IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

Landslide Detection of Using Ensemble Classifiers

R.Sindhuja[1]*, A. Padmapriya[2]

[1,2]Department of Computer Science, Alagappa University, Karaikudi, Tamilnadu, India

Corresponding Author: sindhu.ramu.202@gmail.com, Tel.: +91 8098704288

DOI: https://doi.org/10.26438/ijcse/v7i7.202206 | Available online at: www.ijcseonline.org

Accepted: 17/Jul/2019, Published: 31/Jul/2019

***Abstract*—** Landslides play an important role in this world. The landslide affects hundreds and thousands of people and is economically vulnerable. The causes of landslides are mainly caused by rain, earthquake and so on. This paper helps to use the method for classification of landslide detection. Subsequently, the construction of the classification algorithm depends on the global-landslide dataset. The main purpose of this method to improve the performance of machine learning ensemble classifiers is to perform better in terms of classification accuracy and execution time based on multiboost, bagging, subspace discrimination and subspace KNN.

***Keywords*—** landslide, clustering techniques, Machine Learning, k-means, ensemble classifiers**.**

I. INTRODUCTION

The goal of machine learning is to adapt the system to their experience. A branch of Artificial Intelligence is concerned with the study and analysis of systems that can learn from the machine and data [19]. With the help of machine learning we can find out the future information. In the Industry 4.0 revolution, the impact of Machine Learning is a vital one. It is usually applied to more complex real world problem scenarios. This research work considers the Landslides as the area of interest. Landslides are occurring around the world. This is the most dangerous part of human life. A landslide is defined as a large rock, debris or earth movement beneath the earth. An earthquake and other factors can trigger underwater landslides[10,12,17]. These landslides are called submarines. Submarine landslides sometimes occur in volatile damaging tsunamis.

One of the techniques used in data processing is clustering, a method of finding related groups of objects or data in a data set. Clustering is the process of dividing a set of objects or data into groups of related subclasses [18]. Due to the diversity of applications, clustering analysis is a major issue in data mining. The main goal of the clustering technique is to separate the different parts of the cluster with their feature. The number of groups is represented by the K variable [15]. This means that clustering analysis is one of the best techniques for analysing a dataset. Recently, ensemble classifiers have received increasing attention in many fields because they improve the accuracy of the models and deal with complex and high-dimensional data. Various ensemble classifiers such as bagging, boosting, subspace KNN, and multiboost have been proposed. Ensemble classifiers perform better in terms of classification accuracy and implementation time based on this performance.

The structure of the paper is as follows: Section II describes the landslide detection in various areas as a literature review. In Section III the algorithm of the proposed method with flow diagram is illustrated. Section IV describes the dataset. Section V presents a comparison of the accuracy of classifying all sizes of landslides using different methods. Section VI presents an analysis of the results, while section VII presents the conclusion.

II. RELATED WORK

Maneesha.V. Ramesh proposed a research work on landslide detection using Wi-Fi, multichannel network and satellite communication [1].

Yang Hong, Robert F. Adler, George Huffman assesses the state of rainfall that triggered landslides around the world, which died in a landslide disaster of 8 years. The dataset is compiled from geological datasets and compares the results produced from that global landslide map [2].

S. Karthik, K. Yogesh, Y.M. Jagadish, RK Satheendran implemented a low cost energy harvesting wireless sensor network for landslide detection. This system uses solar energy harvesting to provide the sensor. The system uses a super condenser for the purpose of saving energy harvested. The system is all handled by the base station [3].

VN Deekshit, Maneesha, Vinodoni Ramesh, P.K. Indukala, G Jayachandran Nair proposed a research wok on network design and algorithm. It uses the Arduino-based data acquisition system with the Geophone network [4].

Paraskevas Sangaratos Ioana Ilia presented a paper to monitor hazardous landslides and alert them to avoid the risk of landslides by measuring parameters. This paper discusses a supervised machine learning spatial tool using a pixel-based technique representing a naive Bayes classifier [5].

Dieu Tien Bui, Tien-chung Ho proposed a method to predict rain-induced landslides using the ensemble framework. Here the data are obtained from the Geographic Information System. Adaboost, bagging and multiboost ensemble frameworks are used. The results of this study are used to predict landslide detection areas [6].

