Deep learning models for the prediction of earthquake magnitudes

MSc Research Project
Cloud Computing

Yaseen Ali Khan
Student ID: 22236252

School of Computing
National College of Ireland

Supervisor: Dr Giovani Estrada

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Yaseen Ali Khan
Student ID:	22236252
Programme:	Cloud Computing
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr Giovani Estrada
Submission Due Date:	24/04/2025
Project Title:	Deep learning models for the prediction of earthquake magnitudes
Word Count:	7718
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Yaseen Ali Khan
Date:	24 April 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deep learning models for the prediction of earthquake magnitudes

Yaseen Ali Khan
22236252

Abstract

Earthquake prediction is a critical yet complex task due to the stochastic nature of seismic activities and the multitude of factors influencing tectonic movements. This study aims to leverage advanced machine learning and deep learning techniques to predict earthquake magnitudes based on spatial, temporal, and geophysical parameters. Multiple baseline models were implemented and evaluated, including Random Forest Regressor, Feedforward Neural Network, SVR, Extra Trees Regressor, and Gradient Boosting Regressor. Hyperparameter tuning using RandomizedSearchCV was employed. As part of the research, a number of deep learning architectures were studied in detail, including ones with GRU, Conv1D, Bidirectional LSTM, a hybrid Bidirectional GRU + Conv1D layers. Performance was assessed using MAE, MSE, RMSE, and R^2 score. Over the course of this document, it will be shown that Gradient Boosting Regressor emerged as the best-performing model with the lowest MAE (0.2817), lowest MSE (0.1514), lowest RMSE (0.3891), and the highest R^2 Score (0.1734), demonstrating its robustness and reliability. The work did not find comparable results with any deep learning technique. However, this study underscores the potential of ensemble learning models, particularly Gradient Boosting, instead of much-hyped deep learning techniques.

Keywords— Earthquake Prediction, Machine Learning, Deep Learning, Random Forest Regressor, GRU (Gated Recurrent Unit)

1 Introduction

1.1 Background

Earthquake prediction refers to the scientific effort to determine the likelihood, location, and magnitude of a future seismic event before it occurs (Bhardwaj et al.; 2021). Unlike general seismic hazard assessments, which estimate long-term probabilities over regions, earthquake prediction focuses on pinpointing specific events in terms of time, place, and intensity. The process involves analyzing various geological, geophysical, and geochemical signals such as ground deformations, foreshocks, gas emissions, changes in groundwater levels, and historical seismic data (Martinelli; 2020). Despite decades of research, accurately predicting earthquakes remains a significant challenge due to the complexity and chaotic nature of tectonic processes. However, advances in data science AI and ML have opened new avenues for identifying patterns in large seismic datasets, leading to more promising results in forecasting earthquake magnitude or probability with greater precision. Predictive models now integrate real-time data streams from seismographs, satellites, and sensor networks to improve early warning systems (Esposito et al.; 2022). Although true short-term earthquake prediction is still not fully reliable or universally accepted in the scientific community, these models contribute significantly to risk mitigation by improving preparedness, evacuation planning, and infrastructure resilience. Earthquake prediction thus remains a multidisciplinary challenge aimed at minimizing the devastating impact of seismic events on human lives and property through advanced computational and scientific strategies. However, recent developments in data science, artificial intelligence, and ML have provided a different perspective on identifying patterns in big seismic datasets, often resulting in more promising outcomes in predicting the magnitude or probability of the earthquake with relatively higher accuracy. Real-time data streams from seismographs, satellites, and sensor networks are now integrated into predictive models that improve early warning systems. Although nothing quite like true short-term prediction exists, and even models are still not fully reliable or universally accepted by the scientific community, it goes without saying that they help mitigate risk, made all the more important given that a huge portion of population density is concentrated near tectonic activity. As such, earthquake prediction is thus a matter of interdisciplinary efforts to reduce human and material damage caused by seismic phenomena using computational and scientific tools.

