Predicting Damage to Buildings Caused by Earthquakes Using Machine Learning Techniques

Kuldeep Chaurasia
Bennett University
Greater Noida

Samiksha Kanse
Smt. Indira Gandhi College of Engineering
Mumbai

Aishwarya Yewale
Smt. Indira Gandhi College of Engineering
Mumbai

Vivek Kumar Singh
ABES Engineering
Gaziabad

Bhavnish Sharma
Thapar University
Patiala

B. R. Dattu
Bennett University
Greater Noida

Abstract- This paper presents the level of damage prediction to buildings caused by Gorkha Earthquake in Nepal using machine learning techniques. The predictions have been made based on mathematically calculated eight tectonic indicators and past vibrational activity records. In this research the objective is to predict earthquake damage on existing data set of seismic activity by using machine learning techniques. In this study, two well-known approaches of machine learning viz. Neural Network (NN) and Random Forest (RF) have been implemented and optimal parameters for accurate prediction are investigated. The analysis reveals that Random forest method has outperformed the neural network approach for building damage prediction. The F1 score using the random forest classification has been obtained as 74.32%.

I. INTRODUCTION

Earthquake is nothing but the shaking movement of Earth's crust. Elastic energy produced due to sudden crack and it get store in rocks that are subjected to great strain [1]. Energy produced during earthquake is getting store over long time and it will release in minutes or in seconds. Cracks on the rocks results in more elastic energy being stored which tends large possibilities of an earthquake. The Seismic waves are nothing but elastic waves which are produced by an earthquake. Seismic waves are low frequency waves those release energy during Earthquake can cause tremendous loss of human life. It results in serious damage to wide variety of civil engineering structure. Our project is based on Nepal earthquake which was held on April 25th, 2015 at Gorkha. It's also named as 'Gorkha earthquake'. earthquake was natural disaster to destruction Nepal [1, 4, 10].

The mortality rate due to earthquake in Nepal has never been less. In 2015 the mortality rate due to earthquake was nearly 9000 and 22,000 people was got injured. Hundreds of thousands Nepalese were made houseless due to these natural calamities. World heritage sites in Kathmandu valley and some other regions get destroyed due to damage caused by an earthquake. We are able to categories the damage rate and intensity of that earthquake using different reports. In this study, we are using artificial intelligence. Prediction keeps a set of input which decreases relative number of earthquakes from temporal distribution of part earthquake. Relative decrease in number of earthquakes is break in the normal seismic energy released from the region. Grades are used to represent the damage caused by an earthquake. Grades represent a damage to the building which was cause by the earthquake. There are three grades of the damage: Grade1, Grade 2 and Grade 3 are used to represent low, medium and complete destruction respectively [2].

The Nepal earthquakes of April and May 2015 killed around 9000 of people and injured around 20000. The Destruction caused was severe. Around 2 million were left homeless after the disaster. On April 25, an earthquake of magnitude of 7.8Mw strikes followed by an earthquake of magnitude 7.5 which caused massive destruction resulting in damage of schools, public health centers, water system, power systems, roads and bridges along with homes of people [2]. In the surveys conducted by various organization, the results had some very unusual findings such as those villages seemed to observe a less amount of damage were the ones located near the epicentre and to the west of the fault.

The earthquake showed very comparatively less magnitude of 5-6Mw as compared to northern region which had around earthquake of magnitude 7Mw. This showed that the intensity and magnitude of the earthquake was increasing in the North direction.

978-1-7281-4392-7/18/$31.00 ©2019 IEEE

II. RELATED WORK

The present paper consists of basic information about earthquake damage prediction. In this, we use Random forest algorithm as supervised classification algorithm. Random forest algorithm can be used both for classification and for regression kind of problems. Random forest algorithm is the most popular classification algorithm. Hosokawa et al. (2001) have introduced an approach for Landform classification method using self-organizing map and its application to earthquake damage evaluation. They have to classify typical landforms based on a land cover map and a digital elevation model (DEM). They have used supervised classification method with self-organizing map (SOM) [8]. Xu et al. (2008) have presented an approach for Earthquake disaster simulation for an urban area, with GIS, CAD, FEA, and VR integration. They have aim to develop an integrated urban earthquake simulation system (UESS) that uses GIS as the model source, CAD as the model generating tools, FEA as damage prediction, and virtual reality (VR) as the post-process platform [6]. Bhargava et al. (2009) have presented an approach for earthquake prediction using abnormal animal behavior. Earthquake prediction is a social importance and there is need to carry out research with Indian context [3].

