced landslide sample conditions. The accuracy and recall value calculation equations are as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (4)$$

3. Results and Discussion

3.1. Important Landslide variables

The selection of predictor variables is one of the essential aspects of landslide susceptibility evaluation and mapping [51]. GIS is frequently utilized to extract important variables for susceptibility evaluation from digital elevation models, such as Topographic Wetness Index (TWI), curvature, Elevation, aspect, and Slope [10, 52]. This research selected nine contributing factors in line with previous literature [29]: Slope, Elevation, Lithology, Curvature, Aspect, TWI, NDVI, SPI, and Landcover (Table 1). The R Studio (Window 4.0.3) was used to determine whether the landslide factors were significant in this analysis or not.

The Slope and Elevation were the essential landslide parameters, while the influence rate of aspect was the minimum (Figure 8). Therefore, in the research of Dai et al. and Ohlmacher et al. [54, 55], it was found that Elevation and Slope were the best predictor variables for estimating the possibility of landslide events. Xiong et al. [53] showed numerous geomorphologic factors (e.g., Elevation) as the maximum significant parameters for predicting the LSM. Hong et al. [49] developed the LSM using J48 Decision Trees, and the results showed that the slope parameter significance was greater than other parameters. Similarly, Dou et al. [54] generated the LSM using advanced random forest and decision tree algorithms, and findings showed that the slope variable was the primary cause of landslides.

Furthermore, this study found that the Muzzafarabad formation was the most prone to landslides, pursued by the Slope. Therefore, the results of our LSM corroborated the findings of Riaz et al. [29] at a certain level. Similarly, Torizin et al. [55] created the LSM for the Torghar and Mansehra districts of Pakistan. Their LSM was created by combining four variables: Landcover, Slope, Lithology, and Distance to the fault. The included variables were similar to the contributive variables considered in this study; nevertheless, we developed the LSM using nine causative variables.

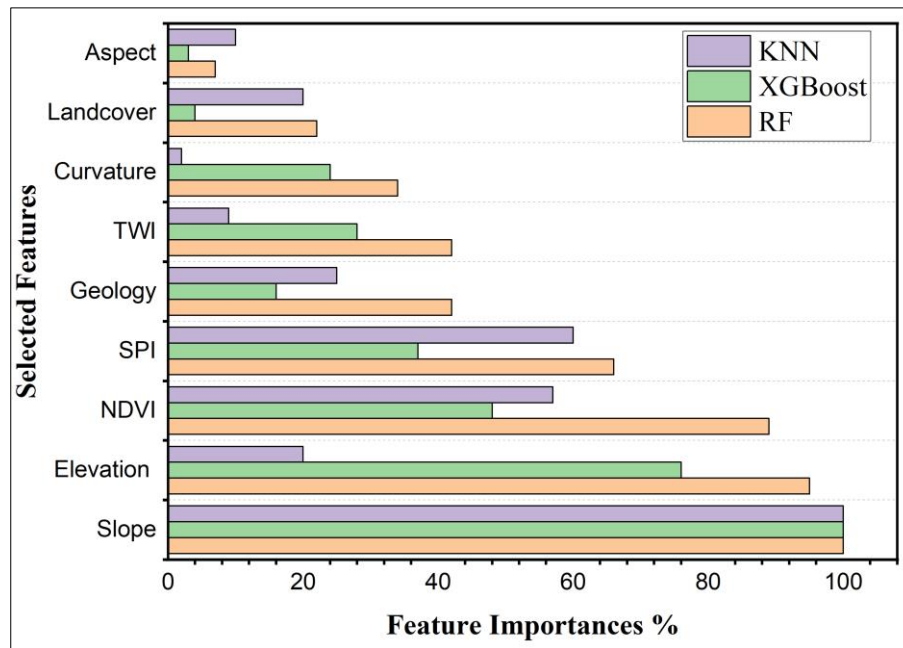


Figure 8. Feature Importance of Landslide Variables graphs according to (a) XGBoost model (b) KNN Model (c) RF Model

3.2. Mapping Results of Landslide Susceptibility

Figure 9 indicates using the three mapping responses graded by the landslide index (LPI). The greater the landslide chance, the greater the LPI. The natural breakpoint method [18] used the class-division probability distribution histogram to divide the LPI into four classes (Jenks method). The decreasing value should be the distribution of possibility, the groups with a loss of balance (<http://www.svds.com>). In this analysis, the LPI was divided into four categories using the natural breakpoint procedure, which considers the histogram.

