Original Article

Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques

Aditya Saxena[1], Rishabh Chauhan[2], Devansh Chauhan[3], Dr Shilpi Sharma[4],Dr. Dolly Sharma[5] and Vipul Narayan[6]

[1-5]Amity School of Engineering and Technology, Amity University Noida
[6]School of Computing Science and Engineering, Galgotias University, Greater Nida Uttar Pradesh

{Adityaarchie1234, rishabh97.chauhan, devanshchauhan19990 }@Gmail.com, ssharma22@amity.edu, dolly.azure@gmail.com, vipulupsainian2470@gmail.com
DOI: 10.47750/pnr.2022.13.S10.767

Abstract

This paper proposes a machine-learning model for earthquake prediction. Earthquakes are complex and unpredictable natural phenomena, making it challenging to predict them accurately. However, recent advances in machine learning techniques have shown promise in predicting earthquakes by analyzing various factors such as seismic activity, geospatial data, and weather patterns. In this study, we collected earthquake data from multiple sources and used it to train a machine-learning model. We evaluated the model's performance using accuracy, precision, and recall metrics. Our results demonstrate that our machine-learning model can accurately predict earthquakes with high precision and recall. The model has the potential to provide early warnings of earthquakes, which can help reduce the damage caused by these disasters. Overall, this study highlights the potential of machine learning in earthquake prediction and provides a roadmap for future research in this field.

Keywords: — Earthquake Prediction, Machie Learning, Regression.

1 Introduction

The natural phenomena of earthquakes have been troublesome for Human civilization since the development of human consciousness and society. The movement of the tectonic plates is what causes this natural phenomenon. Evidence of earthquakes can even be found in historical texts which date back hundreds of years[1]. The phenomenon of earthquakes is more common than we expect it to be. Earthquakes occur every day around the world. However, these may need to be noticed as the magnitude of such earthquakes is usually less and cannot be felt, generally appearing as small tremors. Earthquakes of high volume and intensity cause tremendous damage to life and property. What makes earthquakes dangerous is that the movement of tectonic plates, the sole reason for significant earthquakes, is unpredictable. Earthquakes are especially devastating in coastal areas as the movement in tectonic plates is often accompanied by tsunamis and aftershocks, taking a considerable toll on human life and property. The point where the movement of tectonic plates occurs is called the epicenter of an earthquake. The earthquakes occur due to the shifting of tectonic plates, vast fragments of the Earth's crust that fit together like a puzzle. These plates can move and collide with each other, causing friction and building up energy that is eventually released as an earthquake. This energy release happens when the plates suddenly move along a fault line, a crack in the Earth's surface. This movement creates seismic waves that spread through the Earth's crust, causing the ground to shake. Additionally, volcanic eruptions can also lead to earthquakes. When lava, ash, and gas escape from a volcano, it can cause the ground to tremble, resulting in an earthquake. Moreover, human activities such as underground mining and large construction projects can also cause tremors, called induced earthquakes. In Table 1, we can see the various reasons which may cause earthquakes.

Table 1. Causes of Earthquakes

S. no	Activity	Descriptions
1.	Movements along tectonic plate borders are what create the majority of earthquakes. An earthquake can occur when two plates pass by or collide and lock, storing energy until it is unexpectedly released**.**	Tectonic Activity
2.	Volcanic activity, such as the flow of magma beneath the Earth's surface or the collapse of a volcanic structure, can also cause earthquakes.	Volcanic activity

3.	Building massive dams or structures, drilling for oil, or underground nuclear explosions can cause or amplify some earthquakes.	Human activities
4.	Landslides, glacial isostatic adjustment, and rebounding of the Earth's crust after an ice age are other natural occurrences that might result in earthquakes.	Other natural causes
5.	Giant meteor impacts that unleash enormous amounts of energy can result in extraordinarily unusual but highly violent earthquakes.	Meteor impacts

The abrupt energy released in the Earth's crust causes earthquake waves. Typically, this energy is held in rocks along faults or the boundary of tectonic plates, accumulating over time due to tectonic pressures. Rocks break when pressure on them surpasses their capacity for resistance, releasing the tension as seismic waves radiate outward from the earthquake's epicenter. The seismic waves force the ground to quiver and shake as they move through the Earth, creating the impacts of an earthquake. The kind and scale of the earthquake and the features of the rocks and soil that the waves pass through determine the seismic waves' characteristics, such as their speed, direction, and amplitude.

