Machine Learning Based Early Prediction of Rainfall Induced Landslide

V. Aarthi
Department of Electronics and Communication Engineering
National Engineering College, Kovilpatti, Tamil Nadu, India
Email- aarthivino13@gmail.com

V. Vijayarangan
Department of Electronics and Communication Engineering
National Engineering College, Kovilpatti, Tamil Nadu, India
Email- vijayvrangan@gmail.com

Abstract- The Landslide is a worse and most destructive natural disaster. In this world, many peoples lost their lives due to landslides. Hence, to overcome this problem, the early landslide prediction model is developed to save the lives of the peoples near hilly regions. Some other model uses machine learning technique to analyze rainfall data to predict these hazards. In the proposed system, Logistic Regression(LR) algorithm is used. Moreover the prediction performance of this approach is improved by various factors including landslide such as rainfall, slope, geology, geomorphology, distance from road, distance from stream, distance from lineaments and deforestation. These landslide cases are measured and are evaluated using the receiver operating characteristics, area under the curve (ROC-AUC) and false negative rate (FNR).

Keywords - Machine Learning Technique, Logistic Regression algorithm, Rainfall, Receiver Operating Characteristics, Area Under the curve.

I. INTRODUCTION

The landslide is one of the most dangerous natural disaster. In general, landslide is a movement of rock and debris or soil flow down to the earth's surface. It affects at least 15% of the land area of our country exceeding 0.49 million sq. km. Landslides are initiated by earthquake, heavy rainfall, flood and volcanic activities[4][7][10]. The factors considered for landslide triggering factors are rainfall, slope, longitude, latitude, geology, geomorphology, dam, road distance, stream power index, topographic wetness index, distance from streams, distance from lineaments and deforestation. Various studies on landslide occurring widely depicts that majority of landslides in India occurring at Himalayas and western ghats. In India number of landslides are due to heavy rainfall. The impacts of landslides are loss of life, destruction to infrastructure, loss of natural resources, damage the land and even some landslide materials block river and increase the risk of flooding[1].

Landslide incidence, being a ramification of spatial climatic changes, geological factors and precipitation is a highly uncertain phenomenon. Among all the factors inducing landslides, rainfall holds major importance[1].. Landslides cause high destruction, damage and disorder during the monsoon season, incurring human fatalities, considerable damage to surroundings and natural resources often shattering the livelihood of local people (Shikha Srivastava et al. 2020). And nowadays the most widely used methods for the landslide prediction related work is Machine Learning. The machine learning method will be a way of identifying patterns in data and by using this data for automatically making predictions or decisions in the future[5].