1.2 Aim of the Study

The aim of this study is to develop and evaluate strong ML and DL models to accurately predict the magnitude of earthquakes using historical seismic data. Leveraging a comprehensive dataset comprising geospatial and temporal features such as latitude, longitude, depth, time, and seismic sources, this research intends to explore various modeling approaches including traditional regressors like Random Forest, Extra Trees, and Gradient Boosting, alongside deep learning models such as Feedforward Neural Networks, GRU, Conv1D, Bidirectional LSTM, and hybrid architectures. The study has been focused on the importance of preprocessing techniques, missing value imputation and feature encoding to have clean and informative inputs. Also hyperparameter optimization using RandomizedSearchCV is combined to increase model performance which is mainly for ensemble-based methods. The aim is not only to determine which model achieves the highest accuracy based on metrics like MAE, MSE, RMSE and R^2 score but also to evaluate the viability of deep neural networks in capturing the nonlinear relationships in earthquake data. Ultimately, this study aspires to contribute to disaster management by providing a reliable computational framework capable of forecasting earthquake magnitudes, which could support early warning systems and risk mitigation strategies in seismically active regions.

1.3 Research Questions

There are few research questions for this research as follows:

1. Which machine learning or deep learning model demonstrates the highest accuracy and reliability in predicting earthquake magnitudes based on historical seismic data?

2. How does hyperparameter tuning using RandomizedSearchCV influence the predictive performance of the Random Forest Regressor in comparison to other models?

3. Can a hybrid Bidirectional GRU and Conv1D model outperform individual models in learning complex patterns from seismic data for more accurate earthquake magnitude forecasting?

1.4 Research Objectives

There are several research objectives of this research as follows:

1. To evaluate and compare the performance of traditional ML and DL models for earthquake magnitude prediction using geophysical features.

2. To optimize the performance of selected models through hyperparameter tuning techniques such as RandomizedSearchCV and analyze the impact on prediction accuracy.

3. To develop a hybrid deep learning architecture that combines sequential (GRU) and spatial (Conv1D) feature learning to improve the accuracy and robustness of the prediction of magnitudes.

1.5 Report Structure

This report is organised into sections that systematically present the research. Section 1 provides an introduction and background on the challenges of earthquake prediction. Section 2 reviews previous studies related to earthquake prediction using traditional statistical approaches, machine learning (ML), and deep learning (DL). Section 3 presents the methodology, from data sources, preprocessing steps, including handling missing values to feature encoding. Section 5 shows implementation details of each model. Section 6 presents results and evaluations. Finally, Section 7 shows some conclusions and future work.

2 Related Work

2.1 Deep Learning in Disaster Recovery

One approach (Pi et al.; 2020) has been proposed to perform real-time disaster impact assessment based on the defined model of ground object detection using aerial imagery and convolutional neural networks (CNNs). The issue of identifying important infrastructural elements (damaged and undamaged buildings, vehicles, vegetation, debris, flooded zones, among others) in huge aerial datasets recorded via UAVs or choppers is tackled, taking into consideration that many times the amount of data is huge for manual and/offline processing. In order to overcome this issue, the authors create eight CNN models based on the YOLO algorithm, which are trained using transfer learning on pre-training weights from COCO and VOC datasets and fine-tuned on a custom in-house dataset called Volan2018. Volan2018 includes 65,580 annotated frames representing eight drone and helicopter videos taken of hurricanes hitting the U.S. between 2017–2018. Our proposed models yield remarkable results (average mean average precision 80.69% and 74.48% for helicopter (high altitude) and drone (low altitude) footage, respectively. It is found that models trained on the same altitudes as test data perform much better and that the VOC-pretrained model using a balanced drone dataset has the best accuracy vs training time trade-off. Although the findings appear promising, the study is limited by the size of the dataset, the reliance on web-mined data, and potential generalisability issues when applied to new disaster scenarios or used in other geographic areas.

Another study by (Munawar et al.; 2021) that examined a flood detection system deployed with UAVs that uses convolutional neural networks (CNN) to quickly identify flood affected areas and damage to local infrastructure, specifically in developing countries where flood prediction systems are often too expensive. In an effort to improve disaster response by examining aerial images taken before and after flood events to determine the characteristics of floods, the article examines a flood-prone area of the Indus River in Pakistan. By resizing and cropping the UAV-captured images, we generate a total of 2,150 image patches for CNN training for the model to detect the changes that are caused by flooding. Although a UAV-based approach provides a practical solution while reducing infrastructure overhead and extensive installation requirements, UAV-CNN combinations have not been implemented in existing flood detection frameworks due to the critical investment required for effective flood prediction. Trained with data until October 2023, the model performs well, registering a 91% accuracy rate in detection of flooded regions. This allows for timely disaster management things that will help protect important infrastructure and support those impacted. In addition, the companion study is limited by the relatively small size of the dataset and geographical coverage, potentially affecting the transferability of the model between flood types and landscapes.