Clarify et al. [4] have introduced the result for Earthquakes magnitude predication using artificial neural network in northern Red Sea area. They make use of artificial neural network. They have built it based on feedforward neural network model with multi-hidden layers. They show the up to 32% accuracy using neural network. Gičev et al. [5] have deployed the result for not predetermined earthquake damage scenarios for structural health monitoring. They have examined damage scenarios for site specific building models. They have made use of vibration health monitoring methods to kept track on change in frequency and hardness of the characteristic function. Brunner et al. [7] have introduced the result based on Change detection for earthquake damage assessment in built-up areas using very high resolution optical and SAR imagery. They have proposed a novel method that detects buildings destroyed in an earthquake. They have used pre-event very high resolution (VHR) multispectral and post-event detected VHR synthetic aperture radar (SAR) imagery.

W. Jiang et al. [9] have concluded the result based on Forecasting seismic damage in frame-shear wall structure based on Fuzzy neural network. They have used fuzzy neural network model. They have used Fuzzy neural network model to express the past actual earthquake damage experience in theoretical form. Mohanaselvi et al. [10] have introduced Application based on earthquake damage prediction using Fuzzy logic. In this paper, they use fuzzy logic for predicting damage due to earthquake. It deals with damage related to earthquake forecasting problem using fuzzy logic. They consider the fuzzy input like ground motion duration, type of soil, building height and maintenance. Guo et al. [11] have deployed the result of urban buildings damage prediction system. They introduced the analogue prediction-based system which was established for new rapid earthquake damage prediction. They classify buildings by their weight for that they used fuzzy mathematics.

Textual damage description-based analysis of the earthquake impacted buildings using deep learning classification has been presented in literature. To classify damage rate of building they used LSTM that is long term memory deep learning method. LSTM method is rapidly used by building professionals to access building cluster damage [12]. Hu et al. [13] have introduced result for cluster analysis of all buildings according to the damage level using artificial intelligence clustering algorithm. To display the analysis results in the form of graphs they combine the GIS platform.

Lee et al. [14] have presented an approach for earthquake damage prediction system using lattice closeness degree, this method simplifies the calculations and provide high accuracy. Lattice closeness degree is improved, and most accurate method used for simplification of calculation. Artificial neural network for seismic damage prediction of buildings has been adopted in literature [15]. In this system, they have used Multilayer Feedforward Perceptron networks. Multilayer Feedforward Perceptron networks is used for approximation of an unknown function and for pattern recognition. The overall aim of this research is to track actual scenario of earthquake, what type of injuries can be expected and how can these be reduced? Asim et al. [16] suggested a new methodology for modeling and predicting future news events using machine learning and data mining techniques. They have reported that the Pundit algorithm generalizes examples of causality pairs to infer a causality predictor. The extraction of quantitative association rules and regression techniques have been adopted to discover patterns that is used to model the behavior of seismic temporal data to help in earthquakes prediction [17,19]. Deep learning techniques for prediction of different types of disasters have also been utilised by several researchers which can be found in the literature [18, 20, 21]

III. DATA RESOURCES

This data set has been taken from the Driven Data competition platform. There are 3 dataset files provided.

Train values and train labels are used for training the model for the prediction of required valued output, one dataset is for training the model and the other for testing the model for the required output, that is damage grade. Test values is used for testing the model for required output value measuring the accuracy of the trained model [2].

The dataset mainly consists of information on the building's structure. Each row in the dataset represents a specific building in the region that was hit by Gorkha earthquake. In this project we convert categorical data into integer form. There are 39 columns in this dataset, where the building id column is a unique and random identifier. The remaining 38 features are used in testing model. Some of the important features include (i) Land surface condition (ii) Foundation type (iii) Roof type (iv) Ground floor type (v) Position (vi) Plan configuration (vii) Legal ownership status. Detailed description of the other features can be found in the literature [2].