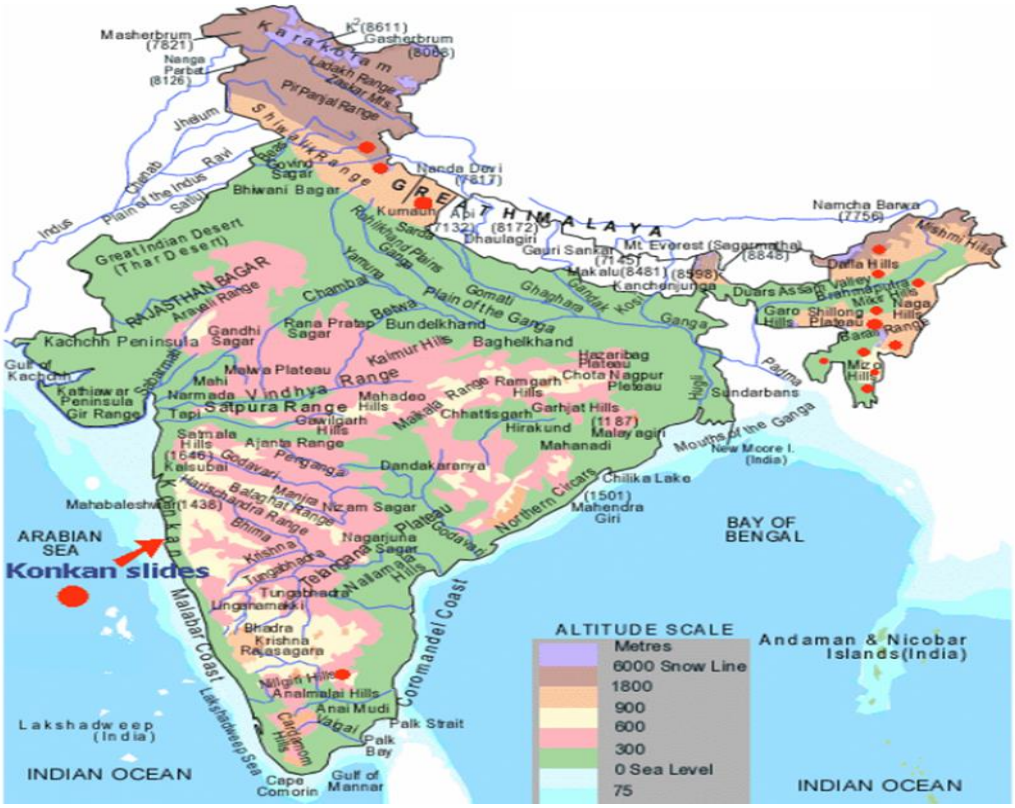


Figure 1. Landslide hazard zonation map in India

In the state of Kerala, a large landslide occurred in Idukki district because of heavy rainfall on **sixth** August 2020. Idukki district becomes the frequently occurring and worst-hit district in Kerala during 2018 and 2020 disaster, with 143 major landslides in the state government records[6]. The slope map of Idukki district shows the geography of the area consists of slopes as steep as 80◦ and the elevation ranges up to 2692 m. A great proportion of the populace of the district had homes in those risky slopes, which had been destroyed withinside **the** 2018 landslides regardless of the constructing typology. 97% of the essentia**l** roads withinside the districts reduce via the rugged mountains and hills, which might be regularl**y** blocked because of landslides withinside the monsoons. Sprawling **t**hroughout a place of 4358 km2 ,Idukki supplies 66% of the electric power requirements of Kerala. Idukki is the second largest district in terms of land area, where one half of the area is covered by forest with 3139 density in Kerala. The Western Ghats can be divided into two segments, north and south, separated by the Gap of Palghat.

