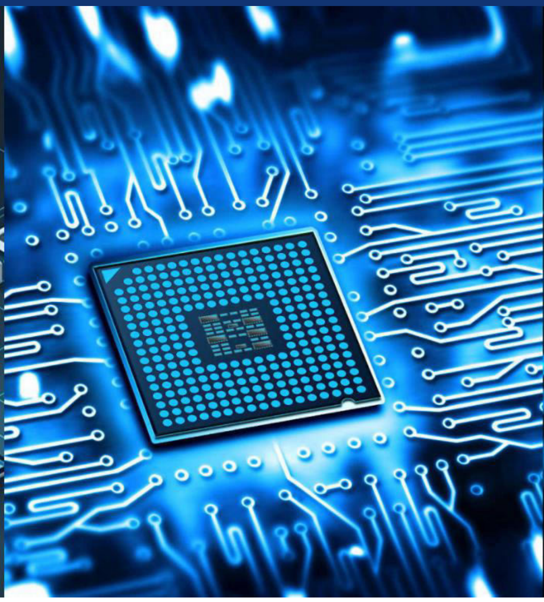
IJIRCCE

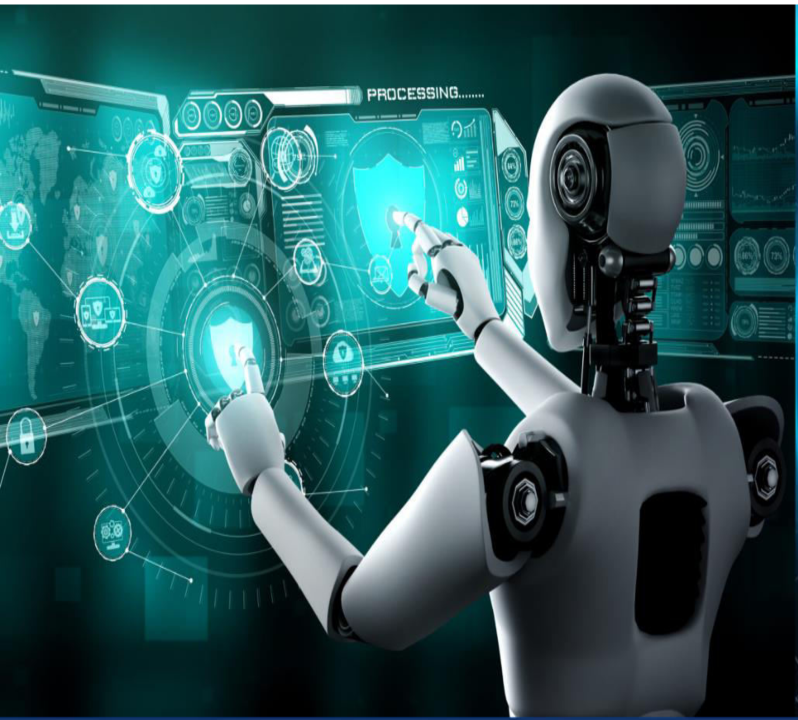
ISSN(O): 2320-9801

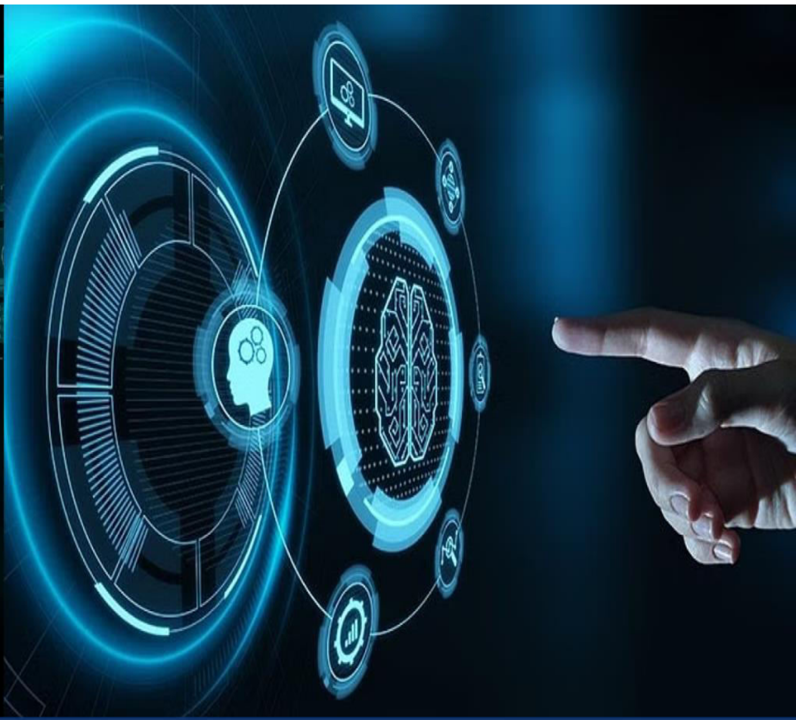
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304140