Abdul Rahaman Wahab Sait, Dr. T. Meyappan proposed a studies work on facts pre-processing and transformation techniques for web log records. This paper discusses a pre-implementation and transformation method and the outcomes display the effectiveness of the work [7].

A.Sumathi,S. Santhosh Kumar, Dr. N.K. Sakthivel proposed a technique an efficient classifier referred to as ICS4-MFMW, which is focusing both the dimensionality reduction and indecisive reputation. The result of this work at is used to carry out higher in comparison with the existing MFMW classifier [8].

S. Neelamegam, Dr. E.Ramaraj proposed a different classification technique in data processing. In this paper, present the basic classification techniques. Decision tree induction, Bayesian networks, K-nearest neighbor classifier, and many other important classification system including neural network [9].

III. METHODOLOGY

In this study, machine learning-based ensemble classifier has been proposed for landslide data. The main idea is illustrated in figure 1. The first step deals with data collection using a dataset from landslide detection areas across India. Landslide dataset is split into training and verification datasets. The second step deals with the extract aspects of machine learning. Attributes of the dataset such as event-time, trigger, landslide level, latitude, and longitude are used to identify the type of landslides. The third step deals with behavioral classification for landslide range impact using selective classifiers. Finally, the landslide limit dataset is classified as low, moderate, and highly flexible to create curve fitting equations to estimate the impact of landslide areas in the form of accuracy.

In the proposed approach, ensemble classifiers such as subspace discriminant, boosting, bagging, subspaces KNN are used. The K-object clustering technique, which uses linear functionality, is applied for the dataset. Similarly, boosting, bagging, subspace KNN and discrimination are the ensemble classifiers that have the potential to significantly improve the performance of prediction models. The predictive performance of ensemble models is approached using training and validation datasets, statistical evaluation measures, receiver operating characteristic curve (*ROC*), and area under the curve (*AUC)*.

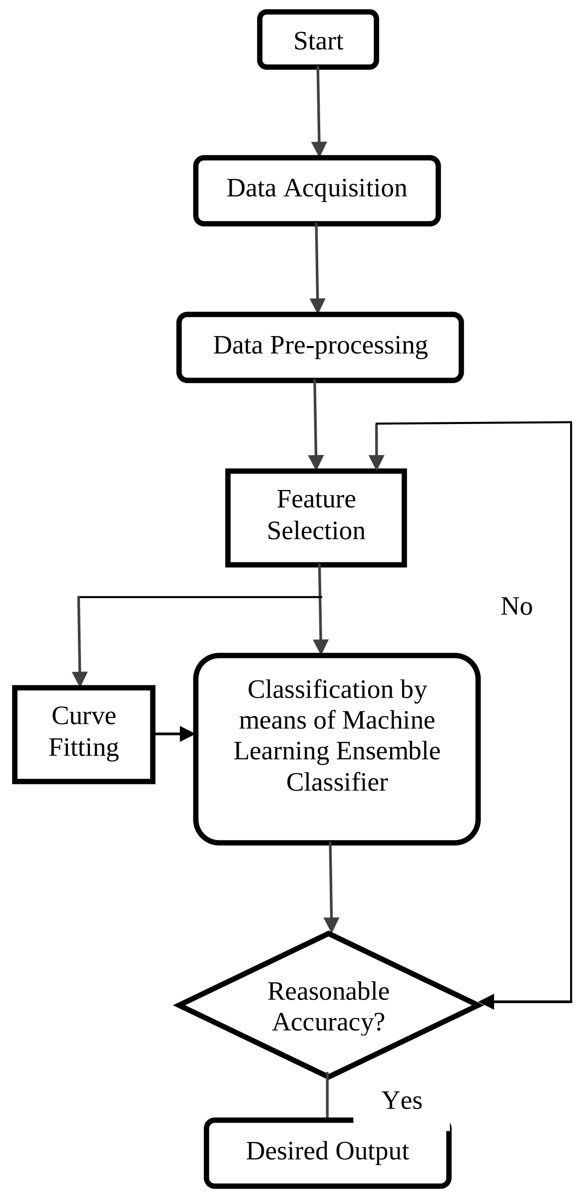

Figure 1. Dataflow Diagram for proposed Methodology

IV. DATASET DESCRIPTION

The landslide dataset contains 29 attributes. Among them, important attributes including ID, date, time, continent-code, country name, code, state, population, city, distance, location, latitude, longitude, geo-location, hazard type, landslide type, magnitude and trigger are identified in the pre-processing stage. The values of the attribute are categorized into two types: namely unique and continuous values.