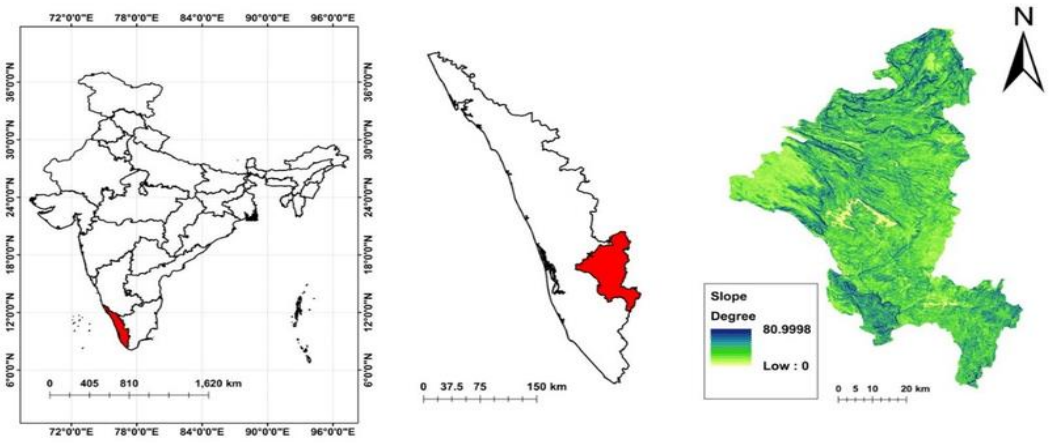


Figure 2. Location and slope map of Idukki district

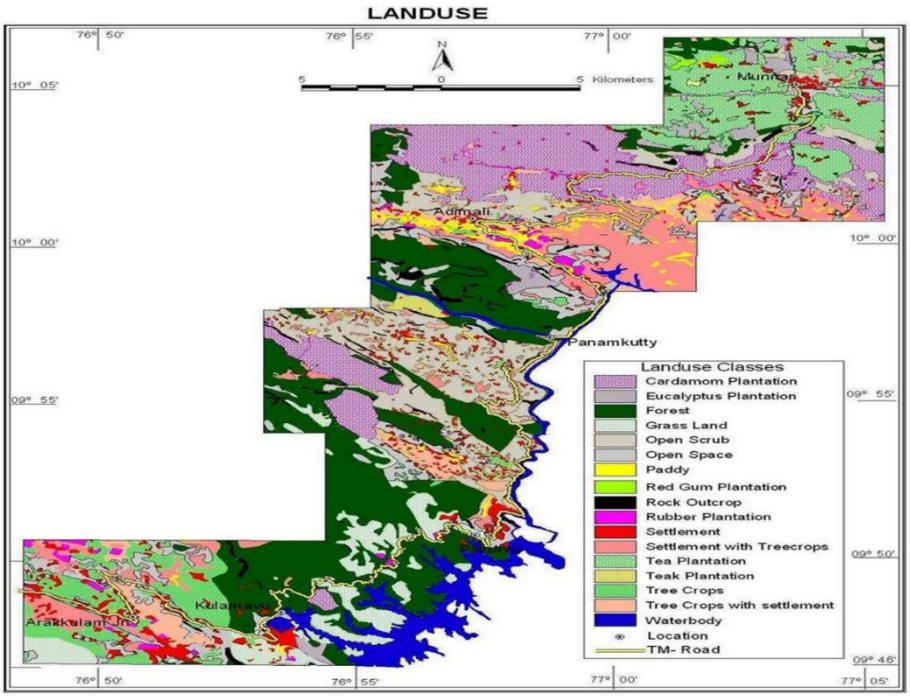
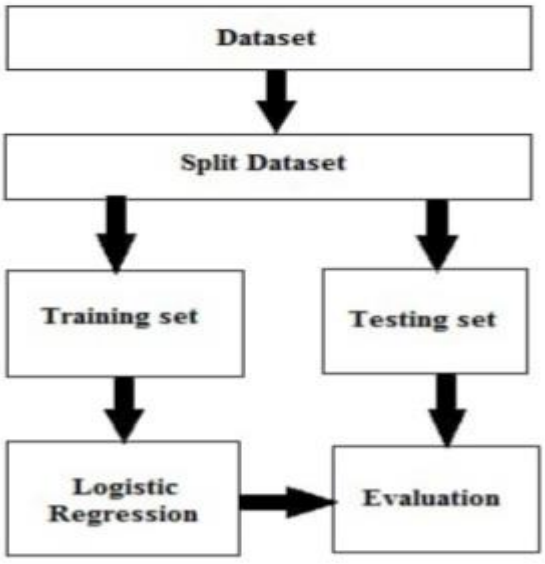


Figure 3. Density of forest in Idukki district

II. PROPOSED ALGORITHM

2.1 Data Collection

The first stage for creating landslide dataset is based on the historical records of any landslide hazard study. The landslide dataset developed for this research has been collected from the Geological Survey of India and from interactions with the people of the area. The deforestation diaries are collected from Kerala Forest Resource Management. The rainfall data of annual resolution collected from 2010 in Meteorological Centre.



.