A flood prediction and detection system using the Internet of Things (IoT), Big Data (BD), and Convolutional Deep Neural Networks (CDNN) has been used to improve disaster things and response (Anbarasan et al.; 2020). To handle these challenges in coping with large-scale, complicated flood-style response data and analysing the prediction accuracy of flood occurrence. The study has been proposed a multi-stage method. First it cleans the input flood data using HDFS MapReduce to remove the redundant data. This data will then go through missing value imputation and normalization after which they will be processed into rules using attribute combinations methods. The rules are then used in a CDNN classifier which classifies the results as flood occurrence or no risk of flood occurrence. This technique is compared with conventional techniques which are Artificial Neural Networks (ANN) and Deep Learning Neural Networks (DNN) on the basis of well-known parameters– Accuracy, Precision, Recall, Sensitivity, Specificity, and Fscore. The details of the findings are summarized in their recent paper, titled 'Flood Prediction Using Machine Learning' Author, which cite a need for improved accuracy and robustness in flood prediction using the model. Nonetheless, the scalability and adaptiveness of the proposal to real-time applications and evolving data patterns are not fully discussed; this may hinder effective use in the context of dynamic and large-scale disaster environments during the predictive modeling task.

In a similar way (Jena et al.; 2021) has been presented a novel framework for the evaluation of the vulnerability to earthquakes of the Indian subcontinent using deep learning models LSTM with geospatial information systems techniques. This study presents new findings that can fill in these gaps with a national-scale model of seismic vulnerability, and overcome the limitations of city- or state-level assessments. This uses several spatial and socio-technical elements such as land use, geology, geomorphology, fault lines, and infrastructure as well as population density to produce socio, structural, and geotechnical vulnerability maps at detailed levels. The LSTM-GIS model that has clearly standouts for the prediction of seal- to very-high seismic vulnerability zones covering the Delhi region, northeastern region, some part (Gujarat region) of plains in West Bengal. Strong model performance metrics were achieved, with 87. 8% precision, 90% sensitivity, and 84.9% specificity, indicating that the model effectively identified areas in need of urgent action to reduce disaster risk. While the study is an important step in the right direction for earthquake preparedness in the country, we must note as significant caveats the static nature of some of the input data used, and the limitations related to working with real-time data, a part of the study about which no significant details were offered in the report and which could limit the ability of the model to dynamically react to the sudden change in the geological status quo.

There is a DL model that has predicted the time and location of earthquakes based on Long Short-Term Memory (LSTM) networks which also implement the attention mechanism stated by (Al Banna et al.; 2021). Focused on enhancing earthquake forecasting, with Bangladesh as a specific case study, the research tackles the issue of locality in earlier models by using regional seismic precursors from the nation's fault database as predictors of regional seismicity one month ahead of time. The researchers used Keras Tuner for the selection of architecture among different combinations of LSTM and dense layers. The LSTM model was modified to include an attention mechanism, which improved the accuracy of predictions, yielding an accuracy of 74.67% in predicting earthquake events. Furthermore, a regression model implemented using LSTM and dense layers predicted the epicentre in relation to a fixed point location with RMSE of 1.25. Despite its experimental results being promising for regional temporal and spatial prediction, the provided model is limited to the regional dataset used and not valid in any geographic context, leaving it ambiguous as to whether it will generalize and scale.