IV. METHODOLOGY

The first step in methodology for research work includes the pre-processing of the input data which consists of noise removal and filling up the empty cells. Next step to balance the data, which involves conversion of all the categorical features in continuous features. Third step is to split the data into training and testing. Now we have to choose a model to train the dataset on it. Once the data is trained it is tested on the testing dataset. After the model is trained and tested, we compute the accuracy of the model to check up to what extent model is predicting correctly.

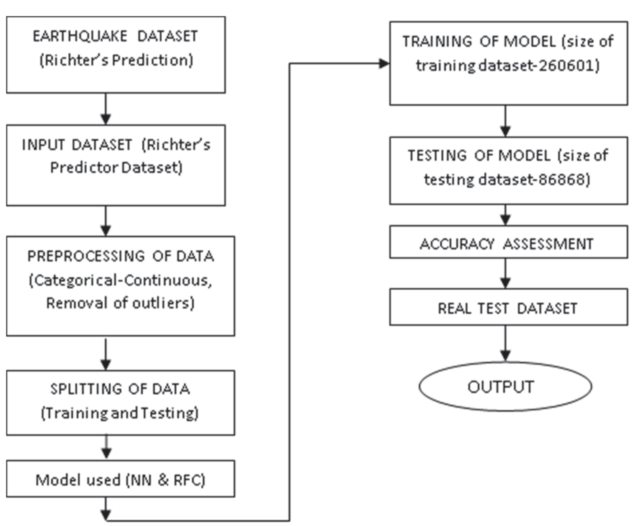

Fig 1. Methodology

Now if the desired accuracy is achieved, we apply the model on the real Test data and generate the output. Methodology for the research has been shown in Fig.1.

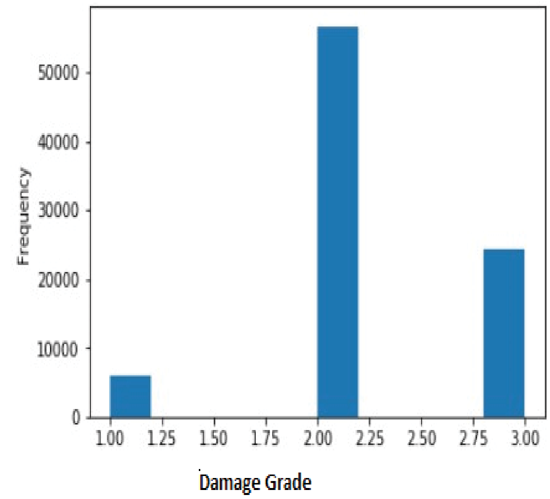

Fig. 2. Relation between frequency and damage grade

V. RESULT AND DISCUSSION

To measure the performance of our algorithms, we used the F1 score which balances the precision and recall of a classifier. Traditionally, the F1 score is used to evaluate performance on a binary classifier, but since we have three possible labels, we used a variant called the micro averaged F1 score as shown in equation (1).

$$F_{\mu} = \frac{2.P\mu\,.R\mu}{P\mu+R\mu} \tag{1}$$

$$\text{Precision } (P_{\mu}) = \frac{\sum_{k=1}^{3} TP}{\sum_{k=1}^{3}(TP+FP)}, \tag{2}$$

$$\text{Recall } (R_{\mu}) = \frac{\sum_{k=1}^{3} TP}{\sum_{K=1}^{3}(TP+FN)} \tag{3}$$

where, TP =True Positive

FP = False Positive
FN = False Negative
k represents each class in 1,2,3.

The formula for Precision and Recall is shown in equation (2) and equation (3) respectively.

Neural Network: - Neural networks are algorithms that represent the working of human brains. It helps in identifying patterns hidden in the data. Neural networks help to classify the data into classes. In the current research work, neural network architecture with six dense hidden layers has been used. The The Activation functions used in the code are tanh, relu, sigmoid, SoftMax functions and the optimizer used is "Adam". Loss Function used is Mean squared error. First, we trained our model on 100 epochs, and the accuracy was 57%. So, we increased the epochs to 1000 and got 62.8% accuracy. Further, we increased the epochs but there was no variation in the prediction accuracy.

