A Study of Landslide Classification Using Machine Learning Approaches

PRIYANGA P[1], Dr.B.ROSILINE JEETHA[2]

[1]PG Student, Department of Computer Science, Dr.N.G.P.Arts and Science College, Tamil Nadu, India.
[2]Professor and Head, Department of Computer Science, Dr.N.G.P.Arts and Science College, Tamil Nadu, India.

Abstract
Landslides can merely be tragic to human life and property. Increase among the speed of human settlement among the mountains has resulted in issues with safety. Landslides have caused economic loss between 1-2% of the value in many developing countries. throughout this study, we've got AN inclination to dialogue a deep learning approach to sight landslides. Convolutional Neural Networks unit used for feature extraction for our planned model. As there was no offer of a specific and precise data set for feature extraction, therefore, a replacement data set was designed for testing the model. we've tested and compared this work with our planned model and with completely wholly completely different machine-learning algorithms like activity Regression, Random Forest, AdaBoost, K-Nearest Neighbors and Support Vector Machine. Our planned deep learning model produces a classification accuracy of ninety six.90% outperforming the classical machine-learning algorithms.
Keywords: Convolutional Neural Network, Deep Learning, Relu, Pooling.

I. INTRODUCTION

A landslide is nothing but a motion of trash, large rocks or parcel down a slope due to gravity [1]. Landslides happen in Associate in Nursing extremely quite surroundings drawn by either delicate or steep slope inclinations ranging from mountains to coastal reefs and to boot even underwater. The factors like degree earthquake, a crucial downfall, floods, construction work on the slope that triggers the landslides by poignant the soundness of the slope. varied factors just like the higher rate of unplanned urbanization, deforestation, and surprising downfall resulting in the increase of landslide problems that seems to be more durable among the longer term. On July 30, 2014, 151 people died and plenty of were missing among the Malin village in region due to landslide [2]. In July 2013 over 5700 people were dead due to landslides caused by Kedarnath floods in Uttarakhand [3]. among the longer term, the number of landslides will increase due to the results of excessive downfall and changing climatic conditions. throughout this fashion, the landslide has clad to be one altogether the foremost well liked subjects for researchers. The precise estimation of the landslide continues to be a troublesome real-world draw back. Landslides ar hazards for human life and economy of a country, therefore, to scale back the damages and dangers to the people and their property many efforts area unit created among the past to grasp the abstraction and temporal distribution of landslides and to know the thanks to management them [4].

Figure1: Photographs of landslides

Landslide classification may well be a brand new approach for detection landslides. Landslide detection study will facilitate in detection the landslides and facilitate early warning signs so as that immediate

safety actions got to be taken. In recent times, deep learning has brought advances in method footage, video, audio, and speech. It permits method models that comprise processing layers to be told with illustration of knowledge with multi-level abstraction. Deep learning models notice patterns in info sets by utilizing backpropagation to specify but a machine got to alter its weights that ar wont to technique the illustration of knowledge in each layer from the past layer [5]. This paper discusses the deep learning framework for landslide classification which may be helpful in classifying and detection landslide areas. we've planned the Convolution Neural Network (CNN) model approach for landslide classification. we've got an inclination to discovered that the CNN approach outperformed the traditional machine learning methodologies like KNN, RF, SVM, provision Regression and AdaBoost. The inconvenience of the knowledge set was a challenge faced by our model. we have a tendency to wished effective and precise resources of feature extraction therefore we've got an inclination to created our own high-resolution image info set of landslide-prone areas.

II. LITERATURE REVIEW
In the past years, many researchers have shown their vast interest among the area of landslide. Landslide classification exploitation the deep neural network may well be a recently pursued subject throughout that landslide is detected among {the footage|the photographs|the images} which may be helpful in characteristic landslide regions in maps and via pictures of internet sites at risk of landslides. Truong et al. [6] have used ensemble technique that might be a fusion of cloth Ensemble technique that might be a fusion of cloth Ensemble and provision Model Tree for refinement the performance of landslide condition model. OH and Lee [7] planned processing approaches like Artificial Neural Network and Boosted Tree for landslide condition modeling giving the validation results of eighty 2.5% and 90.79% severally. Ramesh [8] discusses the design and activity of a landslide sightion system using a Wireless detector Network system to observe landslides. To sight landslides of five differing kinds Martha, Kerle, Western, Hetten, degreed Kumar [9] presents associate degree intelligent classification approach with the thresholds of diagnostic parameters derived from if-means cluster Associate in Nursingalysis with an overall check accuracy of seventy seven.7%.