Similarly, (Malebary; 2023) presents a new hybrid deep learning model called IS-CNN-LSTM, which can learn to predict fire at an early stage and analyze the intensity of fire, addressing existing disadvantages like low accuracy, high false alarm rates, and inability to capture small or far away fire. It uses instance segmentation that accurately differentiates fire from non-fire event by fully merging a deep 57-layer convolutional neural network (CNN) and a long short-term memory (LSTM) network. We also propose a key-frame extraction algorithm, so that CNN including 21 convolutions, 24 ReLU, 6 pooling, 3 fully connected, 2 dropout and a SoftMax layer, will be able to classify video frames fewer than their number. It uses IoT devices to send alerts according to the calculated fire severity. In this study a publicly available data set containing fire and normal video footage was used to validate the proposed model, and the proposed model obtained a correct classification of 95. 25%, a false positive rate of 0. 09% and a false negative rate of 0. 65%, and only takes 0.08 seconds to predict. While qualitatively good performance, its complexity might not be able to deploy on resource-constrained edge devices, and its dependency on video-based input may limit the effective use of the algorithm when the environments have smoke or block.

There is another study who gave a meta-analytic review of 128 studies published between 2010 and 2021 on the use of Artificial Neural Networks (ANN), a category of Deep Learning (DL), in the context of disaster management (DM), which is used to minimize human suffering during emergencies stated by (Guha et al.; 2022). This paper aims to assess the effectiveness of ANN based techniques in different disaster types including floods, earthquakes, storms and wildfires, and different phases of DM including 'Mitigation and Preparedness' and 'Response and Recovery'. The review outlines several prominent trends, such as the widespread application of BPNN for predicting disasters and the significant potential that CNN shows to generate actionable insights from social media activity in response to crises. ANN showed better performance compared to traditional methods, especially in terms of flood prediction and emergency response, as articulated in the study. The review also recognizes some limitations, including possible bias from the selection of literature and a lack of uniform performance benchmarks between studies. Finally, it ends with the future research directions to improve the flexibility, explicability, and scalability of ANN-based solutions in various disaster situations.

At last (Tang et al.; 2023) presents a dynamic evaluation model that assesses the capabilities of meteorological disaster emergency response systems using a recurrent neural network (RNN)-based method. The authors identify the weaknesses of conventional qualitative evaluation and complicated approaches with too many parameters that cannot be easily handled, suffer subjectivity, and are complex in actual implementation; To address these obstacles, the authors propose a recurrent neural network (RNN) autoencoder model based on an encoder–decoder framework that automatically extracts capability features from time-series data. First, before training the model, an evaluation index system is built to define the relevant metrics and indicators. This study proposed a deep learning methodology capable of automatic trial of huge dataset dimensionality reduction with proper feature extraction and providing affinity to complex compared to trivial samples leading to more efficient capability assessments. The experiments prove the great performance of the model in precision and learning focus. However, it may have some limitations, such as the need for large, high-quality data sets and the challenges associated with generalising the model to different types of disasters or regional contexts without extensive customisation.

2.2 Applications of DL in Disaster Recovery

Deep learning (DL) has transformed disaster recovery by automating intelligently and allowing for faster decision making in a number of key areas. Among the most common applications of deep learning is in the field of infrastructure damage assessment, which uses convolutional neural networks (CNNs) to analyse post-disaster images acquired by drones and satellites. These models allow us to better understand the scope of destruction, identifying which buildings, roads and bridges have been damaged, and therefore speeding up emergency and repair efforts. Other important applications include debris and resource mapping, in which U-Net-style semantic segmentation enables the identification of debris zones on satellite imagery, as well as the identification of clean water sources, temporary shelter, and the operative road network. Such knowledge helps to optimize a disaster response logistics and resource allocation. In predictive recovery planning, the recurring neural networks (RNN) and long- and short-term memory (LSTM) models predict recovery timelines, power restoration schedules, and supply chain movements by employing historical and real-time data (Almasoudi; 2023). These models improve the expeditiousness of recovery through proactive planning and mitigating delays. Also, advances in DL models such as BERT and attention-based transformers applied to social media mining allow real-time analysis of sentiments and needs from Twitter, Facebook, and other social media platforms. This enables responders to quickly spot urgent needs - food, medical assistance, or evacuation - via crowd-sourced data. These DL applications are not only accelerating the recovery process but also transforming it into an adaptive and data-driven process. Using multimodal data, from satellite images to text-based posts, DL supports a broad and scalable methodology for post-disaster management. In short, these technologies offer a paradigm shift from reactive to proactive and data-driven disaster recovery measures, enabling decision-makers to act quickly, distribute resources wisely and fortify resilience for the future.

