LANDSLIDE DETECTION AND PREDICTION USING LUBE AND PSOGSA

Archana.S. [1], Abijith B. [2], Anjana G. Nair. [3], Shahanas A. T. [4], Hema P. Menon[5].
Department of Computer Science and Engineering
Sreepathy Institute of Management AndTechnology,
Vavanoor, Kerala, India
archanaarchi2000@gmail.com[1], abhijithbalakrishnan12@gmail.com[2],anjanagswift@gmail.com[3],
shahanasshanu1711@gmail.com[4],hemapmenon@simat.ac.in[5]

Abstract—Natural calamities are the unforeseen events that costs numerous lives and property. It is a well-known fact that occurrences of landslides poses major social and economic threats to areas where geophysical conditions are favorable to such landslide movements. Landslides are universal phenomena, but more than being 'natural hazards', they are induced by human activity. Hence, development of a robust model / system that can be used for continuous monitoring and prediction of landslide is gaining a lot of prominence in the recent days. This becomes very crucial in implementing remedial measures like disaster risk reduction. The main goal of this work is to help in saving lives and properties, in landslide prone areas, by providing timely intimation of the probability of occurrence of a landslide to the citizens and authorities in and around the area. Prediction of the probability of occurrence of landslide is developed using various external and internal parameters that causes the land slide with the help of AI/Machine Learning systems. One aim of this work is to construct the logistic regression (LR) based Prediction intervals (PIs). For the construction of PIs a lower upper bound estimation (LUBE) method has been used. Combining particle swarm optimization (PSO) and gravitational search algorithm (GSA), a hybrid evolutionary algorithm, PSOGSA, is utilized to optimize the parameters. An overall accuracy of 80% has been obtained.

Keywords—Lower Upper Bound Estimation (LUBE), Particle Swarm Optimization (PSO), Gravitational Search Algorithm (GSA)

I. INTRODUCTION

Landslide is a frequently occurring natural hazard. Landslide identification plays an important role in landslide risk assessment and management. An early prediction of landslide occurrence can save lives of people. A landslide is the down slope movement of soil,rock, and organic materials under the effects of gravity and also the land form that results from such movement. Landslides are also known as landslips, slumps or slope failure. Landslides occurs when there is a disturbance over the upper layer of the soil. Nowadays, landslides are becoming the most scariest thing happening during monsoon. The highlands of Kerala experience several types of landslides, of which debris flows are the most common. This year, due to heavy rainfall in the monsoon, severe floods affected Kerala, occurring landslides in some places of Kerala. It resulted in the death of so many people. For the past two years, 145 people died due to landslides in Kerala. This indicates that major landslides would continue to be annual affairs. Thus, our project focuses on the detection of conditions that cause landslides and compute it and predict the occurrence of landslides to minimize the impact imposed by landslide on human, damage to property and loss of lives.

II. LITERATURESURVEY

Landslide occurrences are caused by long periods of environmental and human interaction that cause slippage of top soil. Factors such as the amount of rain, soil type, altitude from sea level and slope of the plain can be some of the reasons to give in to the contribution of the disaster. Data concerning to landslide can be easily over fittedor under fitted, provided the condition relating to them and their implementation go hand in hand. Throughout the papers referred, an understanding of various models and their accuracies was obtained. Several approaches to the problem was taken in such as statistical or auto regressive methods [1]. However, a lot of them were more of a deterministic point prediction from the loads of data obtained. The main disadvantage of such case is the obtaining of prediction error but such clause for prediction correction[2].

The lower upper bound estimation (LUBE) method [5] constructs a model with two outputs for the estimation of the prediction interval bounds. Prediction intervals(PIs) is a well-known tool for quantifying and representing the uncertainty of predictions [3]. PI deals with the uncertainty in the prediction of a future realization of a random variable [4]. The training process to obtain optimal prediction intervals is rather slow, however optimization of the parameters increases the speed. There are various optimization models such as sample annealing, particle swarm optimization that can be used to here [6]. A heuristic optimization algorithm such as a hybrid PSOGSA algorithm can have a significant impact on the results obtained[8].