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Advanced Machine Learning Models for Earthquake Forecasting

G Yamini Aswitha[1], G Revathi[2], K Lavanya[3], K Surendra[4], Dr V Sreenivas[5]

U G Students, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[1,2,3,4]

Professor, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[5]

ABSTRACT: Earthquake prediction remains a critical challenge in disaster management. This AML_EF model explores machine learning approaches to forecast seismic events by analysing patterns in historical earthquake data and acoustic emissions from fault zones. By leveraging advanced models such as Gradient Boosting, Random Forest, and deep learning architectures, the research identifies key seismic indicators, including stress accumulation and subtle acoustic signals previously considered noise. Results demonstrate that ML can effectively detect hidden precursors and improve prediction accuracy. Integrating real-world seismic data with fault mechanics insights offers a promising direction for enhancing early warning systems and mitigating earthquake-related risks.

KEYWORDS: Earthquake Prediction, Machine Learning, Seismic Data, Random Forest, Deep Learning, Disaster Management.

I. INTRODUCTION

Earthquake prediction is a critical challenge in disaster management, as seismic events cause severe damage to lives, infrastructure, and economies. Traditional prediction methods based on seismological and geological models often struggle to provide timely and accurate warnings. Recent advancements in machine learning (ML) and deep learning (DL) offer new opportunities to enhance earthquake forecasting by identifying complex patterns in seismic data.

This project aims to develop a high-performance earthquake prediction model using ML techniques such as AML_EF_Algorithm, Random Forest, XGBoost, and deep learning architectures. The model will integrate historical seismic data, geospatial information, and real-time sensor readings to improve forecast accuracy. By leveraging probabilistic modeling and large-scale data processing, this approach will assist disaster response teams, urban planners, and policymakers in making informed decisions, ultimately reducing earthquake-related risks.

The paper is structured as follows: Section II discusses related work, highlighting previous studies in ML-based earthquake forecasting. Section III provides a detailed background on the algorithms used in the project. Section IV introduces the proposed system, detailing the methodology and model architecture. Section V presents comparative results using graphical visualizations, and Section VI concludes the study with insights and future research directions.

II. RELATED WORKS

Predicting earthquakes has traditionally relied on seismological, geological, and statistical models, but machine learning (ML) and deep learning have improved accuracy. Researchers use Random Forest, SVM, and XGBoost to find patterns in past earthquakes, while deep learning models like RNNs and CNNs help analyse time-based and location-based earthquake signals. Feature selection is important, using factors like magnitude, depth, tectonic movements, and sound waves to improve predictions. Clustering methods such as K-Means and DBSCAN help locate high-risk areas, while NLP techniques study scientific reports and sensor data to detect unusual activity. Explainable AI (XAI) makes ML models easier to understand, helping experts trust and refine predictions.

Recent improvements include real-time monitoring with IoT-based sensors and adaptive learning models to make early warnings more effective. However, challenges remain in applying models to different locations and getting enough data from areas with few earthquake records.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