3 Methodology

3.1 Dataset Description

The Earthquake Prediction dataset, sourced from Kaggle, contains detailed seismic event records intended to support the prediction and analysis of earthquakes. The dataset comprises multiple attributes, including spatial, temporal, and seismic characteristics. Key columns include Date and Time, which timestamp each seismic event, while Latitude and Longitude mark its geographical occurrence. The Type field identifies the nature of the seismic event (e.g., earthquake, explosion), and Depth captures the focal depth at which it occurred. There are some fields like Depth Error, Magnitude Error and Horizontal Error which reflect measurement uncertainties while Depth Seismic Stations, Magnitude Seismic Stations, and Azimuthal Gap provide information on data collection quality and geographical coverage. The Magnitude field indicates the strength of the earthquake, accompanied by its Magnitude Type (e.g., ML, MW). Root Mean Square reflects the average residuals of seismic recordings. The dataset also includes unique identifiers like ID, and metadata such as Source, Location Source, Magnitude Source, and Status, which describe the origin and verification status of the entries. Notably, some attributes have substantial missing values, such as Magnitude Error and Horizontal Error, requiring preprocessing before modeling. Overall, the dataset is a comprehensive collection of seismic data suited for predictive modeling and trend analysis in disaster management.

3.2 Data Cleaning and Preprocessing

To secure data quality and reliability for predictive modelling there is a good data cleaning process was undertaken on the Earthquake Prediction dataset. Initially columns with more than 70% missing values were identified and dropped as such high levels of missingness could undermine the integrity of the analysis. This threshold-based removal helped eliminate non-informative features like Depth Error, Magnitude Error, and Horizontal Error, among others. Following this, the remaining missing values in the dataset were imputed using the median of each respective numeric column. Median imputation is a string type of choice which is mainly in the presence of outliers by ensuring that the central tendency of the data is preserved without being skewed. The ID column, serving only as a unique identifier and contributing no predictive value, was also removed to smooth the dataset. As a result of this cleaning process, the dataset retained important attributes like Latitude, Longitude, Type, Depth, Magnitude, Magnitude Type and various source-related columns. Only a few entries remained missing, including three in the newly engineered Datetime column, which were addressed in subsequent steps. This preprocessing ensures a cleaner, more consistent dataset ready for exploratory analysis and machine learning model development in the context of earthquake prediction.

3.3 Data Visualization

Figure 1 illustrates the distribution of earthquake magnitudes based on observed seismic data, highlighting the prevalence of smaller magnitude events over larger ones. The histogram displays the frequency of earthquakes binned by magnitude, with the majority concentrated between magnitudes 5.5 and 6.0. As magnitude increases, the frequency of occurrences declines sharply, forming a long-tail distribution. This pattern aligns with the power-law distribution typically seen in natural phenomena such as earthquakes, where small events occur far more frequently than large ones. The histogram has a kernel density estimate (KDE) curve, which provides a smoothed visualisation of the underlying distribution. The curve further emphasizes the steep drop in frequency as magnitudes increase, with a pronounced peak at the lower end. The decline in frequency becomes progressively gradual, indicating that high-magnitude earthquakes, while significantly less frequent, are still part of the overall distribution. The exponential decay pattern evident in the plot reflects the Gutenberg-Richter law, which quantifies the inverse relationship between the magnitude of the earthquake and its frequency.