Table 1. Optimized Parameters and Performance Score

S.N.	MODEL USED	PARAMETERS USED	F1-Score
1.	NERUAL NETWORK	Hidden Layer- six dense layers Activation Function- tanh, relu, softmax, sigmoid Optimizer-Adam Epochs- 1000 Loss Function- mean square error	0.628
2.	RANDOM FOREST	Number of estimators- 750 max_parameter-25 leaf_size-5	0.743

Random Forest Classifier- It predicts based on the majority of the votes from each of the decision trees made. We use the no. of estimators to be 750 which decides number of trees in the forest. The higher the number of the estimators the better accuracy we will get, but it becomes more difficult to compute, all the enlarged data and it will take a lot of time for the code to finish and providing output. The criterion is of two types, Gini or entropy, and it help in measuring the quality of each split in data in trees. We have used the Gini impurity, here in our model because it does require us to calculate logarithmic functions, which are difficult to compute, takes much longer time. We have used maximum parameters to be 25. These specify the number of features used whilst looking for the best split. Increasing its value increases the accuracy but once again, by increasing its value makes it difficult to process and it will take a lot of time for getting output results. We have used min samples leaf value to be five as this parameter helps us in determining the size of node of all the decision trees. We set the Random state perimeter to be 42. This parameter helps us to reproduce the same results if we provide the same training data and all the other parameters are also kept same. The relation between the frequency and damage grade has been shown in Fig. 2. Performance of the two models using F1 score is shown in Fig.3.

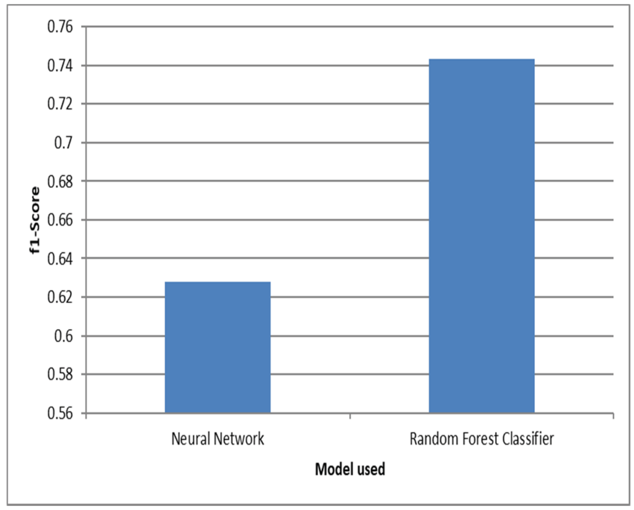

Fig. 3. Relation between F1 score and model used

VI. CONCLUSION

This study successfully built different scenarios of earthquake damages using different features. In this study, we are examining the different damage grades due to earthquake in Nepal. In this research work, Neural Network (NN) and Random Forest algorithm with parameter tuning to assess the condition of the building structures. The analysis reveals that Random forest method has outperformed the neural network approach for building damage prediction. The F1 score using the random forest classification has been obtained as 74.32%. This approach is simple, convenient and easy for computer program implementation. It also contributes to rapid earthquake damage loss assessment.

ACKNOWLEDGEMENT

We would also like to thank Driven Data for providing all the data requirement for this internship project. We would like to thank Dr. Madhushi Verma for providing all the facilities helpful for our research project. We would also like to show our gratitude to HOD, Computer Science Engineering, Bennett University for providing the resources required for the project.

REFERENCES

[1] Goda, K., Kiyota, T., Pokhrel, R.M., Chiaro, G., Katagiri, T., Sharma, K. and Wilkinson, S., 2015. The 2015 Gorkha Nepal earthquake: insights from earthquake damage survey. Frontiers in Built Environment, 1, p.8.