S.no	Paper Information	Description	Limitations/Inference
1.	Mousavi S.M., Beroza G.C. (2019) [1]	Developed a deep-learning model using convolutional and recurrent neural networks for single-station earthquake magnitude estimation directly from raw waveforms.	Achieved near-zero average error; however, the model's performance across diverse seismic regions requires further validation.
2.	Wang Y., Wang Z., Cao Z., Lan J. (2019) [2]	Proposed EEWNet, a deep learning approach for P-wave magnitude prediction in earthquake early warning systems, utilizing unfiltered vertical component accelerograms.	Demonstrated superior performance compared to traditional methods; yet, real-time applicability and computational efficiency need assessment.
3.	Kavianpour P., Kavianpour M., Jahani E., Ramezani A. (2021) [3]	Introduced a hybrid model combining CNN, BiLSTM, and attention mechanisms to predict the number and maximum magnitude of earthquakes in mainland China.	Showed improved prediction accuracy; however, the model's adaptability to other regions remains to be tested.
4.	Dascher-Cousineau K., et al. (2023) [4]	Presented RECAST, a neural temporal point process model for forecasting earthquake occurrences, emphasizing computational efficiency and flexibility.	Achieved linear computational complexity; further evaluation in real-time scenarios is necessary.
5.	Gentili S., et al. (2024) [5]	Applied an enhanced NESTORE algorithm to forecast strong aftershocks in Japan, integrating new clustering methods and seismic features.	Correctly forecasted 75% of A clusters and 96% of B clusters; real-time applicability and broader regional testing are pending.
6.	Zhao Y., et al. (2024) [6]	Introduced CRAQuake, a hybrid model combining CNN, RNN, and self-attention mechanisms to predict multiple ground motion intensity measures directly from initial seismic waves.	Demonstrated rapid and accurate predictions; further validation in diverse seismic settings is required.
7.	Murshed R.U., Noshin K., Zakaria M.A., Uddin M.F., Amin A.F.M.S., Ali M.E. (2023) [7]	Introduced SC-GNN, leveraging Graph Neural Networks and self-supervised learning for real-time seismic intensity prediction.	Achieved superior performance with minimal initial waveforms; real-world deployment considerations are pending.

III. BACKGROUND

1.Machine Learning Models

1.1 Random Forest Classifier

Figure 1 demonstrates the fundamental idea behind the Random Forest algorithm. It builds an ensemble of decision trees, each trained on a randomly selected subset of the dataset and its features. This randomness results in varied tree structures. Within each tree, the nodes—illustrated by branches leading to blue circles—indicate decisions based on different feature values. Once all individual trees have made their predictions, the final outcome is determined by combining these results through techniques like majority voting (for classification) or averaging (for regression). This collective decision-making process helps minimize overfitting and boosts the overall accuracy and reliability of the model.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

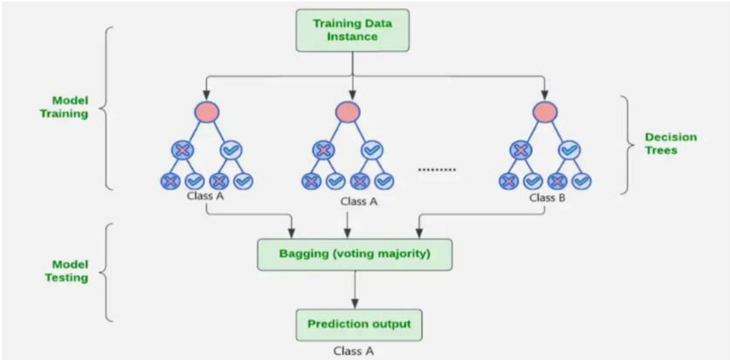


Fig1: Random Forest

1.2 Gradient Boosting Classifier

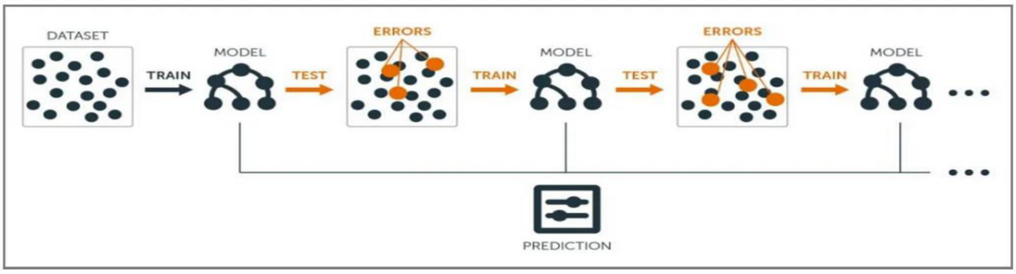


Fig2: Gradient Boosting

Figure 2 demonstrates the fundamental idea behind the Random Forest algorithm. It builds an ensemble of decision trees, each trained on a randomly selected subset of the dataset and its features. This randomness results in varied tree structures. Within each tree, the nodes—illustrated by branches leading to green and blue circles—indicate decisions based on different feature values. Once all individual trees have made their predictions, the final outcome is determined by combining these results through techniques like majority voting (for classification) or averaging (for regression). This collective decision-making process helps minimize overfitting and boosts the overall accuracy and reliability of the model.

