Earthquake prediction with machine learning models based on peak of radon anomalies

Maria Heinrich
2110837746@stud.fh-kufstein.ac.at
FH Kufstein Tirol
Kufstein, Austria

ABSTRACT

Earthquake prediction is currently the most important task required for probability, hazard, risk mapping, and mitigation. In the past, various traditional and machine learning models have been used for risk assessment. It is unlikely that anyone will ever be able to predict earthquakes accurately, but with advancements in deep learning algorithms, predictions can become more precise and closer to the actual natural disaster. Different machine learning approaches and deep learning models based on radon anomaly detection have been compared, opening the field for further developments.[16]

KEYWORDS

neural network, machine learning, deep learning, earthquake prediction, radon anomalies time series prediction, earthquake likelihood, natural science

1 INTRODUCTION

The earthquake is considered a natural disaster, whose unpredictable magnitude results in the loss of human life and infrastructure. Earthquakes can be devastating, causing billions of dollars in damage and with more accurate prediction of deadly earthquakes, the worst damage can be prevented through countermeasures such as evacuation and securing the area. [9]

The issue of earthquake prediction involves predicting the minimum Richter magnitude scale in a specific geographic area and, as a result, determining the probability of an earthquake occurring in that region within a specified time frame. [2]

Earthquakes occur when the edges of the faults interlock and the rest of the block moves, storing the energy that would otherwise cause the blocks to slide past. When the force of the moving blocks finally overcomes the friction of the jagged edges of the fault and it disengages, all of the stored energy is released. The energy radiates from the fault in the form of seismic waves in all directions, like ripples on a pond. The seismic waves shake the earth as they move through it, and when the waves reach the earth's surface, they shake the ground and everything on it. [16]
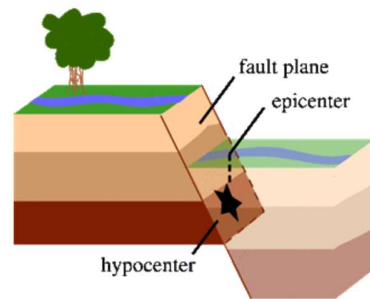

Figure 1: A normal (dip-slip) fault is an inclined fracture where the rock mass above an inclined fault moves down and can causes an earthquake. [16]

In the words of Charles Richter, the father of the Gutenberg-Richter law and the Richter scale for earthquakes, "only fools, charlatans and liars predict earthquakes."

Earthquake prediction is the long-sought holy grail of earthquake research and continues to puzzle geoscientists, but not much progress has been made until the last decade. Significant results have been achieved with the advancement of computer systems, machine learning and deep learning models. [4]

State-of-the-art techniques from the field of artificial neural networks and deep learning for predicting rare phenomena of interest play an important role.

During the preparation for an earthquake beneath the earth's surface, various geophysical and seismological processes occur. Radon and its radioactive isotope, thoron, extracted from uranium and thorium sources deep in the earth, have the potential to predict upcoming earthquakes. [1]

Many researchers have attempted to link radon anomalies in soil and water to seismic activity with statistically significant results. [9]

Consequently, with recent advances in computer science, various computational intelligence techniques have been successfully introduced to predict radon concentration using meteorological parameters. Regression trees were used to predict radon soil gas

concentration using environmental data such as air pressure, precipitation, air temperature, and soil temperature, and concluded that the prediction error increases one week before earthquakes with magnitude from 0.8 to 3.3. [1]

2   SEISMIC DATA

Muzaffarabad is seismically active area and has history of occurrence of regular devastating earthquakes, so forecasting possible earthquake in future is an attractive field of study. Department of Physics and Institute of Geology of The University of Azad Jammu & Kashmir in Muzaffarbad, Pakistan have been involved in the study of soil radon gas and have therefore conducted radon gas measurements at 60 different locations to map the geohazardous zone of the city of Muzaffarabad. [15]
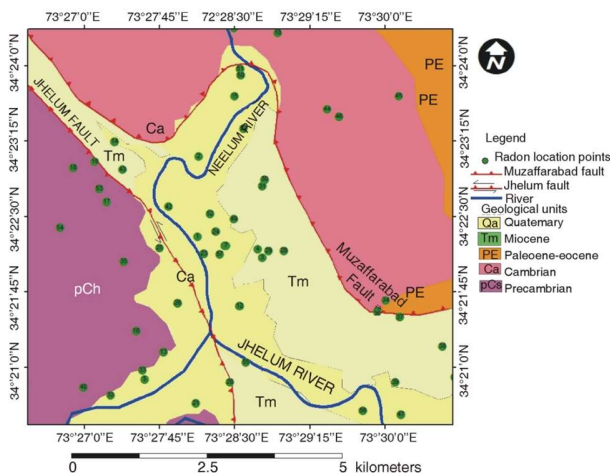

Figure 2: Geological map of Muzaffarabad city and sur rounding area. Tareen, A. et al. 2016. [15]

These data and self-generated data from the same area are used in a variety of publications and as the basis for a wide range of research approaches. A recent publication focuses on predicting earthquake probabilities in radon time series using an ensemble model based on machine learning.