3.2.1. KNN Landslide Susceptibility Map

The research area, 55.2%, is covered by the KNN model's very high and high susceptibility classes (Figure 9-a). In contrast, the low and moderate vulnerability classes have 23.6% and 21.2% respective areas.

3.2.2. XGBoost Landslide Susceptibility Map

The XGBoost model classified 18.1% of the research area shows very high susceptibility, with values of 17.5%, 24.2%, and 40.2% for low, moderate, and high vulnerability, respectively (Figure 9-b).

3.2.3. RF Landslide Susceptibility Map

The research area, 44.3%, is covered by the RF model's very high and high vulnerability classes (Figure 9-c). On the other hand, the moderate and low vulnerability classes have respective areas of 26.2% and 29.5%.

3.3. Verification and Comparison

Based on the effects of the three algorithms, the Jenks natural breakpoint approach [18] was used to identify landslide vulnerability into four classes [40]: low, moderate, high, and very high, based on the results obtained using the three algorithms. For each algorithm, landslide susceptibility zoning maps were developed using this tool. For the XGBoost model, the LPI ranged from 0 to 1. In normal break conditions, the map of landslide vulnerability was divided into four susceptibility classes: low (0.00–0.17), moderate (0.17–0.43), high (0.43–0.71), and very high (0.71–1). The LPI for the KNN model varied from 0 to 1; four categories of landslide susceptibility maps were reclassified: low (0.00–0.01), moderate (0.01–0.49), high (0.49–0.85), and very high (0.85–1). The LPI ranged from 0 to 1 for the RF model, and the natural split system was used to reclassify a four-grade, low (0.00–0.19), moderate (0.19–0.40), high (0.40–0.63), and very high (0.63–1) susceptibility chart. The ROC curve was being designed to test the model's predictive capability.

To compare the prediction precision of the model, the confusion matrix was used (Table 4). We observed from Table 4 that all models were highly reliable. The precision of these models is 0.858(XGBoost), 0.861(KNN), and 0.856(RF), correspondingly. However, the accuracy was identical, and all three versions tend to be strong on the surface. The evaluation metrics seen in Figure 10 and Table 5 are the ROC curve and AUC. With values of 0.790, the KNN Model had the lowest AUC efficiency. The RF and XGBoost models displayed an AUC value more remarkable than the KNN, with AUC values of 0.895 and 0.893; correspondingly, the RF approach is best predictive.