Figure 4. Fliw chart of proposed system

2.2 Logistic Regression

There are many different algorithms and methods which could be used for the landslide prediction research. However, few methods perform better than others; no single method proves to be superior in all given conditions. So, just by selecting a model to get better performance using various factors triggering a landslide an input. The logistic regression model is one of the classification algorithm in machine learning technology, in which it determines the probability that the predicted output[9]. These predicted outputs belongs to a particular target variable or class (landslide and non-landslide). Logistic regression consists of three categories: Binary (dichotomous or has two possible classes coded as 0/1: absence/present), Multinomial (more than two categories, without ordering), and ordinal LR (more than 2 categories with ordering). The simplest form of sigmoid function for Logistic Regression is:

$$\mathrm{P(Y=1)} = \frac{1}{1 + \mathrm{e} - \mathrm{y}}$$

$$\mathrm{y} = \beta 0 + \beta 1 \mathrm{X1} + \beta 2 \mathrm{X2} + \cdots \beta \mathrm{nXn}$$

Where y is dependent variable and $X1, X2, \ldots and\ Xn$ are explanatory variables. Then:

$$\mathrm{P(Y=1)} = \frac{1}{1 + \mathrm{e}^{-(\beta 0 + \beta 1 \mathrm{X1} + \beta 2 \mathrm{X2} + \cdots \beta \mathrm{nXn})}} = \frac{\mathrm{e}^{\beta 0 + \beta 1 \mathrm{X1} + \beta 2 \mathrm{X2} + \cdots \beta \mathrm{nXn}}}{1 + \mathrm{e}^{\beta 0 + \beta 1 \mathrm{X1} + \beta 2 \mathrm{X2} + \cdots \beta \mathrm{nXn}}}$$

where β0 is an intercept of the model and β1, β2, ...βn are regression coefficients that measure the contribution of x1,x2, . . . xn to the prediction model. With LR mathematical equation can be generated from regression coefficients and intercept, hence the probability of landslide incidence can be computed.

III. Experiment and Result

In this study, the evaluation of explained variance of landslide triggering factors to landslide predictive model has been carried out using Logistic Regression algorithmIt is an efficient feature selection method in evaluating the variance of input data to models [2]. Explained variance is commonly used to measure the discrepancy between a model and actual data. Higher percentage of explained variance shows a strong strength of association. Hence, high explained variance shows the better predictive result.