Therefore, a RADON measuring station with RTM-1688-2 SARAD nuclear Instrument has been installed over a fault line passing beneath the Muzaffarabad to generate data in the period of March 1, 2017 to May 11, 2018 and the dataset contains 15692 valid observations of radon concentration along with environmental parameters such as thoron (Bq/m^3), temperature (^{0}C), relative humidity, and pressure (mbar), and nine seismic activities were observed during the data collection period. [1]

3   FORECASTING METHODOLOGIES

Various ensemble and individual machine learning methods have been used to predict radon gas concentration in soil based on various environmental features. The ensemble method in machine learning and statistics refers to the collection of multiple models that combine different hypotheses to determine the best hypothesis by combining weak learners into a strong learner. In experiments, ensemble methods have been shown to produce more accurate results when there is a large amount of diversity between models. [1]

Boosting and bagging are two methods used in ensemble learning to generate base learners. Boosting is a sequential ensemble method that improves the performance of a base learning algorithm by focusing on hard-to-predict examples and increasing the weight of misclassified examples. Bagging, a parallel ensemble method, reduces error by combining independent base learners. In bagging, multiple bootstraps of the training data are created, and a base learning algorithm is trained on each of these bootstraps, resulting in multiple base learners. The final predictions are created by combining the predictions of each base learner through voting or averaging. Bagging helps reduce the variance of a single base learner and minimizes overfitting, especially for high variance base learners. [1]

The K-NN algorithm is a non-parametric method used for prediction. It involves determining the similarity of features between test and training samples. The distance is calculated using either Euclidean or Manhattan distance. Then, the number K of nearest neighbors is selected and used to predict or classify the new sample. The steps of the algorithm are to read the training and test data set, initialize the value of K, calculate the distance between the test sample and the training set, sort the distances in ascending order, select the first K entries, and return the mean of the K response values as the prediction value for the current test sample. [1]

The support vector machine (SVM) is a machine learning tool used for classification and regression tasks. It was designed for binary classification but can be modified for regression (Support Vector Regression). In SVM regression, the goal is to find a function g(x) that doesn't deviate more than a certain value (epsilon) for each training point. The process involves solving a convex optimization problem to minimize the function through residuals with values less than epsilon. Slack variables are introduced for instances when there is no such function g(x) that satisfies all constraints. The C parameter helps prevent overfitting. The loss is measured from the distance between the epsilon boundary and the observed value. The Lagrange dual is obtained by introducing non-negative multipliers for each instance. The function used to predict test or new values is obtained by summing the values of all instances. [1]

The simulation is performed for two different groups of machine learning methods. Group 1 consists of ensemble methods for learning such as are Boosted Tree Model, Bagged Cart Model, and Boosted Linear Model, while Group 2 contains individual learning methods such as K-NN and SVMs with linear and radial kernels. The simulation is performed in four different settings, ranging from setting 1 to 4. The basic purpose of introducing these settings is to investigate the predictive ability of the learned

models on different test data sets containing almost any seismic activity. In addition to the different distributions of the training and test data, the time window is also considered.
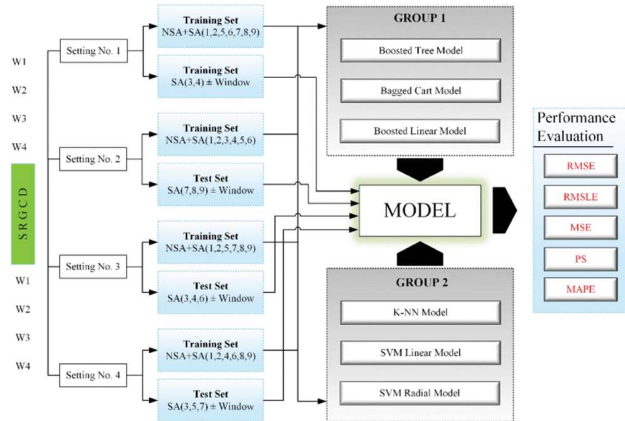

Figure 3: Proposed framework of A. A. Mir et al. 2022. [1]