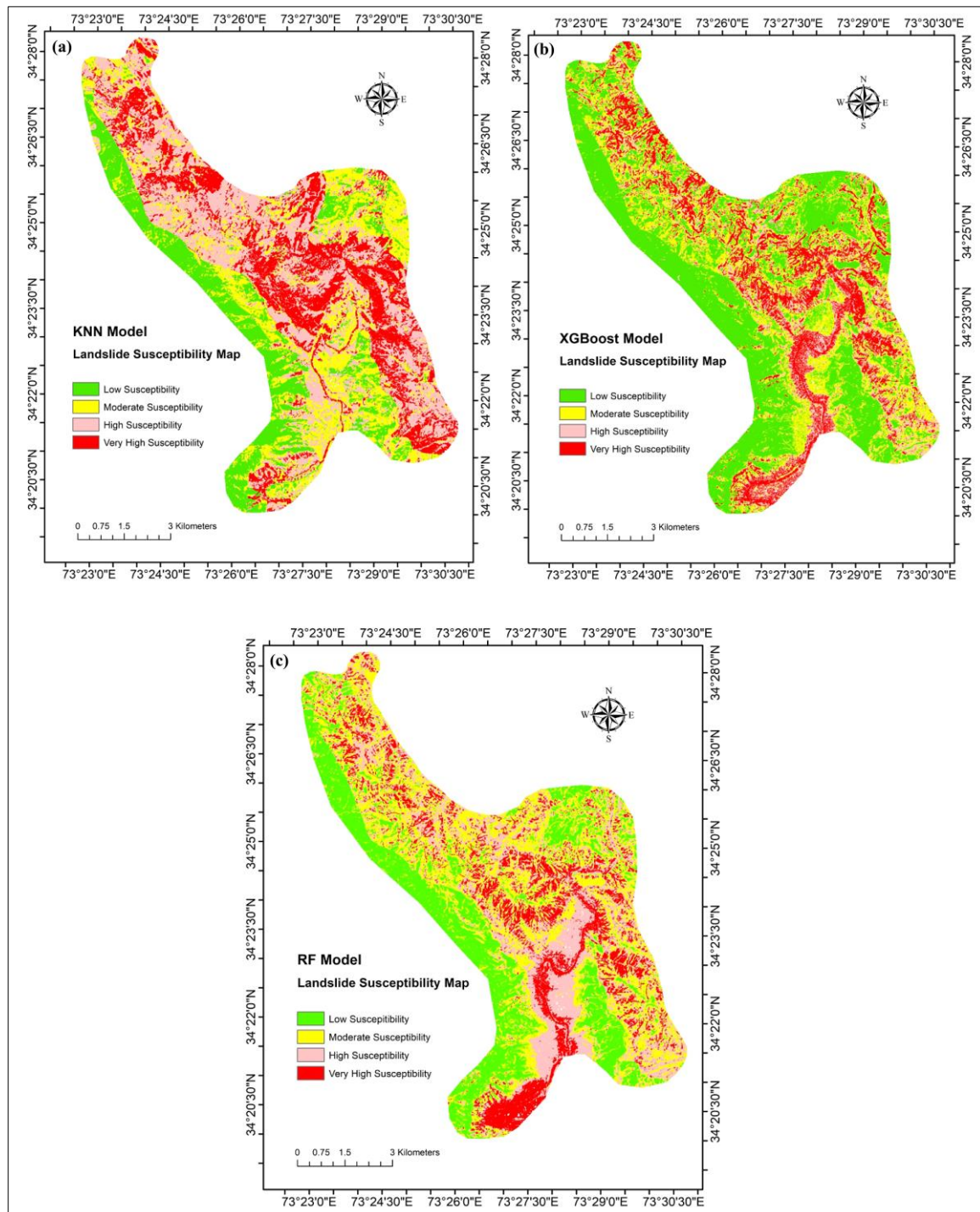


Figure 9. Mapping Results of Landslide Susceptibility of Three Models. (a) LSM using XGBoost Model (b) LSM using KNN Model (c) LSM using RF Model

Table 4. Confusion Matrix of XGBoost, KNN, and RF

Models	Label	Predicted Label		Kappa	Accuracy
		No	Yes		
XGBoost	No	42	11	0.494	0.858
	Yes	52	341		
KNN	No	43	13	0.493	0.861
	Yes	51	339		
RF	No	41	9	0.495	0.856
	Yes	53	343		