III. FUNCTIONAL REQUIREMENTS
3.1. Data collection
The information assortment technique involves the selection of quality knowledge for analysis. Here we've got an inclination to used dataset with eleven attributes for learning. the work of a knowledge analyst is to look out ways that within which and sources of aggregation relevant and comprehensive knowledge, decipherment it, and analyzing results with the help of mathematics techniques.

3.2. Data visualization
A large amount of knowledge delineate in graphic kind is a smaller amount difficult to know and analyze. Some companies specify that a information analyst ought to acumen to create slides, diagrams, charts, and templates. In our approach, the detection rates of landslide is shown as info mental image FIG[].
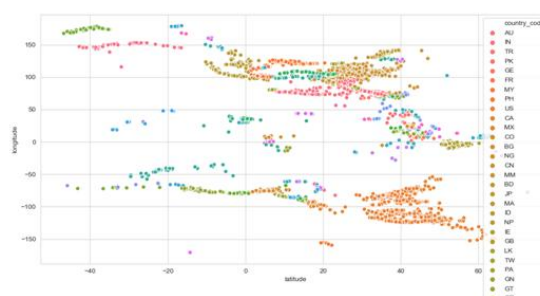

Figure2: Data Visualization of Countrywise Landslide

3.3. Data preprocessing
The purpose of preprocessing is to convert knowledge into a sort that matches machine learning. Structured associate degreed clean info permits a information soul to urge further precise results from Associate in Nursing applied machine learning model. The technique includes info, cleaning, and sampling.

3.4. Dataset splitting
A dataset used for machine learning got to be divided into three subsets — employment, test, and validation sets.Training set. a knowledge soul uses a training set to educate a model and description its best parameters it's to be told from knowledge.Test set. A check set is needed for degree analysis of the trained model and its capability for generalization. The latter means a model's ability to identify patterns in new unseen

info once having been trained over a training info. It's crucial to use utterly completely different subsets for employment and testing to avoid model overfitting, that's that the incapacity for generalization we've got an inclination to mentioned on prime of.

IV. NON-FUNCTIONAL REQUIREMENTS

The following may well be a listing of non-functional wants. the precise details will got to be printed by internal stakeholders.

- Interval
- Accessibility
- Stability
- Maintainability
- Usability

V. METHODOLOGY

we have a tendency to used deep learning framework i.e. CNN to extract vital options required for detection the landslide in our categories of knowledge set i.e. landslide and non-landslide. we have a tendency to enforced our model on a machine consisting of eighth generation Intel i5 processor with RAM of size 8GB, and a GPU (4GB Nvidia 940Mx). The deep convolutional neural network processes the multi-array knowledge, like image knowledge, audio, and video knowledge. CNN constitutes primarily 3 sorts of layers. the primary layer is that the convolutional layer within which the parameters comprise of plenty of kernels that may learn. The convolutional layer uses activation perform, like ReLu [10] i.e. rectilineal Unit perform that produces nonlinearity in our model. The Second layer is pooling layer that diminishes the framework's computation by spatial property reduction and moreover controls overfitting. The used pooling is max-pooling that takes the best most price in every block. At last, the complicated reasoning within the network is finished within the third layer by taking all the somatic cells within the preceding layer and links it to every neuron it's.

We compared and contrasted our knowledge attack this state of the art algorithms (Logistic Regression, RF, AdaBoost, KNN, SVM) and projected deep learning design on the premise of their performance.