III. PROPOSEDSYSTEM

A. System Architecture

Components of Landslide Prediction system

- Dataset Construction
- Splitting of data
- Prediction Intervals
- Prediction
- SMS Intimation

In this system, the pre-recorded landslide occurrence details as well as some random non-occurrence landslide details are taken as input and is used for training the model. For this, details of different attributes such as rainfall, humidity, altitude are used. Now, the model is constructed to which different algorithms are applied. This output is used for predicting whether a landslide occurs or not. If landslide occurs then an intimation is sent in the form of an SMS. Fig. 1 shows the working of the whole Landslide Detection and Prediction System.

The system architecture comprises of different components. At first, a NASA dataset is taken from Kaggle website which had landslide details all around the world. From this landslide details of a specific state named ”Kerala” is extracted. In order to get details of landslide non-occurrence, some such data is added randomly. This forms the required dataset. Next, values of all the input attributes of the dataset is converted to integer data type if it is of any other data type. Then, dataset is split into two such as training set and testing set. The model for the developing system is constructed using logistic regression [9] method. The two algorithms used are LUBE (Lower Upper Bound Estimation) and PSOGSA (Particle Swarm Optimisation and Gravitational Search Algorithm) [7]. Prediction intervals are used for prediction [11]. This is used because of it’s efficiency, as it gives a range of possible values. Gradient boosting regressor[10] is the method used for construction of prediction intervals. This generates lower bound and upper bound values corresponding to each input attribute. Later these values are used for predicting whether a landslide occurs or not for the particular input given. If the input given corresponding to all the attributes lies between these lower bound and upper bound values then there is a possibility for landslide occurrence. If there is a possibility for landslide occurrence, then a SMS alert is send to the respective registered authority. If there is no possibility for landslide occurrence, then a message indicating ”safe spot” is just displayed.

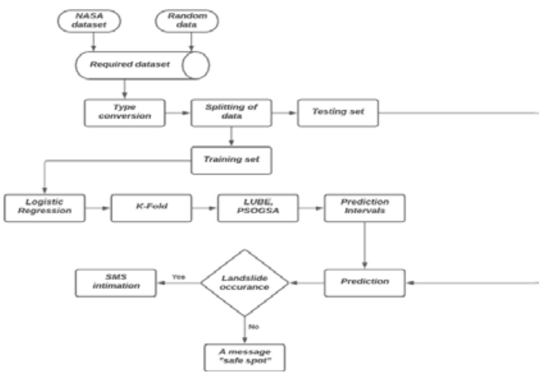

Fig. 1. System Architecture

B. Proposed System Outline

Proposed system consists of four phases. First one is related with dataset and it's pre-processing stages. Dataset description and data pre-processing are the two topics discussed under this.

Second phase is the training, validation and model selection phase. Logistic Regression is used for training and creating the model. K-fold cross validation method [13] is used for validating the model.

The third phase is algorithm implementation phase where two algorithms are used such as LUBE and PSOGSA [14]. The output of this algorithm will be prediction intervals which will be used for prediction. Gradient boosting regressor is used for constructing the prediction intervals. This is the first task done in this phase. The second task done is prediction for which logistic regression is used. It gives two outputs such as landslide occurred and landslide not occurred.

The next and the last phase is the development of message passing system. SMS is the message passing medium used which runs on an android application.

C. Data Pre-Processing

This module deals with dataset and it'spre-processing stages. The data set in this project is previous historical data of landslide occurrences in Kerala. The dataset contains both landslide occurred data and not occurred data.

1) Dataset Description: The historical data about landslide incident's records were collected from the Kaggle website which contains the data from all over the world. The internal factors mostly contributing to the landslide occurrence are slope angle, soil type texture and soil depth while rainfall is the external triggering factor. The rainfall data was collected from the India Meteorological Department(IMD).

2) Data Pre-Processing: Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. A real-world data generally contains noises, missing values, and would be in an unusable format which cannot be directly used for machine learning models. Data pre-processing is a required task for cleaning the data and making it suitable for a machine learning model, which also increases the accuracy and efficiency of the model. The data collected here is geographical details of locations.These environmental factors include latitude, longitude, river distance, soiltype, humidity, altitude and so on. The data which have not been properly screened will cause misleading results. To acquire quality result which will be helpful for information generation and decision making, this raw data need to be pre-processed. After pre- processing the data set is split into two sets which is training and testing set. The splitting of data is done in the ratio of 3:1, that is 70% of the data is used as training set and 30% of the data is used as testing set.