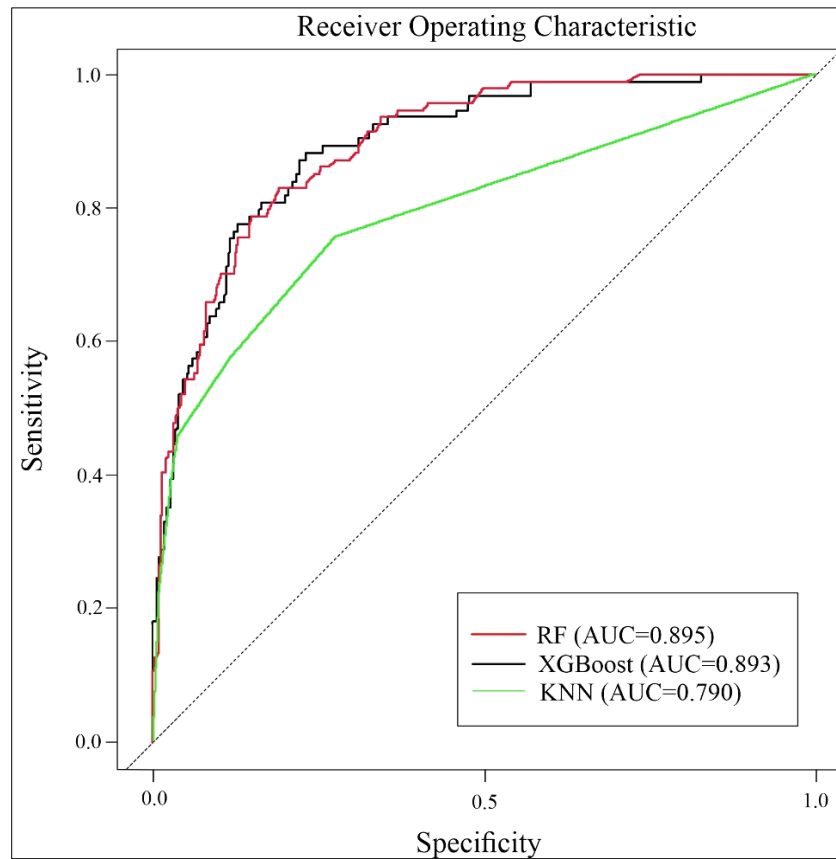


Figure 10. ROC plots of RF, XGBoost, and KNN Models

Table 5. AUC and Recall Value of RF, XGBoost, and KNN models

Parameters	RF	XGBoost	KNN
AUC	0.895	0.893	0.790
Recall Value	0.47	0.48	0.46

The findings show that the XGBoost, KNN, and RF have fair precision for LSM in the studied area. Based on the AUC, RF and XGBoost achieved the highest accuracy compared to KNN, which indicates the results are in line with previous studies [10, 28, 29]. Similarly, other studies have compared several LSM approaches performance to determine the best model [56, 57]. The collaborative model's RF and XGBoost provide exceptional output with a limited range of hyper-parameters to tuning in, but they take a lot of time to run. The XGB, in particular, can process vast amounts of data while using less memory. Unlike the additional tree-based model, the XGB grows trees vertically rather than horizontally [58]. The XGBoost method verified the highest execution and an adequate time of training. The XGBoost is a specific application of the gradient boosting process and uses a memory-efficient technique to improve computational speed and model performance [59]. Whereas the KNN model is easy and basic to perform, its key drawback becomes exponentially slower as the amount of data grows, making it an unfeasible alternative in situations where forecasts must be complete quickly [58]. The Area under curve was used to determine the accuracy of the forecast effects, and AUC denotes the area under the curve bounded by the coordinate axis [60]. The RF model is a well-functioning and reliable model. The strength of RF and XGBoost algorithms is when a large amount of data is missing, RF and XGBoost are suitable methods for estimating missing data and retaining validity [61].

Furthermore, the RF model offers certain benefits and drawbacks. The primary benefits are that the model is not overly sensitive to the variables used to run it, and it is easy to control which variables to use. Random forests tend to exaggerate low values while underestimating high ones, which is its major drawback. When the technique is utilized for regression, this restriction becomes apparent [62].