On the surface, these collisions are often slow and imperceptible, but tremendous tension can develop between the plates. Rapid release of this tension causes seismic waves, potent vibrations, to travel hundreds of miles through the rock and rise to the surface. Earthquakes produce waves, namely P-waves, S-waves, and surface waves. P-waves are the fastest and can travel through solids and liquids, while S-waves are slower and only travel through solids. Surface waves are the slowest and cause the most damage by traveling along the surface of the Earth and producing the shaking felt during an earthquake.

2 Literature Review

In [a], the authors analyze Eight seismic indicators that were mathematically derived from the local earthquake database and have been used to make predictions. These traits have been identified based on well-known geophysical facts, including seismic quiescence, the distribution of typical earthquake magnitudes, and

Gutenberg-inverse Richter's law. Each of the four machine learning techniques used in this study—the pattern recognition neural network, recurrent neural network, random forest, and linear programming boost ensemble classifier—is used to model the relationships between calculated seismic parameters and potential earthquake occurrences. In [2], the authors discuss the short-term and long-term prediction of earthquakes. They also discuss the historical evidence and effects of earthquakes on regions like Japan. The prediction of earthquakes has challenged Humanity since prehistoric times, as the movement of the tectonic plates cannot be predicted.

The authors in [3] try to analyze and predict earthquakes using machine learning in a controlled environment. They demonstrate how machine learning can accurately anticipate how long a laboratory failure will last by listening to the acoustic signal it emits. These forecasts do not consider the acoustical signal's past but are based simply on its current physical properties. In [4], Rui et al. proposed a deep learning-based model for earthquake prediction. They critically review the existing methods for earthquake predictions, i.e., explicit indicators designed by geologists and more implicit deep learning methods. Rui et al. have combined these two methods to develop a deep learning method named DLEP, which utilizes eight pattern-based indicators as explicit vectors and uses CNN as the heart of computation [13][14].

In [5], Sujith et al. analyzes the aftermath caused by earthquakes. They explicitly utilize machine learning to predict the damage caused by earthquakes to artificial structures and buildings. Their work aims to speed up the process of analyzing areas experiencing the most intense earthquake shockwaves. Models such as KNN and Random forest classifiers are used for this purpose. They have also proposed a parameter that classifies existing buildings regarding earthquake tolerance, calculated using machine learning. In [6], Xie et al. have adopted a hierarchical matrix structure to classify existing literature and studies into four categories depicted in Figure 1.

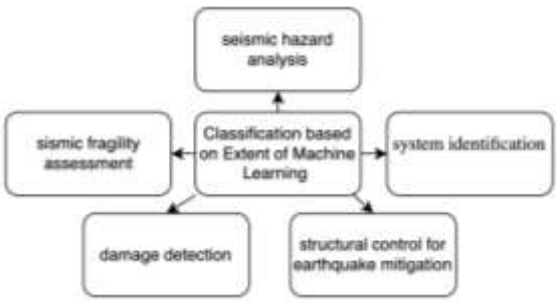

Fig. 1. Classification is done in [6]