D. Training and model selection

This module deals with the training , validation and model selection.

1) Training Phase: Training Phase is the learning phase, the process of training an ML model involves providing an algorithm (that is, the learning algorithm) with training data to learn from. The term model refers to the model artifact that is created by the training process. The training data must contain the correct answer, which is known as a target or target attribute. The learning algorithm finds patterns in the training data that maps the input data attributes to the target, and it outputs a model that captures these patterns. Different machine learning models exist and the choice of which one to use generally depends on the problem at hand.

In this project, LR is used for training and creating a model. The train set is taken to create the model. Logistic regression is a binary classification method. It can be modelled as a function that can take in any number of inputs and constrain the output between 0 and 1. After training, the model is validated and which is done in the next stage.

2) Validation and Model Selection: Validation is a process that estimates the quality of the model. This is used to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform

in general when used to make predictions on data not used during the training of the model.

In this project K-fold cross validation method is used for validating the model. In this, the method splits the dataset into K consecutive folds, each fold is the nusedonce for validation while the K-1 remaining folds form the training set. The result of the cross validation is selected as the best set of parameter values which is called model selection.

E. Prediction

1) Prediction Interval Construction: A prediction interval is an estimate of an interval in which a future observation will fall, with a certain probability [12]. The PI's are constructed using the Lower Upper bound estimation method (LUBE). In this phase, the lower upper bound so feach attribute is calculated. Togenerate PI's, Gradient boosting regressor function is used which is an inbuilt function in sci-kit learn.

Gradient boosting regressor method relies on the intuition that the next best possible model, when combined with previous models minimizes the overall prediction error. The key idea is to set the target outcomes for the next model in-order to minimize the error.

1) For the lower prediction, use Gradient Boosting Regressor(loss= ”quantile”, alpha=lower quantile) with lower quantiler epresenting the lower bound, say 0.1 for the 10^{th} percentile

2) For the upper prediction, use the Gradientn Boosting Regressor(loss= ”quantile”, alpha=upper quantile) with up- per quantile representing the upper bound, say 0.9 for the 90th percentile

The lower quantile and upper quantile values is set as 0.15 and 0.85 respectively. After splitting the data into train and test sets, the model is built. Two separate Gradient Boosting Regressors is used because each model is optimizing a different function and must be trained separately. For the optimization of the output, PSOGSA algorithm is used. The lower and upper bounds of each attribute is calculated. During the prediction the test values of each attribute exceeding the bound values will be considered as negative result. Only the values within the bound gives the positive result.

2) Prediction: In this module, the prediction of data given as input on either there is probability of landslide occurrence or not is done. The prediction is done by LR algorithm. Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature, having the data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Here, the two classes are landslide occurrence and non-occurrence of landslide. 1 is assigned as the landslide occurrence class and 0 is assigned as the landslide not occurred class. After inputting a new data if the system gives an output of 1 that indicates there is a probability of landslide in the given place. If the system returns a 0, then it means there is no probability of landslide in that place.

F. Intimation model

1) Message Passing System: For the purpose of early warning, an intimation model is been developed which is an SMS system that runs with the help of an android application. If the prediction result is ”possibility for landslide” then an intimation is sent to the corresponding authority indicating ”possibility for landslide along with the most probable at eit would take place”.

IV. RESULT

A. Lube Result

After the accumulation of data, normalization of the data sets is done using Lube method where an upper and lower values of interval is formed with the help of the parameters associated. Prediction intervals so made is used to predict he even to ccurrence. The value test ed gives above 80% accuracy, respectively, which indicates tha tthe proposed method can obtain a satisfactory fitting to the data set given here. The interval formed by LUBE is given in Table 1 and Table 2

B. Web-application

To show the prediction simulation a web app is created to which the attributes are added manually. The form to that can be seen in the Fig. 2. The attributes entered are given into the trained network to obtain a particular event occurrence. If the event occurrence corresponds to landslide taking place then the web-app immediately sends a text message to the user about the landslide, else a notification is seen which ensures no landslide takes place. Fig. 3 shows the display message indicating ”possibility for landslide”. This is the output of prediction or can be called as prediction result. Fig. 4 shows the GUI of the android application. Fig. 5 shows the SMS alert sent to authority indicating possibility of landslide.