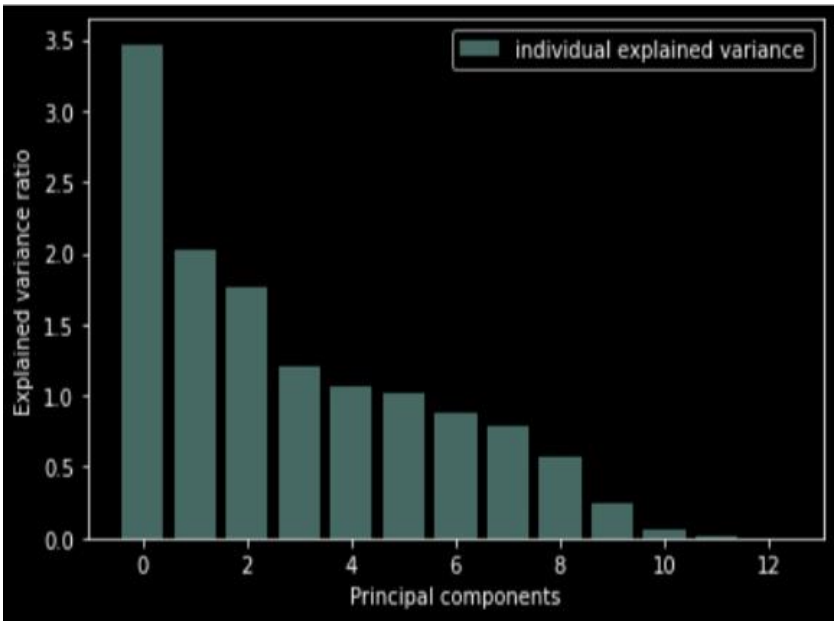
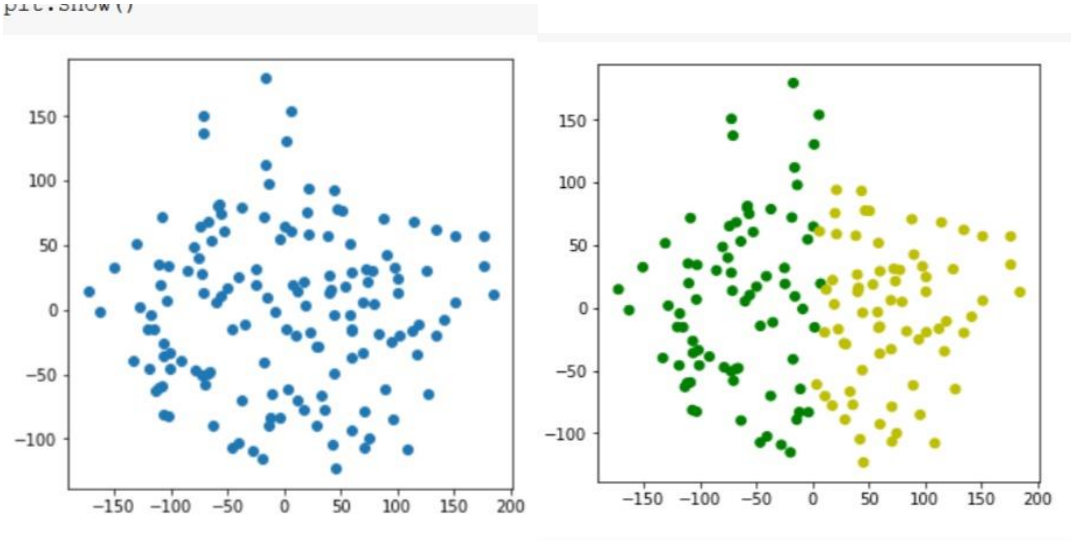


Figure 5. Explained variance of landslide triggering factors



(a) (b)

Figure 6. (a) Visualization of first factor using K-means clustering (b) Visualization of second factor using K-means clusterung

Receiver Operating Curve (ROC curve) is a standard method to evaluate general performance of landslide models. It is built by plotting couple of values ("sensitivity" and "100-specificity")[3]. Area under ROC curve (AUC) is used to evaluate quantitatively the performance of landslide models. If the AUC value is larger than 0.8, the performance of landslide models is good and acceptable. It can be observed that landslide model perform well for sp prediction of landslides in this study (AUC =0.99). The LR model (sensitivity = 5.77%, specificity = 4.46%) achives better performance