The performance of the methods is more vividly evaluated by including different training and test data set distributions through settings from 1 to 4. The training dataset is composed of various seismic activities and normal data, while the test data is based on seismic activities with associated time windows from 1 to 4. In setting 1, Boosted Tree and Support Vector Machine (SVM) with radial kernel performed the same and captured temporal variations in the time series more effectively. For setting 2, the linear Boosted model has the lowest RMSE, and other performance measures did not capture the temporal variations in the time series. In addition, Support Vector Machine with linear kernel and Boosted Tree performed relatively better than other models. In Settings 3 and 4, the Boosted Tree model performed better compared to other ensemble and individual models because it more accurately predicted radon concentration in soil gases. [1]

In a previous work by A. A. Mir et al. 2021. [10], a method for dividing the ground radon gas into seismically active and non-active time series was developed, taking advantage of a stacking ensemble approach, and using an automatic anomaly display function as a post-processing technique.

Radon time series were recorded again over a fault line in Muzaffarabad, a city in the Kashmir region of Pakistan, from March 1, 2017, to February 28, 2018, including four seismic events during period.

The main idea behind the proposed method consists of two layers. The first layer uses a stacking ensemble-based approach that incorporates three learners, i.e., a generalized linear model, linear regression, and K-nearest neighbors, to train on seismically active and inactive periods to predict soil radon gas concentration. These predictions are then combined with the labelled anomaly data to train a meta-learner, i.e., a support vector machine with a radial kernel that categorizes the series into active and non-active radon time series data. In the second layer, these classifications are then passed to an automatic anomaly display function that assigns the

time series to a group of readings for which the degree of display obtained is greater than or equal to the display factor. The conclusion of the study shows that the proposed methodology accurately localizes the anomaly in the radon time series data at different window sizes, i.e., in terms of individual days. The evaluation is based solely on the classification after processing radon time series data by both layers of the proposed methodology. The results show that the proposed methodology efficiently classifies the radon concentration data with a testing accuracy of 0.971 to 0.968. [10]

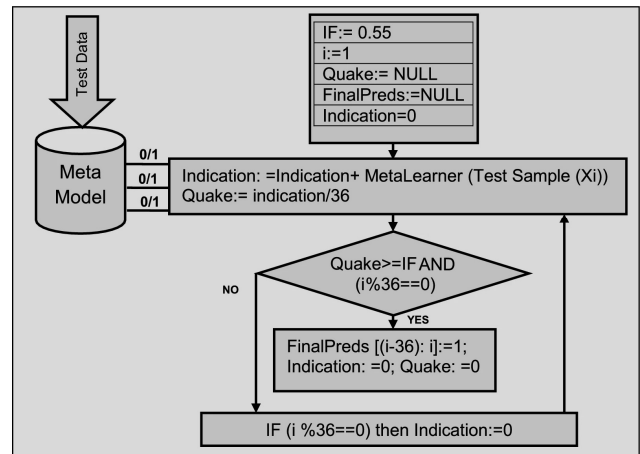**Automated Anomaly Indication Function:**

Figure 4: Schematic of the automated anomaly indication function A. A. Mir et al. 2021. [10]

4 RESULTS

The first paper by A. A. Mir et al. 2022. [1] concludes that ensemble methods lead to relatively better regression models and that the support vector machine with radial kernel performs better than the boosted tree model in settings 3 and 4. In this study, a Boosted-Tree method is proposed for automatic prediction of soil radon gas concentration based on environmental parameters in soil radon time series and it can be extended to classify anomalies in the predicted concentration as performed in the paper by Mir, A.A., Çelebi, F.V., Rafique, M. et al. 2021. [10]

5 CONCLUSION AND OUTLOOK

The focus of the first study is to predict the radon gas concentration in the soil in the presence of anomalies, while the second study classifies the anomalies in the predicted radon concentration. A combination of the two approaches sounds very promising. [1] In addition, a Kaggle competition for the dataset could provide further insights, as was the case in the study *Laboratory Earthquake Prediction: A Machine Learning Competition* [4]

REFERENCES

[1]   A. A. Mir et al. 2022. Anomalies Prediction in Radon Time Series for Earthquake Likelihood Using Machine Learning-Based Ensemble Model. *IEEE Access, vol. 10*, pp. 37984 – 37999. DOI:https://doi.org/10.1109/ACCESS.2022.3163291

[2]   Doğan, A., Demir, E. 2022. Structural recurrent neural network models for earthquake prediction. *Neural Comput & Applic 34*, 11049–11062. DOI:https://doi.org/10.1007/s00521-022-07030-w