In [7], Xiong et al. proposed a machine learning algorithm combining the Adaboost variant and pruning classification trees. The pre-processed data is transferred into an IBPT classification algorithm reducing the impact of less contributed data. In [8], Mousavi et al. have worked on earthquake magnitude prediction. They are using a Stanford Earthquake Dataset to train the network, which contains 1 million one-minute seismographic data. The dataset's seismograms are records of minor events with magnitudes under 2.5. Twenty-three various scales have been used to report volumes. In [9], the authors have used a simple machine-learning approach to predict earthquakes. The main goal of this research in [j]is to use machine learning methods to forecast whether a significant earthquake would be labeled as a negative or positive event. Researchers in [10] talk about using AI in power plants to generate renewable energy with the help of wind speed. In areas with high seismic activity, earthquakes can cause the ground to shake and move, potentially damaging wind turbines and affecting wind power generation. Therefore, when considering the deployment of wind turbines, it is crucial to consider the seismic risk of the area and design structures that can withstand earthquakes. In [11], the researchers dive deep to find security solutions for IoT and big data. The real-time collection and analysis of enormous volumes of data from seismic sensors in earthquake-prone zones are necessary for earthquake early warning systems. By gathering, processing, and transmitting data in real-time, the Internet of Things (IoT) and Big Data technologies can be utilized to increase the effectiveness and precision of earthquake early warning systems. However, security issues exist with the present IoT system's centralized servers. Cyberattacks may compromise the data that the sensors have collected on these sites. It is essential to adopt suitable security measures to ensure the accuracy and integrity of the data gathered by the sensors and communicated to earthquake early warning systems. Researchers in [12] present the use of LSTMs in forecasting sea surface temperature, which might be used to monitor and predict earthquake activity in real-time. According to specific research, seismic activity may be brought on by changes in sea surface temperature brought on by anthropogenic factors such as global warming. An increase in seismic activity could result from changes in mass distribution and pressure on the Earth's crust brought on by melting glaciers and polar ice caps as the planet's climate changes [15][16].

3 Methodology

3.1 Proposed Model and Architecture

Authors have compared the accuracy of different models, both classifiers and regressors. Regression models forecast a continuous target variable, such as the amount of damage produced by an earthquake, using one or more predictor variables. The proposed architecture can be seen in Figure 2.

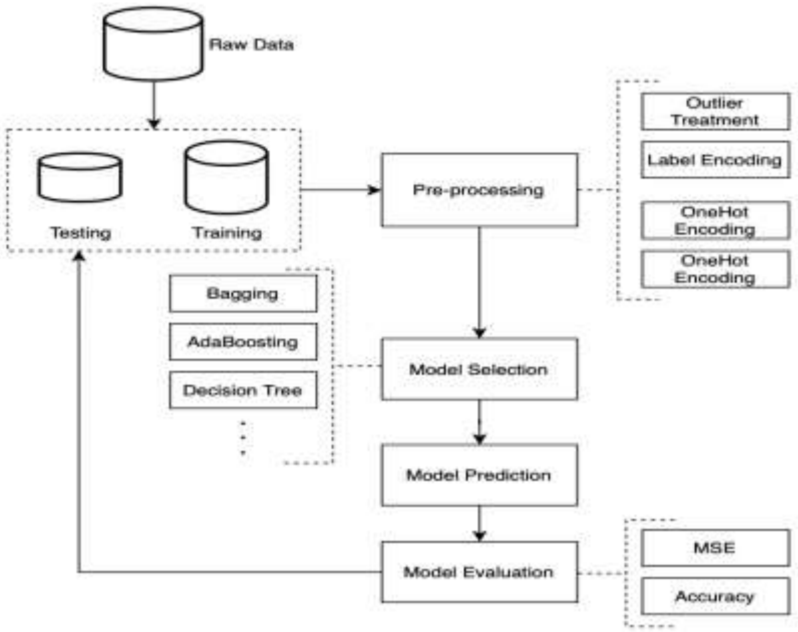

Fig. 2. Proposed Architecture

3.2 Dataset

The data for the research is acquired from an open-source website. Still, it is collected and maintained by the National Planning Commission (NPC) and the Central Bureau of Statistics. These authorities have developed one of the largest databases after a disaster, encompassing essential data on the effects of earthquakes, home situations, and socioeconomic and demographic statistics. The dataset consists of 31 columns describing the houses the Gorkha earthquake(Nepal) hit. Each building in this dataset is given a damage grade ranging from one to five [17][18].