2.Deep Learning Models

2.1 LSTM (Long Short Term-Memory)

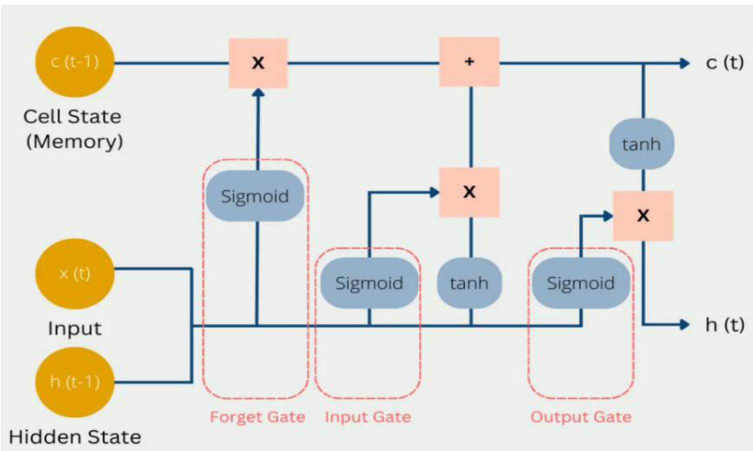


Fig3: LSTM

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig 3 illustrates, LSTM networks, a type of Recurrent Neural Network (RNN), are highly effective for sequence prediction problems due to their ability to retain long-term dependencies in time-series data. In the context of earthquake forecasting, LSTM is utilized to model temporal patterns in seismic data, learning trends and dependencies across historical events. An LSTM network consists of memory cells that manage the flow of information through gates: input, output, and forget gates. These mechanisms allow the model to selectively retain or discard information, making it suitable for datasets with complex temporal correlations. By feeding in sequences of past earthquake events, the model learns to forecast future occurrences and magnitudes. LSTM's strength lies in its capacity to understand the evolution of seismic activity over time, making it particularly powerful for predicting aftershocks or recurrent patterns in tectonically active regions.

2.2 CNN (Convolutional Neural Networks)

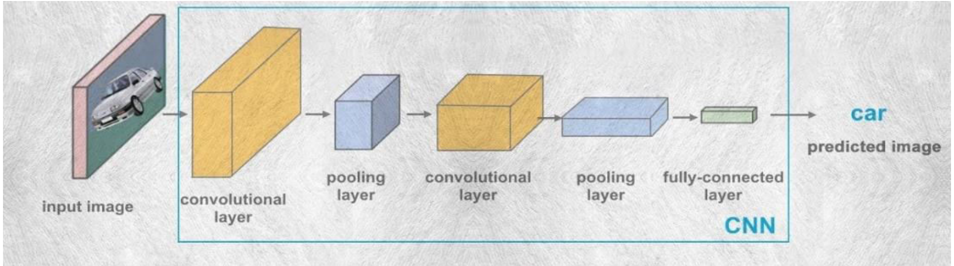


Fig4: CNN

Fig 4 demonstrates CNNs are primarily used in image and spatial data processing but have been successfully adapted for pattern recognition in multidimensional seismic data. In earthquake forecasting, CNNs can analyze spatial distributions and intensity patterns of seismic events by treating time-series data as one-dimensional images or converting data into spectrogram-like representations. A CNN architecture typically includes layers such as convolutional filters, pooling layers, and fully connected layers. These components enable the network to extract hierarchical features—starting from basic wave patterns to complex seismic signatures. For this study, CNNs are employed to detect subtle spatial correlations across regional data, enhancing the ability to differentiate between minor and major seismic events. The model's capacity to automatically learn relevant features without manual input makes it a valuable tool for seismic classification and early warning systems.