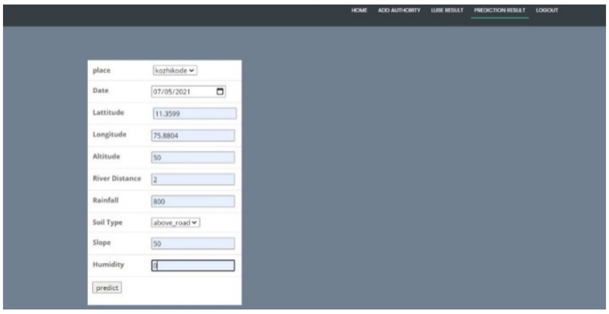

Fig. 2. Form for Prediction Result

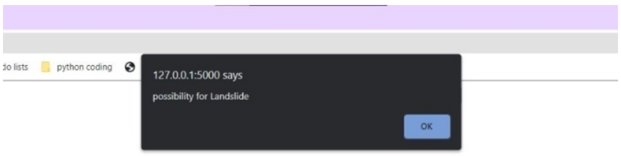

Fig. 3. Prediction Message

C. Comparison Analysis

For the purpose of comparison, LDA, KNN, NB, SVM and Logistic Regression is used on the trained model. The comparison can be seen in the chart in Fig. 6. The results proved that the method proposed by this paper gives more accurate results than all the other methods used to test here. While LDA and NB algorithm gives around 55% accuracy and the other two provides around 48-50% accuracy, the Logistic Regression algorithm provides more than 80% accuracy making it the most suitable option to use on the trained network.

V. CONCLUSION AND FUTUREWORK

Landslide is a subject with major relevance that needs attention. The most viable method to safeguard the economy and the ecosystem from the natural catastrophe is by the identification and prediction of landslides on time. We find ourselves in great peril in midst such a crisis, so the intimation of the predicted occurrence of landslide via a web-app and an app giving an alert message on prediction is discussed here while trying not to compromise with the accuracy in prediction. For normalizing the data set and constructing prediction intervals Lower-Upper Bound method (LUBE) is used along with PSOGSA and logistic regression model for optimization and training. For the classification of prediction values at the intimation stage k-fold is used for evaluating the model and select the best possible classifier. With this simulation, given the perfect training an accuracy of about more than 80% has been obtained and which is comparatively higher than other machine learning models. So, for future work with addition of real- time sensors and training with more wider range of values a better model with an increased accuracy can be developed.

Table I: Lower Bound

DateLB	LatitudeLB	LongitudeLB	AltitudeLB	River distanceLB	RainfallLB	Soil typeLB	SlopeLB	HumidityLB
5.999	11.437	74.988	18.108	103.567	606.524	1.999	55.975	82.744
5.999	11.414	75.124	70.714	230.020	601.649	1.999	55.975	82.745
5.999	11.415	74.885	10.518	230.021	695.167	1.999	55.975	82.612
5.999	11.415	74.905	8.403	149.999	697.645	1.0	21.999	82.732
5.999	11.415	75.152	10.779	228.238	599.222	1.999	49.998	82.745

Table II: Upper Bound

DateUB	LatitudeUB	LongitudeUB	AltitudeUB	River distanceUB	RainfallUB	Soil typeUB	SlopeUB	HumidityUB
8.0	12.940	76.091	23.017	134.263	892.414	2.0	97.660	86.280
8.0	13.079	76.091	69.031	243.200	775.444	2.0	98.955	87.044
8.0	12.922	76.091	20.609	361.509	954.693	4.0	99.136	93.999
8.0	13.106	76.091	28.382	150.017	1009.318	2.0	23.029	89.070
8.0	12.396	76.091	23.017	227.999	719.936	2.0	49.999	89.078