4. Conclusion

Landslides have been one of Pakistan's most severe natural adversities, causing significant deaths and socio-economic losses annually. So far, the process of landslide mapping has been quite complicated and volatile, making it difficult to predict the vulnerability of landslides accurately and timely in the most susceptible areas. For this purpose, various attempts have been made to enhance predictability based on several forecast models for mapping the landslide

vulnerability targeting different regions. It is essential for decision-makers to create more appropriate landslide vulnerability maps, to enhance the efficiency of the prediction model. In this study, various prediction models have been evaluated and compared, such as XGB, RF, and KNN, to select the most suitable and accurate one to forecast the landslide susceptibility. The model's comprehensibility helps determine the potential association between manipulating Parameters and LSM from data, which is critical for considering the parameters that influence landslide susceptibility. The LSM allows decision-makers to choose appropriate sites for the implementation of development projects. This research has defined the relationship between landslides and geo-environmental factors using three tree-based classifiers: XGBoost, RF, and KNN. The collaborative model's RF and XGBoost provide exceptional output with a limited range of hyper-parameters for tuning in. Based on the three machine learning techniques, the innovative consequences show that the performance of the Random Forest model has a maximum AUC value of 0.895 and is more efficient than the other tree-based classifiers. The contributing factors, such as Slope, and Elevation were the essential landslide parameters, while the influence rate of aspect was the minimum. These findings could be employed as fundamental data to help with land-use development and slope management. The robust results imply that the RF is the most appropriate model, based on the AUC, for mapping landslide vulnerability in the study and its vicinity. Therefore, the stakeholders should employ this model to forecast the landslide susceptibility in Azad Jammu and Kashmir and the adjacent districts more accurately. However, these could be less valuable on a particular site scale where regional geological and geographical diversities may dominate. Additional landslide data is required and implemented in other regions for the model to be widely implemented. The current research offers advanced knowledge to create more robust and detailed predictive models for landslides, which can be minimized in highly susceptible regions.

5. Declarations

5.1. Author Contributions

Conceptualization, M.A.H.; methodology, M.A.H.; software, M.A.H.; validation, N.A.; formal analysis, M.S.; investigation, R.W. and C.M.; resources, Z.C.; data curation, S.U.S.; writing—original draft preparation, M.S.; writing—review and editing, R.W.; visualization, S.U.S. and D.X.; supervision, Z.C. and D.X.; project administration, Z.C.; funding acquisition, Z.C. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The study was supported by the National Natural Science Foundation of China (No. 41871305); National key R & D program of China (No.2017YFC0602204); Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (No. CUGQY1945); Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education; and Fundamental Research Funds for the Central Universities (No. GLAB2019ZR02).

5.4. Conflicts of Interest

The authors declare no conflict of interest.

6. References

- [1] Oh, H. J., & Pradhan, B. (2011). Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Computers and Geosciences*, 37(9), 1264–1276. doi:10.1016/j.cageo.2010.10.012.
- [2] Faraji Sabokbar, H., Shadman Roodposhti, M., & Tazik, E. (2014). Landslide susceptibility mapping using geographically-weighted principal component analysis. *Geomorphology*, 226, 15–24. doi:10.1016/j.geomorph.2014.07.026.
- [3] Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, 2(2), 159–178. doi:10.1080/20964471.2018.1472392.
- [4] Shahabi, H., & Hashim, M. (2015). Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Scientific Reports*, 5, 1–15. doi:10.1038/srep09899.
- [5] Bui, D. T., Shahabi, H., Shirzadi, A., Chapi, K., Alizadeh, M., Chen, W., Mohammadi, A., Ahmad, B. Bin, Panahi, M., Hong, H., & Tian, Y. (2018). Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron Highlands, Malaysia. *Remote Sensing*, 10(10), 1527. doi:10.3390/rs10101527.
- [6] Dou, J., Yunus, A. P., Xu, Y., Zhu, Z., Chen, C. W., Sahana, M., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. *Natural Hazards*, 97(2), 579–609. doi:10.1007/s11069-019-03659-4.