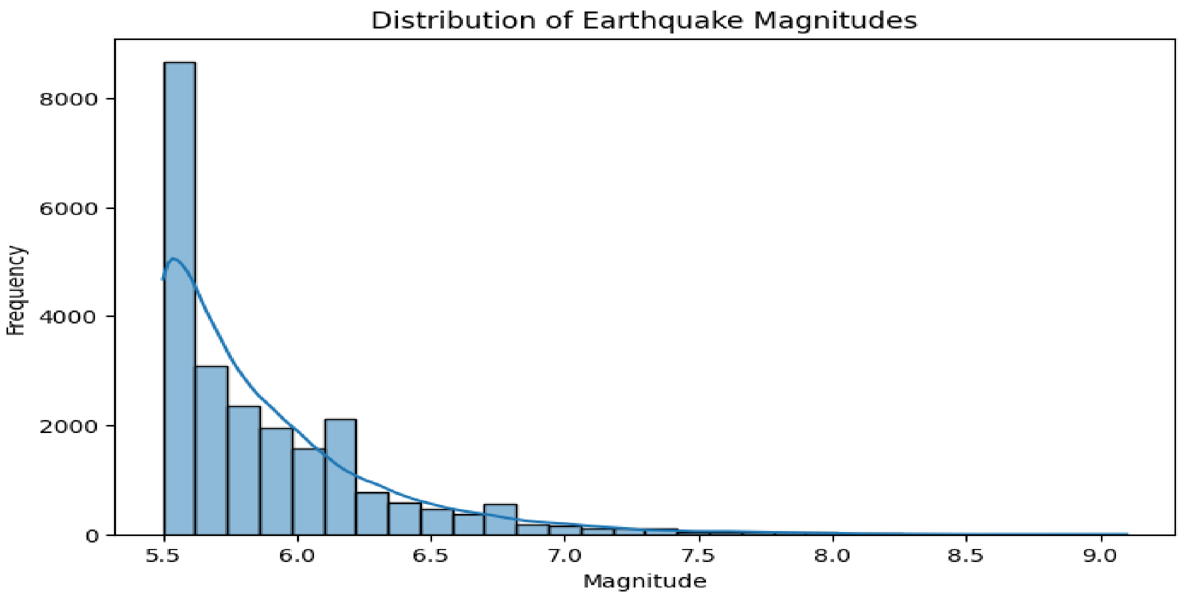

Figure 1: Distribution of Earthquake Magnitudes

Figure 2 presents a boxplot of earthquake depths, providing a visual summary of the central tendency, spread, and presence of outliers in the depth distribution of seismic events. The boxplot reveals that most earthquakes occur at relatively shallow depths, as indicated by the dense clustering of data points near the lower end of the depth scale. The interquartile range (IQR), represented by the central box, shows that the middle 50% of earthquakes typically occur within the top 0 to 70 kilometers of the Earth's crust.

The median depth lies closer to the lower quartile, suggesting a strong skew toward shallow events. Whiskers extending from the box indicate the range of non-outlier depths, beyond which numerous individual points appear as outliers—earthquakes occurring at significantly greater depths, extending up to nearly 700 km. These deep-focus earthquakes, while less common, are geophysical significant as they often occur in subduction zones. The sheer volume of outliers in the plot underscores the importance of accounting for such variability when analyzing seismic data.