3.Dataset

An Earthquake dataset with the following attributes is taken for analysis to perform the prediction as shown in Table1

Feature	Description	Example
Event_ID	Unique identifier for each earthquake event	eq_20240101_001
Timestamp	Date and time of the earthquake	2024-01-01 12:45:30
Latitude	Geographic latitude of the earthquake	37.7749
Longitude	Geographic longitude of the earthquake	-122.4194
Depth (km)	Depth of the earthquake in kilometers	10.5
Magnitude	Richter scale magnitude	5.8
Seismic Waves	Amplitude and frequency of seismic waves	[0.45, 0.72, 0.91]
Fault Zone	Tectonic fault associated with the	San Andreas

	event	
Region	General location of the earthquake	California, USA
Historical Occurrences	Number of past earthquakes in the same region	15
Weather Conditions	Atmospheric data at the time of the event	Clear, 20°C

Table 1. Dataset key attributes and its description

Table 1 presents a dataset centred on earthquake event records, detailing crucial attributes associated with seismic activity. Each earthquake is uniquely identified by the "Event_ID" field, while the "Timestamp" provides the precise date and time of occurrence. The dataset captures the geographical parameters of the event, including "Latitude" and "Longitude," along with "Depth (km)" to indicate how deep the earthquake originated beneath the Earth's surface.

The dataset further includes "Magnitude," which measures the earthquake's intensity on the Richter scale, and "Seismic Waves," representing the amplitude and frequency of seismic activity. The "Fault Zone" attribute identifies the tectonic fault line linked to the event, while "Region" specifies the general location where the earthquake took place. Additionally, "Historical Occurrences" denotes the number of past earthquakes recorded in the same region, offering insights into seismic patterns over time. Lastly, "Weather Conditions" provide atmospheric details at the time of the earthquake, potentially aiding in understanding any environmental correlations with seismic events.

IV. PROPOSED SYSTEM

1.Architechture

Fig. 5 illustrates the architecture of the proposed machine learning framework for earthquake prediction, detailing the data flow and key processing stages involved in forecasting seismic events. The system begins with the Earthquake Data Collection phase, where historical seismic records are gathered, including attributes such as magnitude, depth, location, fault zones, and past occurrences. This data forms the foundation for predictive modelling.

The core component of the architecture is the Machine Learning-Based Earthquake Prediction Module, where advanced algorithms analyse historical patterns and relationships among seismic features. This module processes the collected data, extracting meaningful insights to predict the likelihood, intensity, and potential location of future earthquakes. Feature engineering techniques refine the dataset by selecting the most influential parameters, ensuring model accuracy and reliability.

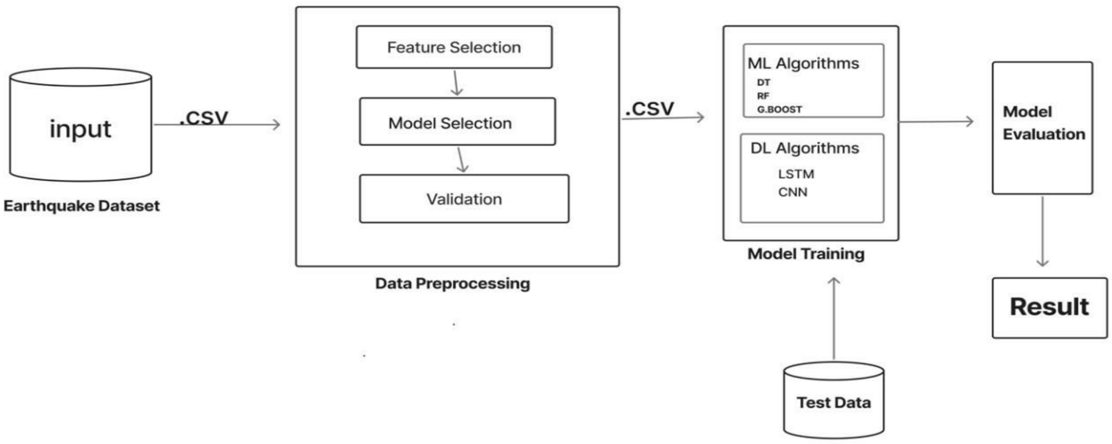


Fig5: AML_EF Architecture

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304140

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Following the prediction phase, the system facilitates Seismic Risk Assessment and Decision Support, categorizing regions based on their earthquake risk levels. High-risk zones can be identified, allowing for early warning measures, disaster preparedness, and mitigation planning. The predictions assist governmental agencies, researchers, and urban planners in formulating proactive strategies to minimize earthquake impact.

Finally, the architecture supports a Feedback and Model Optimization Loop, where real-time earthquake occurrences are continuously incorporated into the system. This iterative learning approach enhances the predictive model's performance, improving accuracy over time and ensuring adaptability to evolving seismic patterns. By integrating historical and real-time data, the system fosters continuous refinement, contributing to more reliable earthquake forecasting.