4 Observation

4.1 Testing Metrics and Environment

The models implemented in the research are classification and regression models and therefore are evaluated with the help of general evaluation metrics. Assessing the performance of a model is crucial for determining its effectiveness and robustness. In machine learning, a metric is a numerical value used to determine how well a model performs. Metrics compare a machine learning model's predictions to the actual results and offer a measurable approach to evaluate the model's recall, accuracy, and precision. Several metrics are employed depending on the type of machine learning problems, such as classification or regression, and the precise objectives of the investigation. The damage grade is an ordinal variable that allows the implementation of ordinal regression models for type. Train accuracy, test Accuracy, and mean squared error of all models are calculated to analyze the models' performance to predict the models' damage grade. Accuracy is one of the most fundamental and popular measures for assessing classification models. It is described as the proportion of correctly categorized cases to all instances. The accuracy metric indicates how well a classification model is performing overall. MSE measures the average squared difference between the target and the projected values. MSE is frequently used to assess regression models since it penalizes significant errors more severely than MAE [19][20][21].

4.2 Accuracy

There are various reasons why the researchers chose this machine learning model over others for this sort of challenge, including the fact that the output was continuous. However, it is crucial to remember that the choice of a machine learning model is contingent on the specific issue and data at hand, and alternative models, such as decision trees or random forests, may also be appropriate for this task. A difference in accuracy can also be seen between the ElasticNet classifier and ElasticNet Regressor because the target variable should be continuous in the case of the ElasticNet Regressor. Still, in the case of our dataset, we can see that our data is persistent. Table 2 and Figure 3 showcase the testing, training accuracy, and mean squared error [22].

Table 2. Causes of Earthquakes

Model	Test	Train	MSE
Bagging Regressor	94.17 %	99.04 %	0.3241
ElasticNet Classifier	92.56 %	92.67 %	0.3663

AdaBoostRegressor	91.86 %	91.96 %	0.3832
Random Forest Regressor	89.76 %	89.87 %	0.4297
DecisionTree Regression	89.3 %	99.94 %	0.4392
Bagging Classifier	89.17 %	99.04 %	0.3241
Linear SVC	87.95 %	88.07 %	0.3679
ExtraTree Regressor	81.67 %	99.94 %	0.5748
AdaBoost Classifier	79.69 %	79.77 %	0.4694
Orthogonal Matching Pursuit	75.74 %	90.92 %	0.6612
ElasticNetCV Regression	0.3 %	0.29 %	1.3406

AdaBoost Regressor works by iteratively creating and merging base models and modifying the weights of the samples in the training data at each iteration to prioritize instances misclassified by the current ensemble of base models, which can be used when the simple base models need to give more accuracy and as the dataset used by the researchers is large. A bit imbalanced, and the ADA boost model works well on datasets like these. ElasticNetCV Regression shows the least accuracy due to the Non-linear relationship between the data points; ElasticNetCV is based on linear regression, which implies a linear connection between features and the target variable. If the relation is non-linear, ElasticNetCV may underperform, and other models, such as non-linear regression or decision trees, as seen in Fig 3 and Figure 4.