Q	R	S	T	U	V	W
latitude	longitude	ctry_nam	ctry_code	div_name	gaz_point	gaz_dist
44.9184	-120.27	United St	US	Oregon	Kinzua	18.34
50.4972	-4.21	United K	GB	England	Plymouth	10.57
31.7501	110.681	China	CN	Hubei	Nan He	13.49
27.0087	88.443	India	IN	West Bei	Kalimpor	8.03
10.3904	124.985	Philippine	PH	Southern	Sogod	0.07
20.8327	-156.14	United St	US	Hawaii	Kailua	11
46.978	-123.8	United St	US	Washingt	Aberdeei	3.58
12.1969	124.553	Philippine	PH	Samar	Oquendo	7.38
25.52	91.27	India	IN	Meghala	Nongstoi	1.66
-7.4011	109.811	Indonesi	ID	Jawa Tei	Wonosob	10.6
41.3639	41.6822	Turkey	TR	Artvin	Borcka	0.75
-33.35	-70.517	Chile	CL	RegiÃ³n	Eulogio S	11.7
30.9202	103.621	China	CN	Sichuan	Guan Xia	9.27
53.3319	-132.41	Canada	CA	British C	Queen Cl	23.96
-7.8972	112.877	Indonesi	ID	Jawa Tim	Wonokiti	3.1
37.3632	-122.17	United St	US	Californi	Moffett I	12.23
-9.4446	159.97	Solomon	SB	Guadalca	Lunga Po	7.99
31.4521	78.4013	India	IN	Himachal	Chitkal	10.72
36.2391	-121.77	United St	US	Californi	Carmel	37.28

Figure 2. Global distribution landslide dataset catalogue

The dataset contains 11,338 records. In this dataset, the country code contains over 145 records; there are 1532 records in India in particular. There are 49 posts divided in India. Gaz_Distance is classified as unique and continuous values in this global landslide dataset.

V. METHODOLOGY

A. Preparation of Training and Validation data

The landslide detection dataset was randomly divided into two subgroups. The first subset (9,806 records) is used for building the prediction model, and the second one (1532 records) is used for testing the model. Landslide detection using data mining methods can be considered a binary classification: Therefore, they require both positive data (e.g. current case, Presence of landslides) and negative data (e.g., landslides absent). Landslide alignment factors are then extracted to create training and validation datasets.

B. Ensemble Learning Algorithms

This section briefly describes three groups of learning instructions, bagging, and Subspace discriminant and boosting, subspace KNN used to establish ensemble models for landslides in this study [6, 11].

1) Bagging:

Bagging has been shown to be effective. It is sensitive in landslide-prone samples Small changes in training data, therefore, may be Ability to improve the predictive ability of the model. The process of bagging process consists of three steps: (1) first, the bootstrap samples are obtained by approximating the training. A dataset to create a set of training subgroups; (2) Then, many Classification-based models are developed based on each Subcommittee; and (3) more recently, the final model was developed Aggregates all classifier-based models.

2) Boosted Trees:

The practices of boost Instruction: (1) First, a subset is created After training dataset and initial classifier based model Events are built in an equally reserved space Weights; (2) The initial model is used to predict all events And unclassified events in the training dataset Weights need to be embedded, whereas weights There are correctly classified events; (3) Next Step, weights of all the events in the training dataset The normalized and new subset is then randomly sampled Develop the next classifier-based model. This process continues until it reaches a stop state [14]. The final model is obtained on a weight basis which is the sum of all the classification-based models.

3) Subspace Models:

The Procedures Subspace algorithm is as follows: (1) using the training dataset; Configuration is carried out to configure random selection [16]. A set of training subgroups, which then use them to create classifiers -Based models; (2) Resetting event weights according to the overall accuracy performance of the classifier Models; (3) Continuous sampling of new subgroups Weight for training new classifier based models. The result is a group of classifiers.