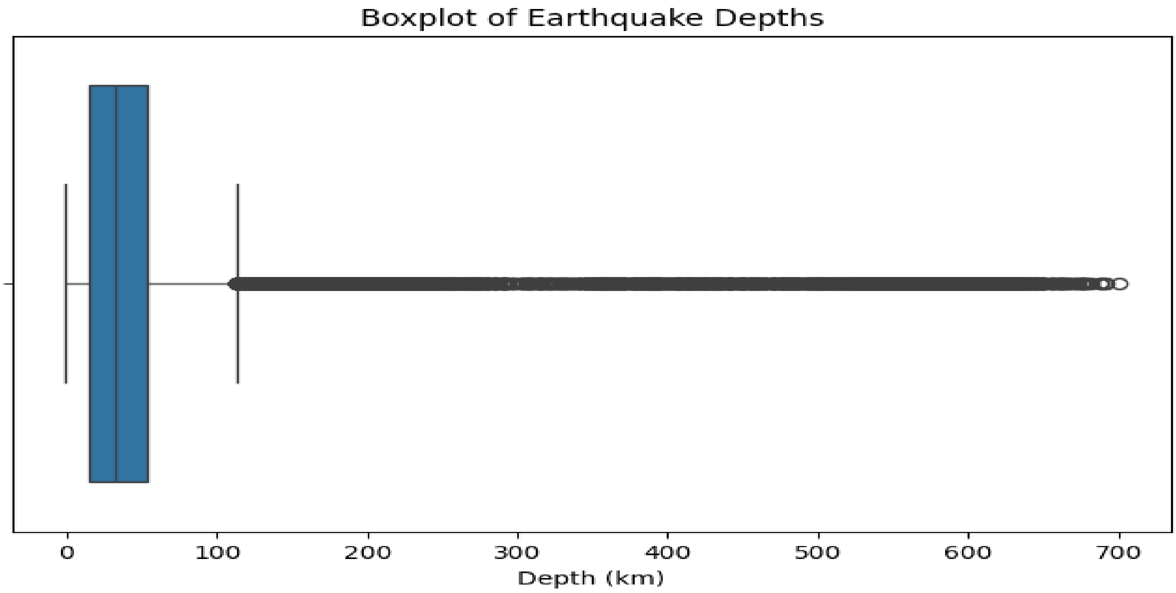

Figure 2: Boxplot of Earthquake Depths

Figure 3 depicts the annual frequency of earthquakes over a period of several decades, providing information on temporal trends and potential changes in the reporting of seismic activity. The line graph plots the number of recorded earthquakes per year, revealing a general upward trend from the early 1960s to the 2010s. While some fluctuations are evident—such as dips in the late 1970s and late 1990s, and a pronounced peak around 2011—overall, the data suggest an increasing number of recorded earthquakes over time. This rise could be attributed not only to actual seismic activity but also to improvements in global seismic monitoring networks, increased population density (resulting in more reported events) and technological advances in earthquake detection. The sharp spike in 2011 likely corresponds to a major seismic event or a cluster of significant earthquakes that year, highlighting the variable nature of tectonic activity. Furthermore, the more stable and higher baseline of earthquake numbers in recent decades indicates improved consistency in data collection.

Figure 4 visualises the global geographic distribution of earthquakes, mapped by plotting latitude against longitude and colour coded according to the magnitude of the earthquake. Each point represents an individual seismic event, with warmer colors (such as