VI. APPROACHES

A convolutional neural specification for the matter of landslide classification is given during this paper. Our projected design consists of seven layers that ar related to weights, out of that there ar two convolutional layers, two pooling layers, one flattening and a pair of totally connected layers. within the Convolution stage, thirty two activation maps ar generated victimisation 3x3 feature detector with input image size of 64x64 pixels and stride of 1-pixel. Convolutional layer uses the activation perform referred to as ReLu to extend the non-linearity in landslide pictures In Maxpooling, the pool size of 2x2 is taken for building the pooled feature map. once these 2 steps, we have a tendency to flatten our pooled feature map into a column-like long feature vector of associate input image. it's gone through 2 totally connected layers that ar fashioned from the multiple dense networks of neurons to predict whether or not the image is of a landslide or not. First, totally connected layer uses ReLu and alternative uses sigmoid activation functions to extend non-linearity. At the dense layer the feature vector is increased by the layer's weights, summed with biases associated gone through an activation perform for higher results. Adam optimisation rule [11] is employed to update network weights unvarying supported coaching knowledge. This CNN design is then compiled to observe the photographs of sophistication landslides or nonlandslide.

VII. MACHINE LEARNING MODELS

This study aims to use a combine of machine learning techniques and select the foremost effective acting one for predicting landslide incidences in land, which might be used for early warning to cut back risks. a combine of classification models that area unit selected to accomplish this analysis space area unit random forest (RF) and provision regression (LR). The simplicity or quality of every MLT differs from one to a definite. Hyper-parameter activity has been done to optimize the performance of every. These classification models were chosen as a results of this analysis is coping with the binary classification balk.
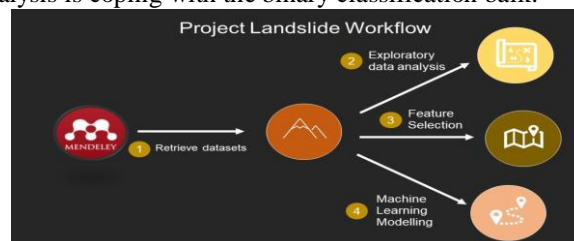

Figure3: Proposed landslide prediction flow chart.

VIII.RANDOM FOREST

Random forest (RF) was initial introduced by Breiman [14] as Associate in Nursing ensemble learning formula. it's a ordinarily used machine learning technique for info prediction for each classification and regression issues [14–15]. it's noted as random forest thanks to randomness at intervals the choice of selections on every call tree and along coaching samples from the dataset [16]. It operates by increase many call trees from random subsets of dataset throughout the coaching time to form up a relationship between selections and labels [17]. call or classification tree (Figure 7) might even be a arboreal diagram which will be a basic structure block of RF accustomed decide a route of action. RF consists of the event of classification rules (tree) from bootstrap samples of the dataset. the choice of the last word classification is computed from the bulk vote of multiple classification trees. Int. J. Environ. Res. Public Health 2020, 17, x nine of nineteen a mix of.3.1. Random Forest Random forest (RF) was initial introduced by Breiman [14] as Associate in Nursing ensemble learning formula. it's a ordinarily used machine learning technique for info prediction for each classification and regression issues [14–15]. it's noted as random forest thanks to randomness at intervals the choice of selections on every call tree and along coaching samples from the dataset [16]. It operates by increase many call trees from random subsets of dataset throughout the coaching time to form up a relationship between selections and labels [17]. call or classification tree (Figure 7) might even be a arboreal diagram which will be a basic structure block of RF accustomed decide a route of action. RF consists of the event of classification rules (tree) from bootstrap samples of the dataset. the choice of the last word classification is computed from the bulk vote of multiple classification trees.
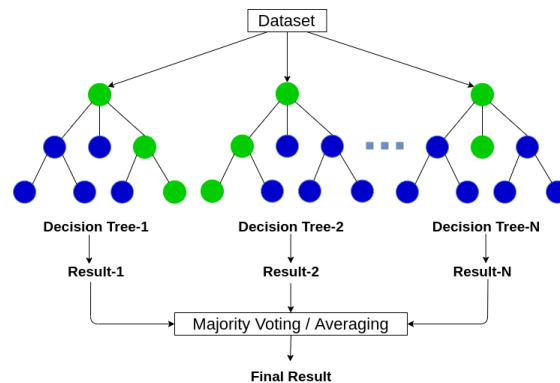

Figure4: Random Forest decision tree

The bigger kind of trees at intervals the model plays a task in up the accuracy and prophetical capability [18] and needs methodology the number of trees (T) and in addition the number of prophetical variables (m). RF will work with each categorical and numerical info. RF was chosen attributable to its power to the traumatize mixed variables (categorical and numerical), over-fitting risks reduced, efficiently functioning on Brobdingnagian databases with high accuracy, hardiness with less coaching time, estimates missing info, and have importance indication[19].