C. Performance Assessment and Comparison of Landslide Ensemble Models

The accuracy measures are commonly used to evaluate the overall performance of any classification model.

Accuracy = TP/ (TP+FN)

□□□

Where, True Positives (*TP)* and True Negatives (*TN*) are the number of correctly classified test cases. False positives (*FP*) and False Negatives (*FN*) are numbers of test cases misclassified. The ROC curve shows true positive rate (*TPR*) versus false positive rate (*FPR*) for a trained classifier, where *TPR* and *FPR* can be calculated as follows

True Positive Rate= Tp /Tp+FN 1-false Negative rate (2)

False Positive Rate= Tp /Tp+FN 1-True Negative rate (3)

Overall effectiveness of landslides Models are rated by the receiver operating characteristic (*ROC*) curve. *ROC* curve maps are built using true positives against false positives in two dimensions Space. Both *TPR* and *FPR* range from *0* to *1*, and the *AUC* (Area Under Curve) ranges from *0.5* to *1*.If the *AUC* value is equal to *1* then there is a perfect model and when *AUC* is equal to *0*, it represents a non-informational model. Since the *AUC* value is close to *1*, it is good for landslides Model [11, 13].

Table 1: AUC (Area Under Curve) information Gain for the Landslide containing Ensemble Models.

Ensemble Models	Area Under Curve	Classification Accuracy in percentage
Boosted Trees	1	100%
Bagged Trees	0.72	72%
Subspace Discriminant	1	100%
Subspace KNN	0.56	56%

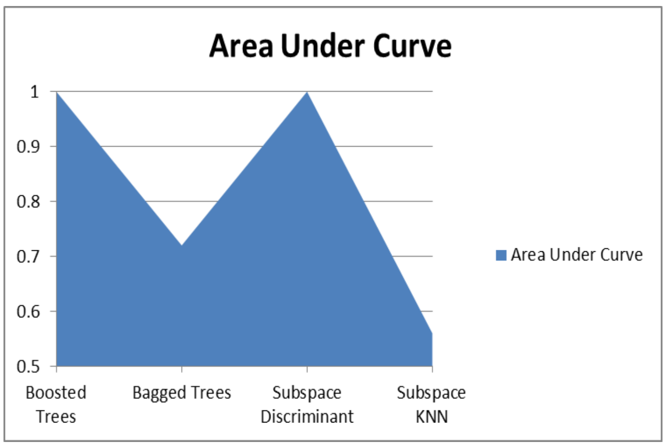

Figure 3: 100% classification accuracy obtained by Boosted Trees & Subspace discriminant.

VI. RESULT AND ANALYSIS

The purpose of this study is to propose and validate the ensemble classifiers (bagging, boosting, subspace, subspace KNN) for machine learning modelling, and K-means clustering as a pre-processing base classifier, and the results are shown in Table 2

Table 2: Accuracy for classification of all size for landslide using various ensemble classifiers.

Parameters	Boosting	Bagging	Subspace Discriminant	Subspace KNN
Large	0.54	0.60	0.55	0.54
Medium	0.51	0.57	0.50	0.60
Small	0.55	0.61	0.53	0.57
Very Large	0.64	0.79	0.80	0.50
Unknown	0.64	0.67	0.64	0.70

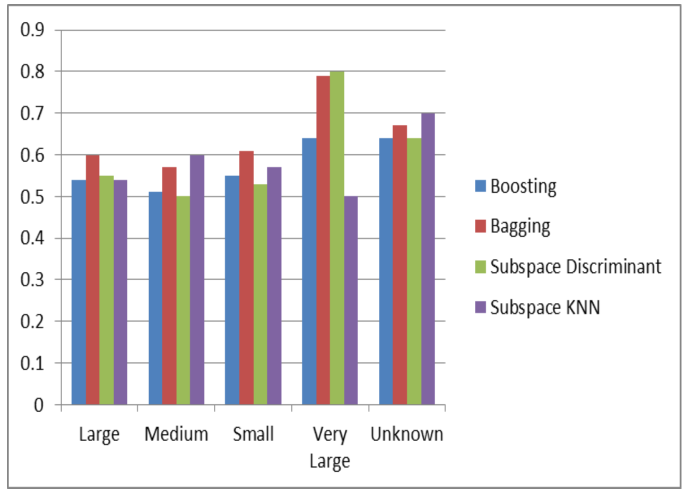

Figure 4: : Bar graphs showing the overall performance of landslide range values using different ensemble classifiers.