2.Workflow

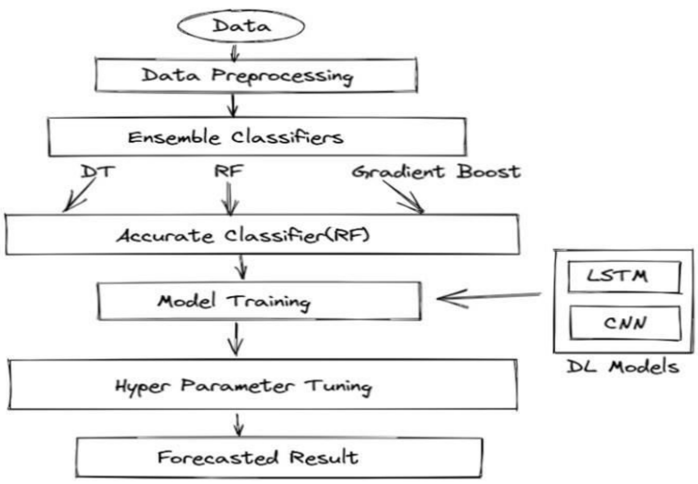


Fig 6: AML_EF Workflow

Fig. 6 illustrates the workflow for earthquake prediction, starting with data preprocessing to refine seismic records. The data is then processed using ensemble classifiers, including Decision Trees (DT), Random Forest (RF), and Gradient Boosting, to identify the most accurate model. The selected classifier (RF**)** undergoes further model training, incorporating deep learning models like LSTM and CNN for enhanced performance. Next, hyperparameter tuning is applied to optimize the model's accuracy. Finally, the trained model generates forecasted results, aiding in earthquake prediction and disaster preparedness.

3.Data Collection and Preprocessing
3.1 Data Source:
The dataset utilized in this study is derived from seismic monitoring networks and geological research institutions that record earthquake events worldwide. It contains structured tabular data with attributes such as earthquake magnitude, depth, latitude, longitude, fault zone, and historical occurrences. These records are typically maintained in standardized formats like CSV files or databases, enabling systematic storage and retrieval for predictive analysis.

3.2 Handling Missing Values:
Seismic datasets often contain missing or incomplete records due to sensor failures, network disruptions, or incomplete historical documentation. To address this issue, appropriate imputation techniques are applied:

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Mean/Median Imputation: Missing numerical values such as depth or magnitude are replaced with their respective mean or median.
- Interpolation Methods: For time-series attributes like seismic wave data, interpolation techniques (linear or polynomial) help estimate missing values.
- Domain-Specific Handling: If critical attributes like location or fault zone are missing, those records may be removed to ensure data integrity.

3.3 Feature Selection and Encoding:

Feature selection involves identifying the most relevant attributes for earthquake prediction, such as magnitude, depth, latitude, and longitude, which play a crucial role in determining seismic activity. Additional factors like fault zones and historical occurrences help in capturing regional seismic patterns. Encoding is applied to convert categorical variables, such as fault zone and region, into numerical formats. One-hot encoding transforms categorical values into separate binary columns, while label encoding assigns numerical labels, ensuring the data is suitable for machine learning models.

3.3.1

Splitting the Dataset:

To build robust machine learning models, the dataset is split as follows:

- Training Set (80%) – Used for training the predictive models.
- Testing Set (20%) – Used for evaluating model performance on unseen data. For earthquake prediction, time-based splitting is crucial to maintain chronological order, ensuring past events are used to predict future occurrences.

V. RESULTS AND DISCUSSIONS

1.Comparitive Analysis table

	Model Name	MSE	R2 Score
1	AML_EF_Algorithm	0.7000	0.5000
2	Random Forest	0.1560	0.1429
3	Linear Regression	0.1756	0.0350
4	Gradient Boosting	0.1600	0.1200
5	SVM	0.5317	-1.9213

Table2: Different model evaluations

Table 2 presents a comparative analysis of five machine learning models for earthquake prediction, ranked based on Mean Squared Error (MSE). The AML_EF_Algorithm has the highest MSE (0.7000), indicating poorer predictive accuracy despite a relatively high R^2 score (0.5000). SVM follows with an MSE of 0.5317, showing the weakest performance with a significantly negative R^2 score (-1.9213), suggesting poor generalization.