[2] https://www.drivendata.org/competitions/57/nepal-earthquake/page/136/ [access date- 24 June, 2019]

[3] Bhargava, N., Katiyar, V. K., Sharma, M. L., & Pradhan, P. (2009). Earthquake prediction through animal behavior: A review. *Indian J. Biomech*, *78*, 159-165.

[4] Alarifi, A. S., Alarifi, N. S., & Al-Humidan, S. (2012). Earthquakes magnitude predication using artificial neural network in northern Red Sea area. *Journal of King Saud University*-Science, *24*(4), 301-313.

[5] Gičev, V., & Trifunac, M. D. (2012). A note on predetermined earthquake damage scenarios for structural health monitoring. Structural control and health monitoring, 19(8), 746-757.

[6] Xu, F., Chen, X., Ren, A., & Lu, X. (2008). Earthquake disaster simulation for an urban area, with GIS, CAD, FEA, and VR integration. Tsinghua Science and Technology, 13(S1), 311-316.

[7] Brunner, D., Bruzzone, L., & Lemoine, G. (2010, July). Change detection for earthquake damage assessment in built-up areas using very high resolution optical and SAR imagery. In 2010 IEEE International Geoscience and Remote Sensing Symposium (pp. 3210-3213).

[8] Hosokawa, M., & Hoshi, T. (2001, July). Landform classification method using self-organizing map and its application to earthquake damage evaluation. In IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217) (Vol. 4, pp. 1684-1686).

[9] W. Jiang and B. X. Gu, "Forecasting Seismic Damage in Frame-Shear Wall Structure Based on Fuzzy Neural Network", Advanced Materials Research, Vols. 594-597, pp. 1692-1695, 2012

[10] Ram, T. D., and Wang, G. (2013). Probabilistic seismic hazard analysis in Nepal. *Earthquake Eng. Eng. Vib.* 12, 577–586.

[10] Mohanaselvi, S., & Hemapriya, G. (2019, June). Application of fuzzy logic to earthquake damage predictions. In *AIP Conference Proceedings* (Vol. 2112, No. 1, p. 020043). AIP Publishing.

[11] Guo, X. D., Su, J. Y., Ma, D. H., & LI, G. (2006). Rapid earthquake damage prediction system of urban buildings. *Journal of Natural Disasters*, *15*(3), 128.

[12] Mangalathu, S., & Burton, H. V. (2019). Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions. International Journal of Disaster Risk Reduction, 36, 101111.

[13] Hu, C., Cui, Z., Lin, J., Huang, M., & Liu, S. (2019, June). Application Research of Clustering Algorithm in Earthquake Disaster Prediction System of Sanhe. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022013). IOP Publishing.

[14] Lee, J., Lin, J., & Liu, J. (2018). An Improved Method for Earthquake Damage Prediction of Highway Subgrade and Pavement Based on Closeness Degree Method. *American Journal of Civil Engineering*, *6*(4), 118-122.

[15] Morfidis, K., & Kostinakis, K. (2018). Approaches to the rapid seismic damage prediction of r/c buildings using artificial neural networks. *Engineering Structures*, *165*, 120-141.

[16] Asim, K. M., Martínez-Álvarez, F., Basit, A., & Iqbal, T. (2017). Earthquake magnitude prediction in Hindukush region using machine learning techniques. Natural Hazards, 85(1), 471-486.

[17] Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. Geophysical Research Letters, 44(18), 9276-9282.

[18] Vetrivel, A., Gerke, M., Kerle, N., Nex, F., & Vosselman, G. (2018). Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS journal of photogrammetry and remote sensing*, *140*, 45-59.

[19] Martínez-Álvarez, F., Troncoso, A., Morales-Esteban, A., & Riquelme, J. C. (2011, May). Computational intelligence techniques for predicting earthquakes. In International Conference on Hybrid Artificial Intelligence Systems (pp. 287-294). Springer, Berlin, Heidelberg.

[20] Bai, Y., Gao, C., Singh, S., Koch, M., Adriano, B., Mas, E., & Koshimura, S. (2017). A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks. IEEE Geoscience and Remote Sensing Letters, 15(1), 43-47.

[21] Ross, Z. E., Meier, M. A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894-2901.