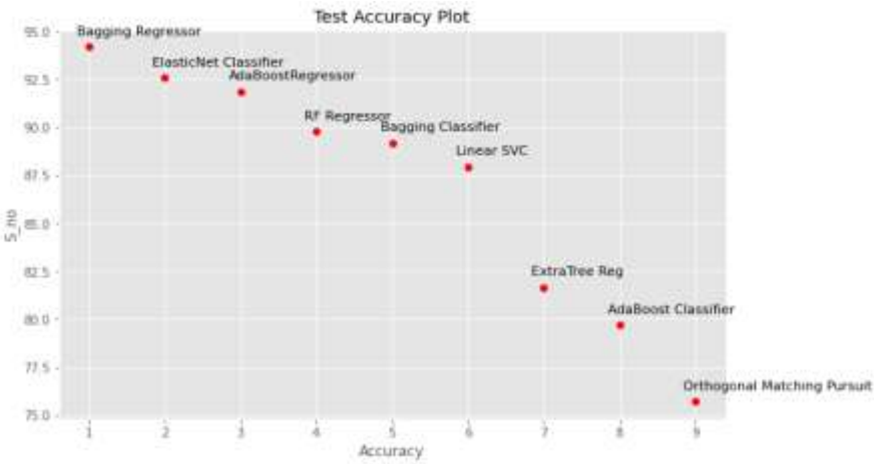

Fig. 3. Test accuracy of Models

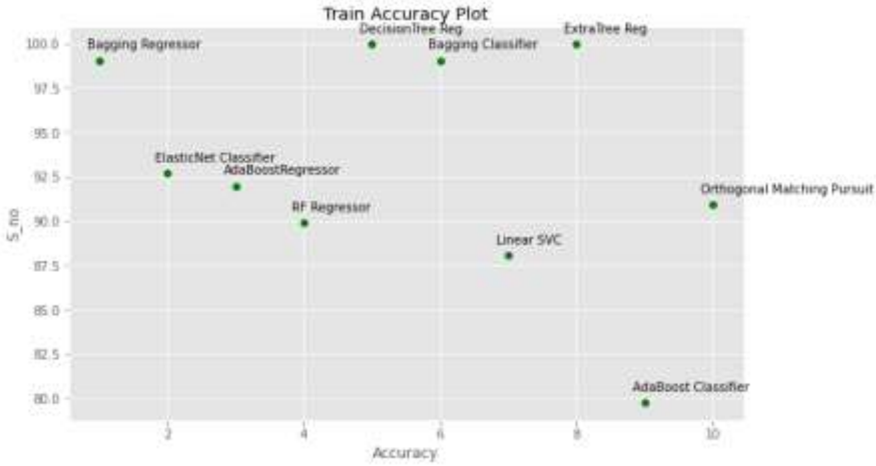

Fig. 4. Accuracy of Models

Bagging Regressor is a machine-learning technique that boosts the performance of a single regression model by blending predictions from various models. Bagging refers to Bootstrapped Aggregating, which involves repeatedly sampling the training data with replacement to train many base models. However, it is essential to acknowledge that the effectiveness of the Bagging Regressor and other machine learning models depends on the specific problem, and it may only sometimes provide the best results. Hence, it is necessary to consider additional models and methods and conduct proper model evaluation and selection to achieve optimal outcomes. It is worth noting that using Bagging Regressor in conjunction with other models can produce more reliable results based on the particular situation. The working of the Bagging Regressor can be seen in Figure 5.