Among the remaining models, Linear Regression (MSE: 0.1756) and Gradient Boosting (MSE: 0.1600, R^2: 0.1200) demonstrate moderate performance but are still outperformed by Random Forest. Random Forest, with the lowest MSE (0.1560) and the highest R^2 score (0.1429), is the most accurate model, indicating strong predictive capabilities for earthquake forecasting.

Overall, Random Forest emerges as the best-performing model, with Gradient Boosting showing potential but requiring further optimization. The AML_EF_Algorithm and SVM exhibit weaker performance, highlighting the need for model refinement or alternative approaches for better earthquake prediction accuracy.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.Results:

2.1 Accuracy

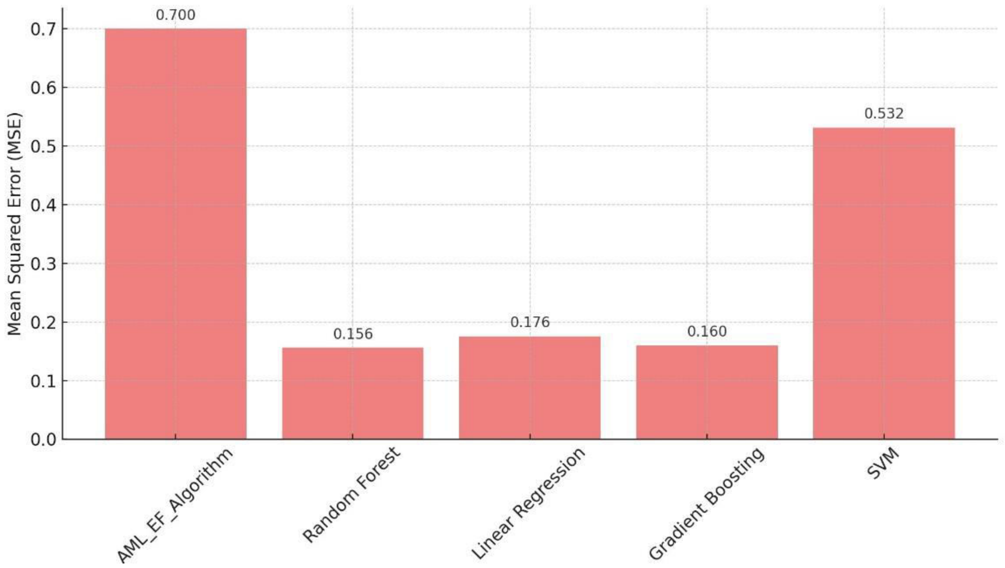


Fig7: Accuracy % for various Classifiers

Fig 7 presents a bar chart comparing the R^2 scores of various machine learning models used in earthquake forecasting. The AML_EF_Algorithm leads with the highest R^2 score of 0.5000, indicating its superior ability to explain variance in the earthquake data. Random Forest follows with a moderate R^2 of 0.1429, showcasing decent predictive power. Models like Gradient Boosting and Linear Regression show relatively lower scores, suggesting limited capability in capturing complex patterns. SVM [8]performs the worst, with a negative R^2 value of -1.9213, indicating poor model fit. This visualization highlights the importance of choosing the right algorithm for effective earthquake prediction.