- [7] Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102(3–4), 99–111. doi:10.1016/j.enggeo.2008.03.014.
- [8] Basharat, M., Rohn, J., Ehret, D., & Baig, M. S. (2012). Lithological and structural control of Hattian Bala rock avalanche triggered by the Kashmir earthquake 2005, sub-Himalayas, northern Pakistan. *Journal of Earth Science*, 23(2), 213–224. doi:10.1007/s12583-012-0248-3.
- [9] Dunning, S. A., Mitchell, W. A., Rosser, N. J., & Petley, D. N. (2007). The Hattian Bala rock avalanche and associated landslides triggered by the Kashmir Earthquake of 8 October 2005. *Engineering Geology*, 93(3–4), 130–144. doi:10.1016/j.enggeo.2007.07.003.
- [10] Basharat, M., Shah, H. R., & Hameed, N. (2016). Landslide susceptibility mapping using GIS and weighted overlay method: a case study from NW Himalayas, Pakistan. *Arabian Journal of Geosciences*, 9(4), 1–19. doi:10.1007/s12517-016-2308-y.
- [11] Basharat, M., & Rohn, J. (2015). Effects of volume on travel distance of mass movements triggered by the 2005 Kashmir earthquake, in the Northeast Himalayas of Pakistan. *Natural Hazards*, 77(1), 273–292. doi:10.1007/s11069-015-1590-4.
- [12] Baeten, N. J., Laberg, J. S., Forwick, M., Vorren, T. O., Vanneste, M., Forsberg, C. F., Kvalstad, T. J., & Ivanov, M. (2013). Morphology and origin of smaller-scale mass movements on the continental slope off northern Norway. *Geomorphology*, 187, 122–134. doi:10.1016/j.geomorph.2013.01.008.
- [13] Shafique, M., van der Meijde, M., & Khan, M. A. (2016). A review of the 2005 Kashmir earthquake-induced landslides; from a remote sensing prospective. *Journal of Asian Earth Sciences*, 118, 68–80. doi:10.1016/j.jseas.2016.01.002.
- [14] Komac, M. (2006). A landslide susceptibility model using the Analytical Hierarchy Process method and multivariate statistics in perialpine Slovenia. *Geomorphology*, 74(1–4), 17–28. doi:10.1016/j.geomorph.2005.07.005.
- [15] Wachal, D. J., & Hudak, P. F. (2000). Mapping landslide susceptibility in Travis County, Texas, USA. *GeoJournal*, 51(3), 245–253. doi:10.1023/A:1017524604463.
- [16] Yilmaz, I. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Computers and Geosciences*, 35(6), 1125–1138. doi:10.1016/j.cageo.2008.08.007.
- [17] Zhao, X., & Chen, W. (2020). GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. *Applied Sciences (Switzerland)*, 10(1), 16. doi:10.3390/app10010016.
- [18] Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology*, 11(4), 909–926. doi:10.1007/s13762-013-0464-0.
- [19] Clerici, A., Perego, S., Tellini, C., & Vescovi, P. (2002). A procedure for landslide susceptibility zonation by the conditional analysis method. *Geomorphology*, 48(4), 349–364. doi:10.1016/S0169-555X(02)00079-X.
- [20] Lee, S., Ryu, J. H., Lee, M. J., & Won, J. S. (2003). Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environmental Geology*, 44(7), 820–833. doi:10.1007/s00254-003-0825-y.
- [21] Yao, X., Tham, L. G., & Dai, F. C. (2008). Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology*, 101(4), 572–582. doi:10.1016/j.geomorph.2008.02.011.
- [22] Jaafari, A., Rezaeian, J., & Omrani, M. S. (2017). Spatial prediction of slope failures in support of forestry operations safety. *Croatian Journal of Forest Engineering*, 38(1), 107–118.
- [23] Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., & Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena*, 187, 104396. doi:10.1016/j.catena.2019.104396.
- [24] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147–160. doi:10.1016/j.catena.2016.11.032.
- [25] Wang, G., Lei, X., Chen, W., Shahabi, H., & Shirzadi, A. (2020). Hybrid computational intelligence methods for landslide susceptibility mapping. *Symmetry*, 12(3), 325. doi:10.3390/sym12030325.
- [26] Rahmati, O., Falah, F., Naghibi, S. A., Biggs, T., Soltani, M., Deo, R. C., Cerdà, A., Mohammadi, F., & Tien Bui, D. (2019). Land subsidence modelling using tree-based machine learning algorithms. *Science of the Total Environment*, 672, 239–252. doi:10.1016/j.scitotenv.2019.03.496.
- [27] Dao, D. Van, Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., Phong, T. Van, Ly, H. B., Le, T. T., Trinh, P. T., Luu, C., Quoc, N. K., Thanh, B. N., & Pham, B. T. (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena*, 188, 104451. doi:10.1016/j.catena.2019.104451.