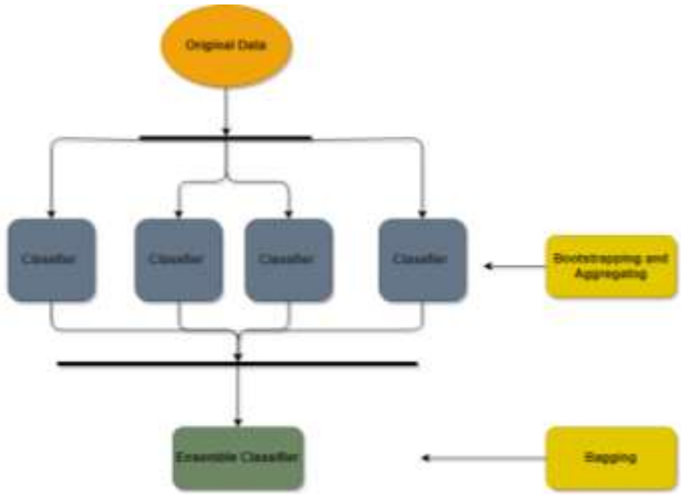

Fig. 5. Bagging Classifier

4 Results:

This work evaluated the efficacy of several classification and ordinal-regression models for estimating the damage grade of dwellings after an earthquake using the pre-processed Gorkha earthquake dataset using various evaluation criteria. The models included Bagging regressor, Linear SVC, ExtraTree Classifier, AdaBoost Classifier, Orthogonal Matching Pursuit, ElasticNetCV Regression, Bagging Regressor, ElasticNet Classifier, Ada Boost Classifier, Random Forest Classifier, Decision Tree Classifier, Bagging Classifier, and Ada Boost Classifier. Accuracy and Mean Squared Error (MSE) measures assessed the models' performance. All models were trained using the same dataset and conditions on the Google Collaboratory research platform. The training and testing sets were each given 80% and 20% of the total dataset, respectively. The test set evaluates the model's performance after the training set has been used to fit the model. In this study, a thorough comparative analysis was also carried out. The performance of various models is evaluated and compared as part of a comparative analysis of machine learning models. This analysis is used to identify the most precise, effective, and appropriate model for a given application. The strengths and limitations of each model can be determined by contrasting the results of many models, which improves decision-making and enables the selection of the model most suited for the task. \The bagging regressor had a training accuracy of 99.04 percent, a testing accuracy of 94.17 percent, and a mean squared error of 0.3241. It is followed by the ElasticNet classifier had a training and testing accuracy of 92.67 percent and 92.56

percent, respectively, and a mean squared error of 0.3663. AdaBoost regressor had a training accuracy of 91.96 percent, testing accuracy of 91.86 percent, and mean squared error of 0.3832. It is followed by a random forest classifier with a mean squared error of 0.4397 and training and testing accuracy of 89.87 percent and 89.76 percent, respectively. Decision tree regression has a mean squared error of 0.4392, train accuracy of 99.94 percent, and testing accuracy of 89.3 percent. An MSE of 0.3241 and train and testing accuracy of 99.04 and 89.17 percent, respectively, were recorded for the bagging classifier. Linear SVC has a mean squared error of 0.3637, train accuracy of 88.07 percent, and testing accuracy of 87.95 percent. The extra tree regressor followed a training and testing accuracy of 99.94 percent and 81.67 percent, respectively, and a mean squared error of 0.5748. AdaBoost classifier had a training accuracy of 79.77 percent, testing accuracy of 79.69 percent, and mean squared error of 0.4694. Orthogonal matching pursuit has a mean squared error of 0.6612, a training accuracy of 90.92 percent, and a testing accuracy of 75.74 percent. Finally, An MSE of 1.3406 and train and testing accuracy of 0.29 and 0.3 percent were recorded for the ElasticNetCV regressor. The best model fit to our data set is the bagging regressor. The plot can be seen in Figure 6.