Table 3: Experimental Data based Accuracy and Relative Errors.

Ensemble Classifiers	Accuracy Rate	Error Rate
Boosted trees	75.6%	24.4%
Bagged trees	72.5%	27.5%
Subspace discriminant	75.6%	24.4%
RUS boost	46.8%	53.2%
Subspace KNN	69.4%	30.6%

Samples evaluated using ensemble methods such as accuracy for validation datasets. Comparison of the sample results indicates the highest precision boosted trees *(75.6%)* followed by subspace discriminant (*75.6%*), bagged trees (*72.5%)* and the lowest subspace KNN (*69.4%*). The categorical accuracy is almost equal to the boosted and subspace discrimination model (*75.6%).*

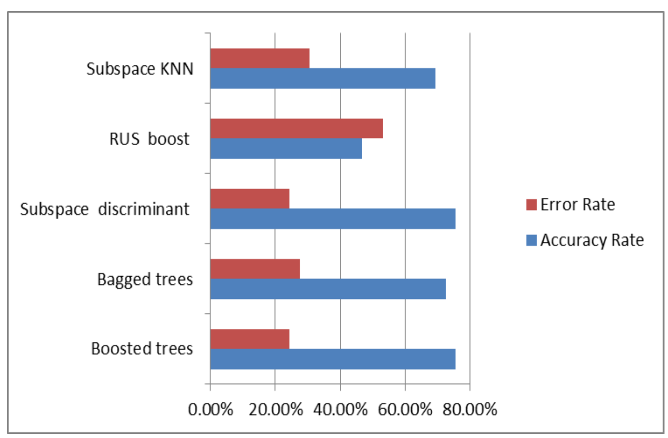

Figure 5: The results shows the data-based accuracy and relative errors of the various ensemble classifiers.

VII. CONCLUSION AND FUTURE SCOPE

In this study, the boosted and subspace discriminant model was used to evaluate the accuracy of the ensemble models. Low, medium, high, moderate: group model obtained using the boosted and subspace methods, respectively. By comparing the relative classification rate of the verification data set and the accuracy value of the *ROC* curve, the boosted trees achieved *75.5*% and subspace discriminant *75.5%*, respectively. Therefore, the boosted and subspace results are selected for the best models to participate in the landslide range optimization. In conclusion, the results of this study are useful for predicting the accuracy of landslide impact areas. Future work for this research is to further improve the accuracy of the assessment results by taking into account recent landslide events to reduce landslide risk.

REFERENCES

[1] Maneesha V. Ramesh, *"Real-time Wireless Sensor Network for Landslide Detection",* 2009 Third International Conference on Sensor Technologies and Applications, 2009.

[2] Yang Hong, Robert F. Adler, and George Huffman,*"An Experimental Global Prediction System for Rainfall-Triggered Landslides Using Satellite Remote Sensing and Geospatial Datasets",* IEEE Transactions On Geoscience And Remote Sensing, Vol. 45, No. 6, June 2007.

[3] S.Karthik, K.Yokesh, Y.M.Jagadeesh, R.K.Sathiendran*, "Smart Autonomous Self Powered Wireless Sensor Networks based Low-cost Landslide Detection System*" 2015 International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2015.

[4] Deekshit V N, Maneesha Vinodoni Ramesh, Indukala P.K, and G. Jayachandran Nair, *"Smart Geophone Sensor Network for Effective Detection of Landslide Induced Geophone Signals"* International Conference on Communication and Signal Processing, India,2016.

[5] Paraskevas Tsangaratos, Ioanna K. Ilia, *"A Supervised Machine Learning Spatial tool for detecting terrain deformation induced by landslide phenomena*", Proceedings of the International Congress of the Hellenic Geographic Society, Oct 2014, Greece.