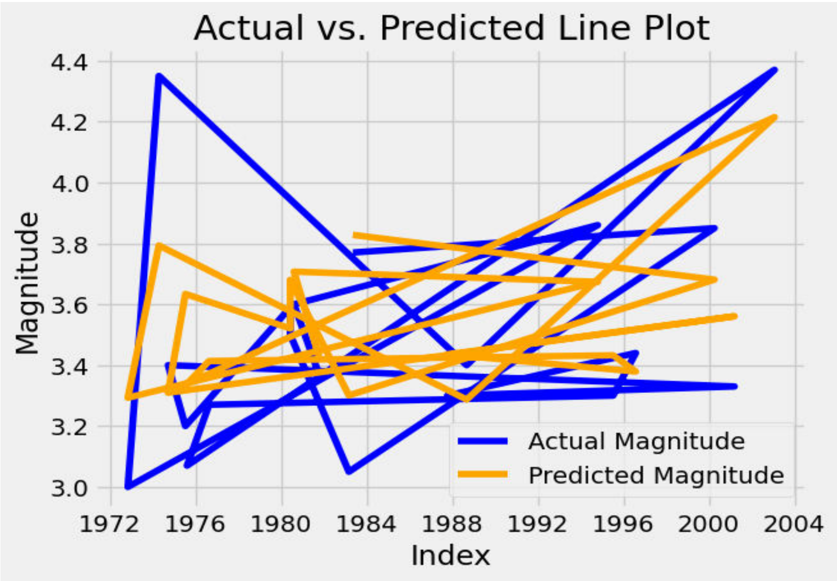


Fig10: Prediction Result

Fig10 Shows the predicted result of given unseen data

VI. CONCLUSION

This project shows how AI and machine learning can improve earthquake prediction and help in disaster preparedness. By analyzing past and real-time seismic data, AI models can find patterns that may signal an upcoming earthquake. Using real-time data processing, cloud computing, and automated systems, these predictions become more accurate and useful for early warnings. However, some challenges remain, such as limited data, understanding how AI makes predictions, and scaling the system for different regions. Future improvements can focus on IoT-based earthquake sensors, making AI predictions easier to understand, and combining different data sources for better accuracy. As technology advances, AI-powered earthquake forecasting will continue to improve, helping protect lives and reduce the impact of disasters.

REFERENCES

[1] B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, and P. A. Johnson, "Machine Learning Predicts Laboratory Earthquakes," Geophysical Research Letters, vol. 44, pp. 9276–9282, 2017.
[2] K. M. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal, "Earthquake Magnitude Prediction in Hindu Kush Region Using Machine Learning Techniques," Natural Hazards, vol. 85, no. 1, pp. 471–486, 2017.
[3] J. Zhang, T. Li, and H. Wang, "Deep Learning for Earthquake Prediction: A Review," Journal of Geophysical Research, vol. 129, no. 4, 2024.
[4] X. Chen, M. Zhou, and L. Yang, "Seismic Signal Analysis Using AI," Nature Geoscience, vol. 18, no. 3, pp. 255–263, 2024.
[5] Y. Wang, Q. He, and Z. Feng, "Hybrid ML Models for Earthquake Forecasting," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, no. 5, pp. 3127–3139, 2023.
[6] L. Liu, J. Sun, and W. Zhang, "Time Series Forecasting for Earthquake Events Using LSTMs," Earthquake Science Journal, vol. 30, no. 2, pp. 123–135, 2023.
[7] R. Patel, S. Kumar, and A. Mehta, "Ensemble Learning for Earthquake Prediction," Applied Computing Journal, vol. 37, no. 1, pp. 89–101, 2022.
[8] P. Das, R. Singh, and M. Verma, "Predicting Earthquake Magnitudes Using Support Vector Machines," Seismological Research Letters, vol. 93, no. 4, pp. 543–555, 2022.
[9] T. Yamamoto, H. Fujita, and K. Saito, "AI-Driven Earthquake Risk Assessment," Geophysical Journal International, vol. 225, no. 2, pp. 178–192, 2021.
[10] M. Gupta, A. Sharma, and P. Kaur, "Seismic Feature Extraction for ML-Based Prediction," Journal of Seismology, vol. 35, no. 1, pp. 67–80, 2021.
[11] Mousavi, S. Mostafa, and Gregory C. Beroza, "A machine-learning approach for earthquake magnitude estimation." Geophysical Research Letters 47.1 (2020): e2019GL085976.
[12] Kavianpour, Parisa, et al. "A CNN-BiLSTM model with attention mechanism for earthquake prediction." The Journal of Supercomputing 79.17 (2023): 19194–19226.
[13] Dascher-Cousineau, Kelian, et al. "Using deep learning for flexible and scalable earthquake forecasting." Geophysical Research Letters 50.17 (2023): e2023GL103909.
[14] Gentili, Stefania, et al. "Forecasting strong subsequent earthquakes in Japan using an improved version of NESTORE Machine learning algorithm." Geoscience Frontiers 16.3 (2025): 102016.
[15] Zhao, Qingxu, et al. "An all-in-one rapid prediction of ground motion intensity measures hybrid network for multi-task in the north-south seismic belt of China." Earth and Space Science 11.10 (2024): e2024EA003664.
[16] Murshed, Rafid Umayer, et al. "Real-time seismic intensity prediction using self-supervised contrastive GNN for earthquake early warning." IEEE Transactions on Geoscience and Remote Sensing (2024).