- [28] Kamp, U., Growley, B. J., Khattak, G. A., & Owen, L. A. (2008). GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology*, 101(4), 631–642. doi:10.1016/j.geomorph.2008.03.003.
- [29] Riaz, M. T., Basharat, M., Hameed, N., Shafique, M., & Luo, J. (2018). A Data-Driven Approach to Landslide-Susceptibility Mapping in Mountainous Terrain: Case Study from the Northwest Himalayas, Pakistan. *Natural Hazards Review*, 19(4), 05018007. doi:10.1061/(asce)nh.1527-6996.0000302.
- [30] Juez, C., Murillo, J., & García-Navarro, P. (2013). 2D simulation of granular flow over irregular steep slopes using global and local coordinates. *Journal of Computational Physics*, 255, 166–204. doi:10.1016/j.jcp.2013.08.002.
- [31] Iverson, R. M., & George, D. L. (2016). Modelling landslide liquefaction, mobility bifurcation and the dynamics of the 2014 Oso disaster. *Geotechnique*, 66(3), 175–187. doi:10.1680/jgeot.15.LM.004.
- [32] Lacasta, A., Juez, C., Murillo, J., & García-Navarro, P. (2015). An efficient solution for hazardous geophysical flows simulation using GPUs. *Computers and Geosciences*, 78, 63–72. doi:10.1016/j.cageo.2015.02.010.
- [33] Li, R., & Wang, N. (2019). Landslide susceptibility mapping for the Muchuan County (China): A comparison between bivariate statistical models (WoE, EBF, and IoE) and their ensembles with logistic regression. *Symmetry*, 11(6), 762. doi:10.3390/sym11060762.
- [34] Wang, Q., Li, W., Chen, W., & Bai, H. (2015). GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang county of Baoji city, China. *Journal of Earth System Science*, 124(7), 1399–1415. doi:10.1007/s12040-015-0624-3.
- [35] Dietterich, T. G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1857 LNCS, 1–15. doi:10.1007/3-540-45014-9_1.
- [36] Graczyk, M., Lasota, T., Trawiński, B., & Trawiński, K. (2010). Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5991 LNAI(PART 2), 340–350. doi:10.1007/978-3-642-12101-2_35.
- [37] Pham, B. T., Tien Bui, D., & Prakash, I. (2017). Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. *Geotechnical and Geological Engineering*, 35(6), 2597–2611. doi:10.1007/s10706-017-0264-2.
- [38] Pham, B. T., Phong, T. Van, Nguyen-Thoi, T., Parial, K., K. Singh, S., Ly, H. B., Nguyen, K. T., Ho, L. S., Le, H. Van, & Prakash, I. (2020). Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto International*, 1–23. doi:10.1080/10106049.2020.1737972.
- [39] Khan, U., & Rehman, D. H. (Eds.). (2017). *Landslide Hazard, Vulnerability and Risk Analysis of Muzaffarabad City*. doi:10.4172/978-1-63278-001-0-002.
- [40] Chen, J., Yang, S., Li, H., Zhang, B., & Lv, J. (2013). Research on geographical environment unit division based on the method of natural breaks (Jenks). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(4W3), 47–50. doi:10.5194/isprsarchives-XL-4-W3-47-2013.
- [41] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. doi:10.1007/bf00058655.
- [42] Alsahaf, A., Azzopardi, G., Ducro, B., Veerkamp, R. F., & Petkov, N. (2018). Predicting Slaughter Weight in Pigs with Regression Tree Ensembles. In *Proceedings of APPIS*; pp. 1-9, doi:10.3233/978-1-61499-949-4-1.
- [43] Sahin, E. K. (2020). Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto International*, 1–25. doi:10.1080/10106049.2020.1831623.
- [44] Marjanović, M., Bajat, B., & Kovačević, M. (2009). Landslide susceptibility assessment with machine learning algorithms. *International Conference on Intelligent Networking and Collaborative Systems, INCoS 2009*, 273–278. doi:10.1109/INCOS.2009.25.
- [45] Miner, A., P. Vamplew, D. J. Windle, Flentje, P., & P. Warner. (2010). A Comparative Study of Various Data Mining Techniques as applied to the Modeling of Landslide Susceptibility on the Bellarine Peninsula. In *Faculty of Engineering - Papers* (pp. 1327–1336).
- [46] Bröcker, J., & Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, 22(3), 651–661. doi:10.1175/WAF993.1.
- [47] Chen, J. S., Huang, H. Y., & Hsu, C. Y. (2020). A kNN Based Position Prediction Method for SNS Places. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12034 LNAI, 266–273. doi:10.1007/978-3-030-42058-1_22.
- [48] Kala, R. (2012). Multi-robot path planning using co-evolutionary genetic programming. *Expert Systems with Applications*, 39(3), 3817–3831. doi:10.1016/j.eswa.2011.09.090.