[6] Dieu Tien Bui, Tien-Chung Ho, Biswajeet Pradhan, Binh-Thai Pham, Viet-Ha Nhu, Inge Revhaug, *"GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks*", Enviromental Earth Sciences 75(14), July 2016.

[7] Abdul Rahaman Wahab Sait, Dr.T.Meyappan, *"Data Preprocessing and Transformation technique to Generate Pattern from the Web Log"* International Journal of Computer Sciences and Information Systems (ICSIS 2014), Oct 17-18, 2014 Dubai (UAE).

[8] A. Sumathi,S. Santhoshkumar ,Dr. N. K. Sakthivel,*"Development Of An Efficient Data Mining Classifier With Microarray Data Set For Gene Selection And Classification"* Journal Of Theoretical And Applied Information Technology, Vol. 35 No.2,January 2012.

[9] S. Neelamegam, Dr. E. Ramaraj, *"Classification Algorithm in Data Mining: An Overview",* International Journel of P2P Network Trends and Technology (IJPTT), Vol.3, Issue.5,Sep-Oct 2013.

[10] N. L. Ravi Teja1, V.K.R. Harish, D. Nayeem Muddin Khan ,R. Bhargava Krishna, Rajesh Singh, S Chaudhary*, "Land Slide Detection and Monitoring System using Wireless Sensor Networks (WSN*)", IEEE 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, India, 2014.

[11] Mohammad Zawad Ali, Md Nasmus Sakib Khan Shabbir, Xiaodong Liang, Yu Zhang, Ting Hu, *"Machine Learning based Fault Diagnosis for Single- and Multi-Faults in Induction Motors Using Measured Stator Currents and Vibration Signals",* IEEE Transaction, Vol.55, Issue.3, pp.2373-2391, 2019.

[12] Satishkumar Chavan, Shobha Pangotra, Sneha Nair, Vinayak More, Vineeth Nair*, "Effective and Efficient Landslide Detection System to Monitor Konkan Railway Tracks"* 2015 International Conference on Technologies for Sustainable Development (ICTSD-2015), 2015, Mumbai, India.

[13] Huang Qingqing, Meng Yu, Chen Jingbo, Yue Anzhi, Lin Lei, *"Landslide Change Detection Based On Spatiotemporal Context",* IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017.

[14] Siti Khairunniza-Bej, Siti Rusaniza Jusoh, *"Integrated change detection method for landslide monitoring",* International Conference on signal Acquisition and Processing, 2009.

[15] Thomas L., *"A Scheme to Eliminate Redundant Rebroadcast and Reduce Transmission Delay Using Binary Exponential Algorithm in Ad-Hoc Wireless Networks",* International Journal of Computer Sciences and Engineering, Vol.3, Issue.8, pp.1-6, 2017.

[16] Gagandeep Kau , Harmanpreet Kaur, "*Ensemble based J48 and random forest based C6H6 air pollution detection* ", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.2, pp.41-50, April (2018).

[17] Rohini M , Arsha P, *"Detection of Microaneurysm using Machine Learning Techniques",* International Journal of Scientific Research in Network Security and Communication, Volume-7, Issue-3, Jun 2019.

[18] Hemant Kumar Soni, "Machine Learning-"*A New Paradigm of AI",* International Journal of Scientific Research in Network Security and Communication, Volume-7, Issue-3, Jun 2019.

[19] Afzal Ahmad, Mohammad Asif, Shaikh Rohan Ali, *"Review Paper on Predicting Mood Disorder Risk Using Machine Learning",* International Journal of Scientific Research in Computer Science and Engineering, Vol.7, Issue.1, pp.16-22, February (2019).

Authors Profile

Ms.R.Sindhuja, is a Research scholar in the Department of Computer Science, Alagappa University, Karaikudi, Tamilnadu, India. She has received her M.Sc in Computer Science from Alagappa University, Karaikudi, Tamilnadu in the year of 2018. Her areas of research interest includes Data Mining, Machine Learning.

Mrs. A. Padmapriya, is working as Associate Professor in the Department of Computer Science, Alagappa University, Karaikudi – 630 003, Tamil Nadu, India. She has 15 years of teaching experience and 11 years research experience. She has published many papers in reputed journals and conferences. Her research areas include Data analytics, Communication Networks and Information security.