[3]   Chetia, T., Baruah, S., Dey, C. et al. 2022. Seismic induced soil gas radon anomalies observed at multiparametric geophysical observatory, Tezpur (Eastern Himalaya), India: an appraisal of probable model for earthquake forecasting based on peak of radon anomalies. *Nat Hazards 111*, 3071–3098. DOI:https://doi.org/10.1007/s11069-021-05168-9

[4]   Johnson, P. A., Rouet-Leduc, B., Pyrak-Nolte, L. J., Beroza, G. C., Marone, C. J., Hulbert, C., ... & Reade, W. 2021. Laboratory earthquake forecasting: A machine learning competition. Proceedings of the National Academy of Sciences, 118(5). DOI:https://doi.org/10.1073/pnas.2011362118

[5]   Khan, M. M., Ghaffar, B., Shahzad, R., Khan, M. R., Shah, M., Amin, A. H., ... & Ali, R. 2022. Atmospheric Anomalies Associated with the 2021 M w 7.2 Haiti Earthquake Using Machine Learning from Multiple Satellites. Sustainability, 14(22), 14782. DOI:https://doi.org/10.3390/su142214782

[6]   Pulinets, S. A. 2007. Natural radioactivity, earthquakes, and the ionosphere. *Eos, Transactions American Geophysical Union, 88(20)*, 217-218. DOI:https://doi.org/10.1029/eost2007EO20

[7]   Won, J., Park, J., Park, J. W., & Kim, I. H. 2020. BLESeis: Low-cost IOT sensor for smart earthquake detection and notification. Sensors, 20(10), 2963. DOI:https://doi.org/10.3390/s20102963

[8]   Tareen, A. D. K., Nadeem, M. S. A., Kearfott, K. J., Abbas, K., Khawaja, M. A., & Rafique, M. 2019. Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data. Applied radiation and isotopes, 154, 108861. DOI:https://doi.org/10.1016/j.apradiso.2019.108861

[9]   Keskin, S., Külahcı, F. 2022. ARIMA model simulation for total electron content, earthquake and radon relationship identification. Nat Hazards. DOI:https://doi.org/10.1007/s11069-022-05622-2

[10]   Mir, A.A., Çelebi, F.V., Rafique, M. et al. 2021. Anomaly Classification for Earthquake Prediction in Radon Time Series Data Using Stacking and Automatic Anomaly Indication Function. Pure Appl. Geophys. 178, 1593–1607. DOI:https://doi.org/10.1007/s00024-021-02736-9

[11]   Rafique, M., Tareen, A.D.K., Mir, A.A. et al. 2020. Delegated Regressor, A Robust Approach for Automated Anomaly Detection in the Soil Radon Time Series Data. Sci Rep 10, 3004. DOI:https://doi.org/10.1038/s41598-020-59881-9

[12]   Jena, R., Pradhan, B., Naik, S. P., & Alamri, A. M. 2021. Earthquake risk assessment in NE India using deep learning and geospatial analysis. Geoscience Frontiers, 12(3). DOI:https://doi.org/10.1016/j.gsf.2020.11.007

[13]   Duong, V. H., Ly, H. B., Trinh, D. H., Nguyen, T. S., & Pham, B. T. 2021. Development of Artificial Neural Network for prediction of radon dispersion released from Sinquyen Mine, Vietnam. Environmental Pollution, 282. DOI:https://doi.org/10.1016/j.envpol.2021.116973

[14]   Weiss, J. R., Qiu, Q., Barbot, S., Wright, T. J., Foster, J. H., Saunders, A., ... & Echalar, A. 2019. Illuminating subduction zone rheological properties in the wake of a giant earthquake. Science advances, 5(12), eaax6720. DOI:https://doi.org/10.1126/sciadv.aax6720

[15]   Tareen, A. D. K., Rafique, M., Kearfot, K. J., Basharat, M., & Shafique, B. 2016. Soil gas radon mapping of Muzaffarabad city, Pakistan. Nuclear Technology and Radiation Protection, 31(3), 291-298. DOI:https://doi.org/10.2298/NTRP1603291T

[16]   Earthquake Hazards | U.S. Geological Survey. 2022. Lisa Wald (U.S. Geological Survey) for "The Green Frog News". https://www.usgs.gov/programs/earthquake-hazards/science-earthquakes

[17]   LANL Earthquake Prediction | Kaggle. 2022. Kaggle Home Page. https://www.kaggle.com/competitions/LANL-Earthquake-Prediction

[18]   Global Observatories: GFZ. 2022. GEOFON Home Page. https://geofon.gfz-potsdam.de/

[19]   Erdbebenarchiv — ZAMG. 2022. Erdbebenarchiv Home Page. https://www.zamg.ac.at/cms/de/geophysik/erdbeben/erdbebenarchiv