- [49] Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A. X., Chen, W., & Ahmad, B. Bin. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399–413. doi:10.1016/j.catena.2018.01.005.
- [50] Tien Bui, D., Tuan, T. A., 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(2), 361–378. doi:10.1007/s10346-015-0557-6.
- [51] Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., & Chacón, J. (2012). Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: Application to the river Beiro basin (Spain). *Natural Hazards and Earth System Science*, 12(2), 327–340. doi:10.5194/nhess-12-327-2012.
- [52] Süzen, M. L., & Doyuran, V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environmental Geology*, 45(5), 665–679. doi:10.1007/s00254-003-0917-8.
- [53] Xiong, K., Adhikari, B. R., Stamatopoulos, C. A., Zhan, Y., Wu, S., Dong, Z., & Di, B. (2020). Comparison of different machine learning methods for debris flow susceptibility mapping: A case study in the Sichuan Province, China. *Remote Sensing*, 12(2), 295. doi:10.3390/rs12020295.
- [54] Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C. W., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, 662, 332–346. doi:10.1016/j.scitotenv.2019.01.221.
- [55] Torizin, J., Fuchs, M., Awan, A. A., Ahmad, I., Akhtar, S. S., Sadiq, S., Razzak, A., Weggenmann, D., Fawad, F., Khalid, N., Sabir, F., & Khan, A. J. (2017). Statistical landslide susceptibility assessment of the Mansehra and Torgar districts, Khyber Pakhtunkhwa Province, Pakistan. *Natural Hazards*, 89(2), 757–784. doi:10.1007/s11069-017-2992-2.
- [56] Dou, J., Yunus, A. P., Bui, D. T., Sahana, M., Chen, C. W., Zhu, Z., Wang, W., & Pham, B. T. (2019). Evaluating gis-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the lidar dem. *Remote Sensing*, 11(6), 638. doi:10.3390/rs11060638.
- [57] Thai Pham, B., Tien Bui, D., & Prakash, I. (2018). Landslide susceptibility modelling using different advanced decision trees methods. *Civil Engineering and Environmental Systems*, 35(1–4), 139–157. doi:10.1080/10286608.2019.1568418.
- [58] Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225. doi:10.1016/j.earscirev.2020.103225.
- [59] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794. doi:10.1145/2939672.2939785.
- [60] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. doi:10.1016/j.patrec.2005.10.010.
- [61] Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, NJ, United States.
- [62] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324.