Fig. 4. SMS app

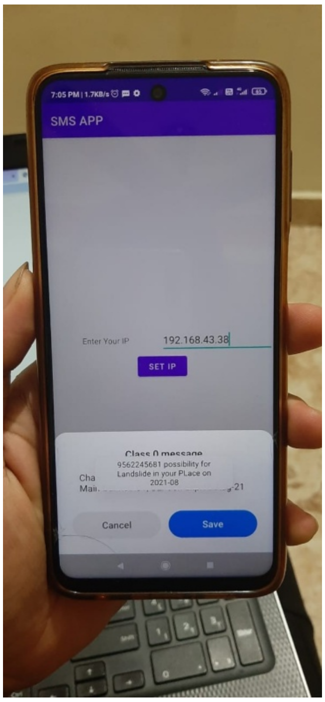

Fig.5.SMSalertsentfromApptoauthority

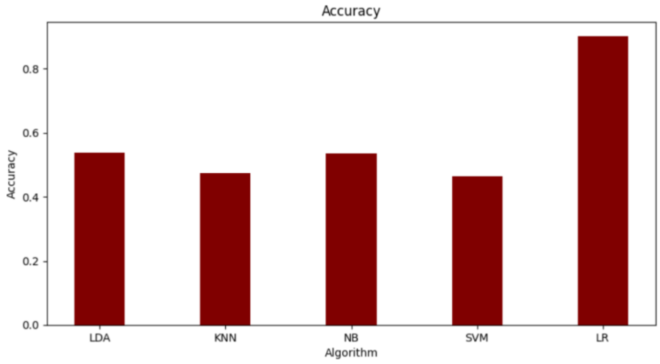

Fig. 6. Comparisons among differentmodels

REFERENCES

[1]Cheng Lian, ZhigangZeng, Senior Member, IEEE, Wei Yao, Huiming Tang, and Chun Lung Philip Chen, Fellow, IEEE ”Landslide Dis- placement Prediction With Uncertainty Based on Neural Networks With Random Hidden Weights”,IEEE Transactions on Neural Networks and Learning Systems,December 20, 2015

[2]N. Meade and T. Islam, “Prediction intervals for growth curve forecasts,” J. Forecast., vol. 14, no. 5, pp. 413–430, Sep. 1995.

[3]T. Heskes, “Practical confidence and prediction intervals,” in Advances inNeuralInformationProcessingSystems,vol.9,T.P.M.Mozerand

[4]M. Jordan, Eds. Cambridge, MA: MIT Press, 1997, pp. 176–182. [4]A. Khosravi, S. Nahavandi, and D. Creighton, “A prediction interval based approach to determine optimal structures of neural network metamodels,” Expert Syst. Appl., vol. 37, no. 3, pp. 2377–2387, Mar. 2010

[5]A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya,“Lower upper bound estimation method for construction of neural network-based pre- diction intervals”, IEEE Trans. Neural Netw., vol. 22, no. 3,pp. 337–346, Mar. 2011.

[6]J. Kennedy and RC. Eberhart, ”Particle swarm optimization,” in Proceed- ings of IEEE international conference on neural networks, vol. 4, 1995, pp. 1942–1948.

[7]E. Rashedi, S. Nezamabadi, and S. Saryazdi, ”GSA: A Gravitational Search Algorithm,” Information Sciences, vol.

179, no. 13, pp.22322248, 2009.

[8]S. Mirjalili and S. Z. M. Hashim, "A new hybrid PSOGSA algorithm for function optimization," in Proc. Int. Conf. Comput. Inf. Appl., Tianjin, China, Dec. 2010, pp. 374–377.

[9] ZHU Lei, HUANG Jing-feng," GIS-based logistic regression method for landslide susceptibility mapping in regional scale" Journal of Zhejiang University SCIENCE A,2006

[10] AnkushPathania , Praveen Kumar, Priyanka, AakashMaurya, K. V. Uday and VarunDutt," Development of an Ensemble Gradient Boosting AlgorithmforGeneratingAlertsaboutimpending Soil Movements"ResearchGate August 2020

[11]C. Chatfield, "Prediction intervals," Journal of Business and Economic Statistics, vol. 11, pp. 121–135, 1993.

[12]A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," IEEE Transactions on Power Systems, vol. 25, no. 3, pp. 1496 –1503, Aug. 2010.

[13]Ping Jiang, ZhigangZeng, Jiejie Chen and Tingwen Huang, "Gener- alized Regression Neural Networks with K-Fold Cross-Validation for Displacement of Landslide Forecasting" Conference: International Sym- posium on Neural Networks November 2014.

[14]R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in 1995. MHS '95., Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Oct. 1995, pp. 39–43.