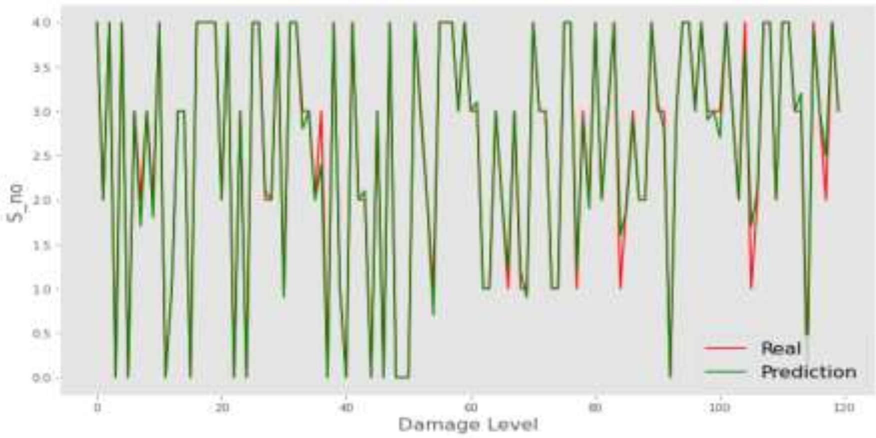

Fig. 6. Plot of Bagging Classifier

5 Conclusion

In conclusion, the study used a pre-processed dataset and a variety of machine-learning models to estimate the degree of damage to homes following the Gorkha earthquake. Bagging Classifier, Linear SVC, ExtraTree Classifier, AdaBoost Classifier, Orthogonal Matching Pursuit, ElasticNetCV Regression, Bagging Regressor, ElasticNet Classifier, Ada Boost Classifier, Random Forest Classifier, Decision Tree Classifier, and Ada Boost Classifier were some of the models tested in this study. According to the data, the Bagging Regressor outperformed the other examined models, having the highest test accuracy (94.14%). The study also discovered that the ordinal regression models of the Bagging Regressor were more accurate in predicting the degree of damage to houses after an earthquake than the classification models. This shows that employing ordinal regression models and considering the damage grade's ordinal nature can improve prediction outcomes. The findings show that machine learning systems can accurately anticipate the degree of earthquake damage to homes. The models' robustness and accuracy show that these algorithms can efficiently learn from input information and predict the damage grade accurately. The outcomes of this study demonstrate how crucial feature selection and pre-processing are in machine learning models. The study offers valuable insights into how well various machine learning models perform in estimating the degree of damage to houses as well. Future research should examine the effects of multiple input variables and assess how deep learning models predict the degree of damage to dwellings following an earthquake. The findings of this study serve as a starting point for additional research in this area and can be used by disaster management organizations to improve earthquake preparedness and response.

References

1. Asim, K. M., Martínez-Álvarez, F., Basit, A., & Iqbal, T. (2017). Earthquake magnitude prediction in Hindukush region using machine learning techniques. Natural Hazards, 85, 471-486.
2. Kanamori, H. (2003). Earthquake prediction: An overview.
3. 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.
4. Li, R., Lu, X., Li, S., Yang, H., Qiu, J., & Zhang, L. (2020, July). DLEP: A deep learning model for earthquake prediction. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE
5. Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. Earthquake Spectra, 36(1), 183-208.
6. Xie, Y., Ebad Sichani, M., Padgett, J. E., & DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Earthquake Spectra, 36(4), 1769-1801.
7. Xiong, P., Tong, L., Zhang, K., Shen, X., Battiston, R., Ouzounov, D., ... & Zhou, H. (2021). Towards advancing earthquake forecasting by machine learning of satellite data. Science of The Total Environment, 771, 145256.
8. Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. Geophysical Research Letters, 47(1), e2019GL085976.
9. Mallouhy, R., Abou Jaoude, C., Guyeux, C., & Makhoul, A. (2019, December). Major earthquake event prediction using various machine learning algorithms. In 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) (pp. 1-7). IEEE.
10. Deepta, E., Juyal, N., & Sharma, S. (2022). Wind Speed Prediction in the Region of India Using

Artificial Intelligence. In Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2021 (pp. 723-731). Singapore: Springer Nature Singapore.
11. Abhijith, H., Gupta, D., & Sharma, S. (2021). Blockchain Security Solutions for IoT and Big Data. Blockchain Applications for Secure IoT Frameworks: Technologies Shaping the Future, 1, 248.
12. A. Saxena, R. Chauhan, D. Chauhan, and G. Aggarwal, "Time Series Prediction of Sea Surface Temperature Using LSTM," 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2023, pp. 458-461, doi: 10.1109/Confluence56041.2023.10048855.
13. Mall, Pawan Kumar, et al. "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities." Healthcare Analytics (2023): 100216.
14. Narayan, Vipul, et al. "Severity of Lumpy Disease detection based on Deep Learning Technique." 2023 International Conference on Disruptive Technologies (ICDT). IEEE, 2023.
15. Narayan, Vipul, et al. "Enhance-Net: An Approach to Boost the Performance of Deep Learning Model Based on Real-Time Medical Images." Journal of Sensors 2023 (2023).
16. Sawhney, Rahul, et al. "A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease." Decision Analytics Journal 6 (2023): 100169.
17. Narayan, Vipul, et al. "FuzzyNet: Medical Image Classification based on GLCM Texture Feature." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.
18. Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.
19. Srivastava, Swapnita, et al. "An Ensemble Learning Approach For Chronic Kidney Disease Classification.
20. Mall, Pawan Kumar, et al. "Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning Technique." Journal of Pharmaceutical Negative Results (2022): 4784-4792.
21. Narayan, Vipul, and A. K. Daniel. "A novel approach for cluster head selection using trust function in WSN." Scalable Computing: Practice and Experience 22.1 (2021): 1-13.
22. Narayan, Vipul, A. K. Daniel, and Ashok Kumar Rai. "Energy efficient two tier cluster based protocol for wireless sensor network." 2020 international conference on electrical and electronics engineering (ICE3). IEEE, 2020.