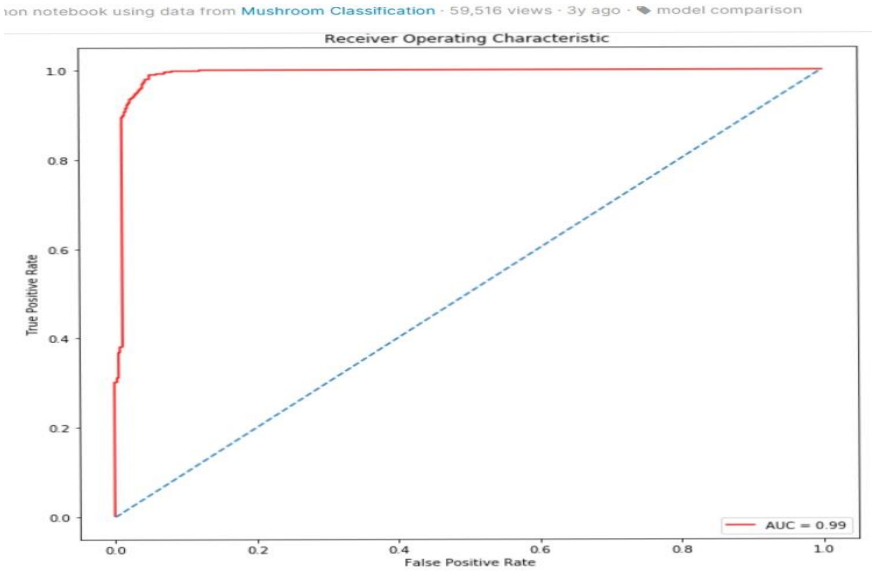


Figure 7. ROC-AUC curve of Logistic Regression model

IV.CONCLUSION

The natural hazards are beyond the control of human beings, but their destruction is often reduced if prediction mechanisms are carried out in advance. In this paper a detailed description of Logistic Regression algorithm for landslide prediction was discussed. The developed models are capable of predicting with better predictive result.The predictive accuracy of the models can be increased by introducing other input variable such as density of forest. The study is conducted specifically for Idukki district, Kerala but can be generalized to any area vulnerable to rainfall-induced landslides.

REFERENCES

[1] S. Renuga Devi, Pranay Agarwal. C. Venkatesh, P. Arulmozhivarman, "Daily Rainfall Forecasting using Artificial Neural Networks for Early Warning of Landslides", *International Conference on Advances in Computing, Communications and Informatics*, pp.2218-2224,2014.

[2] Binh Thai Pham1,2, DieuTien Bui3 , Indra Prakash4 and M.B. Dholakia1, "Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS", *Journal of Geomatics*, pp. 71-79, 2016.

[3] Mateja JEMEC AUFLIČ1 , Jasna ŠINIGOJ1 , Matija KRIVIC1 , Martin PODBOJ1 , Tina PETERNEL1 & Marko KOMAC2, "Landslide prediction system for rainfall induced landslides in Slovenia (Masprem)", *Geologij*a, pp.259-271,2016.

[4] Janko Logar, Goran Turk, Peter Marsden, Tomaž Ambrožič, "Prediction of rainfall induced landslide movements by artificial neural networks", *Hazards Earth Syst. Sci. Discussion*, pp.1-18, 2017.

[5] K. Uma, C. Ramesh Kumar, T. R. Saravanan, M. Basha Khaja, "Prediction of Rainfall Induced Landslides using Machine Learning Algorithms", *International Journal of Innovative Technology and Exploring Engineering*, pp. 4274-4278, 2019.

[6] Minu Treesa Abraham, Deekshith Pothuraju and Neelima Satyam, "Rainfall Thresholds for Prediction of Landslides in Idukki, India: An Empirical Approach", *MDPI Journal*, pp.1-16, 2019.

[7] Jaishankar Bhatt, Akash Gangwar, Rahul Nijhawan, Durgaprasad Gangodkar, "A Research on Deep Learning Advance for Landslide Classification using Convolutional Neural Networks", *International Journal of Innovative Technology and Exploring Engineering*, pp.903-906,2019.

[8] F. S. Tehrani, G. Santinelli, M. Herrera, "A framework for predicting rainfall-induced landslides using machine learning methods", *Proceedings of the XVII ECSMGE*, pp.1-8,2019.

[9] Martin Kuradusenge, , Santhi Kumaran and Marco Zennaro, "Rainfall-Induced Landslide Prediction Using Machine Learning Models: The Case of Ngororero District, Rwanda*". International Journal of Environmental Research and Public Health*, pp. 1-20, 2020.

[10] Shikha Srivastava, Nishchay Anand, Sumit Sharma, Sunil Dhar, Lokesh Kumar Sinha, "Monthly Rainfall Prediction Using Various Machine Learning Algorithms for Early Warning of Landslide Occurrence", *International Conference for Emerging Technology*, pp.1-7, 2020.