IX. CONCLUSION

In our paper, we've an inclination to used classical machine learning algorithms and a deep learning model to ascertain their performance on our information set. we've an inclination to tested associate degreed compared their performance results ANd concluded that deep convolutional neural networks approach to note landslides succeeded in getting an accuracy of ninety six.90%. The planned model is evaluated on various parameters like mythical being curve (AUC=99.20%), Sensitivity (96.90%), accuracy (96.90%) and F1-Score (96.90%). the aim of this study would possibly embody early detection of the landslides thus rescue operations and safety measures could also be taken quickly to scale back the loss of life in landslide-hit areas. This work can facilitate in making associate early decision and land use developing with in areas vulnerable to landslides.

REFERENCE

[1]. Cruden, D.M., 1991. A simple definition of a landslide. Bulletin of Engineering Geology and the Environment, 43(1), pp.27-29.

[2]. Parkash, S., 2011. Historical records of socio-economically significant landslides in India. J South Asia Disaster Studies, 4(2), pp.177-204.

[3]. Jiménez-Perálvarez, J.D., El Hamdouni, R., Palenzuela, J.A., Irigaray, C. and Chacón, J., 2017. Landslide-hazard mapping through multi-technique activity assessment: an example from the Betic Cordillera (southern Spain). Landslides, 14(6), pp.1975-1991.

[4]. Varnes, D.J., 1984. Landslide hazard zonation: a review of principles and practice (No. 3).

[5]. Nijhawan, R., Verma, R., Bhushan, S., Dua, R. and Mittal, A., 2017, December. An Integrated Deep Learning Framework Approach for Nail Disease Identification. In 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 197- 202). IEEE.

[6]. Truong, X., Mitamura, M., Kono, Y., Raghavan, V., Yonezawa, G., Do, T., Tien Bui, D. and Lee, S., 2018. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. Applied Sciences, 8(7), p.1046.

[7]. Oh, H.J. and Lee, S., 2017. Shallow landslide susceptibility modeling using the data mining models artificial neural network and boosted tree. Applied Sciences, 7(10), p.1000.

[8]. Ramesh, M.V., 2009, June. Real-time wireless sensor network for landslide detection. In 2009 Third International Conference on Sensor Technologies and Applications (pp. 405-409). IEEE.

[9].   Martha, T.R., Kerle, N., van Westen, C.J., Jetten, V. and Kumar, K.V., 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. IEEE Transactions on Geoscience and Remote Sensing, 49(12), pp.4928-4943.

[10]. Dahl, G.E., Sainath, T.N. and Hinton, G.E., 2013, May. Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8609-8613). IEEE.

[11]. Le, Q.V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B. and Ng, A.Y., 2011, June. On optimization methods for deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning (pp. 265-272). Omnipress.

[12]. Liao, X., Xue, Y. and Carin, L., 2005, August. Logistic regression with an auxiliary data source. In Proceedings of the 22nd international conference on Machine learning (pp. 505-512). ACM.

[13]. Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), pp.1145-1159.

[14]. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]

[15]. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? Catena 2018, 162, 177–192. [CrossRef]

[16]. Wang, Y.; Wu, X.; Chen, Z.; Ren, F.; Feng, L.; Du, Q. Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using smote for Lishui city in Zhejiang province, China. Int. J. Environ. Res. Public Health 2019, 16, 368. [CrossRef] [PubMed]

[17]. Kim, J.C.; Lee, S.; Jung, H.S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocarto Int. 2018, 33, 1000–1015. [CrossRef]

[18]. Lin, W.; Wu, Z.; Lin, L.; Wen, A.; Li, J. An ensemble random forest algorithm for insurance big data analysis. IEEE Access 2017, 5, 16568–16575. [CrossRef]

[19]. Louppe, G. Understanding random forests: From theory to practice. Ph.D. Thesis, University of Liege, Liege, Belgium, July 2014.