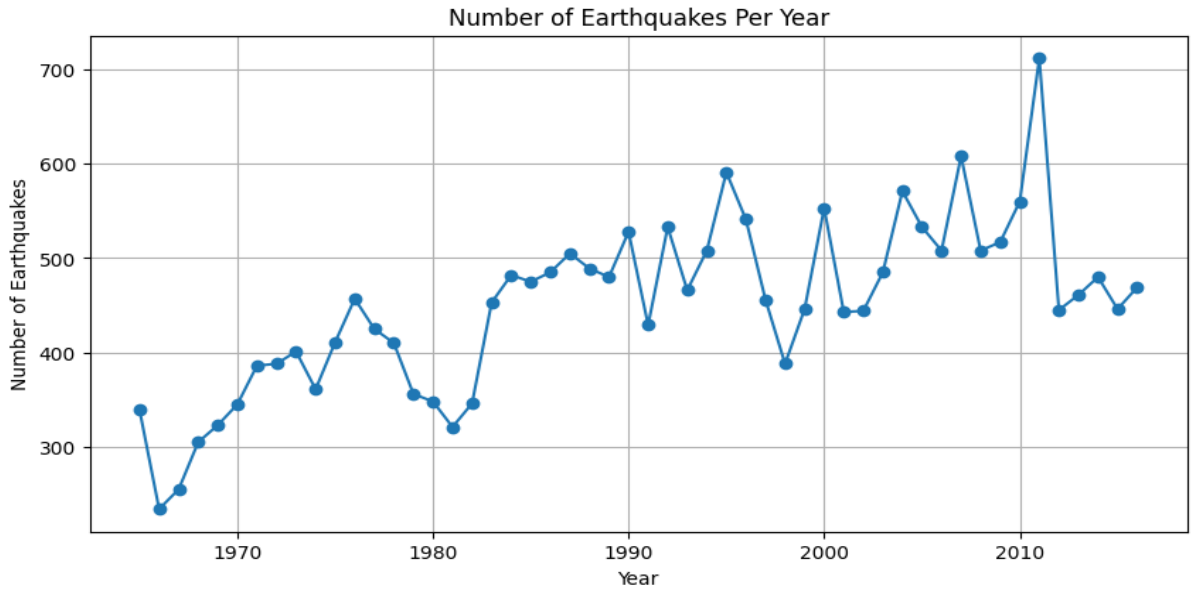

Figure 3: Number of Earthquakes Per Year

red and orange) indicating higher-magnitude earthquakes, and cooler colors (blue tones) representing lower magnitudes. The scatterplot clearly outlines major tectonic plate boundaries, as earthquakes are densely clustered along well-known fault lines such as the Pacific Ring of Fire, Mid-Atlantic Ridge, and Himalayan Belt. The high concentration of seismic events along the Pacific Rim—particularly near Japan, Indonesia, the west coasts of the Americas, and New Zealand—emphasizes the tectonic volatility of this region. In contrast, the interiors of continental plates exhibit sparse activity, reinforcing the idea that earthquakes primarily result from interactions at plate boundaries. Notably, some of the highest-magnitude events (depicted in red hues) are found in subduction zones where oceanic plates dive beneath continental plates. This spatial distribution aligns with geological theory and provides critical information for the assessment of seismic hazards.

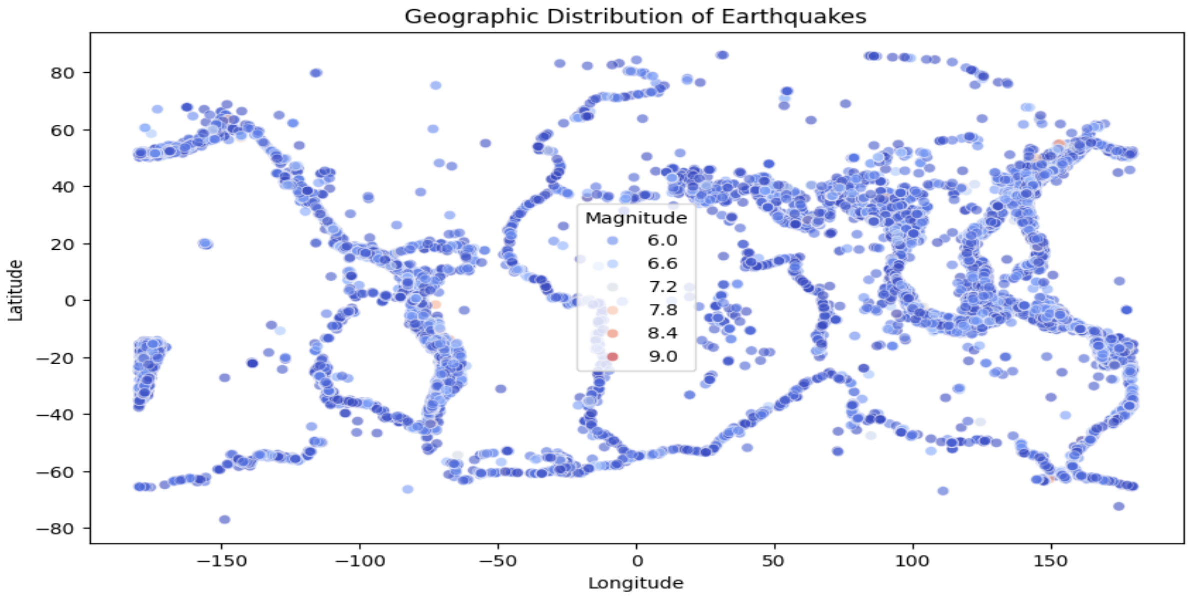

Figure 4: Geographic Distribution of Earthquakes

Figure 5 presents a correlation heatmap of numerical features in the earthquake dataset, offering insight into the relationships between key variables such as latitude, longitude, depth, magnitude, root mean square (RMS), and year. Each cell in the heatmap displays a correlation coefficient ranging from -1 to 1, with values closer to 1 indicating strong positive correlation, values near -1 indicating strong negative correlation, and values around 0 indicating no linear relationship. The diagonal line of ones reflects perfect self-correlation. In particular, most pair of characteristics exhibit very weak or negligible correlations, suggesting a high degree of independence among variables. For example, magnitude does not show almost no meaningful correlation with depth (0.02), latitude (0.03), or time (–0.03), indicating that larger earthquakes do not systematically occur deeper, at specific latitudes, or in particular years. The strongest correlation observed is between latitude and longitude (0.20), which is still weak, likely reflecting geographic clustering in seismically active zones. The Root Mean Square (RMS), which typically measures waveform amplitude or energy, also shows only weak relationships with other features.

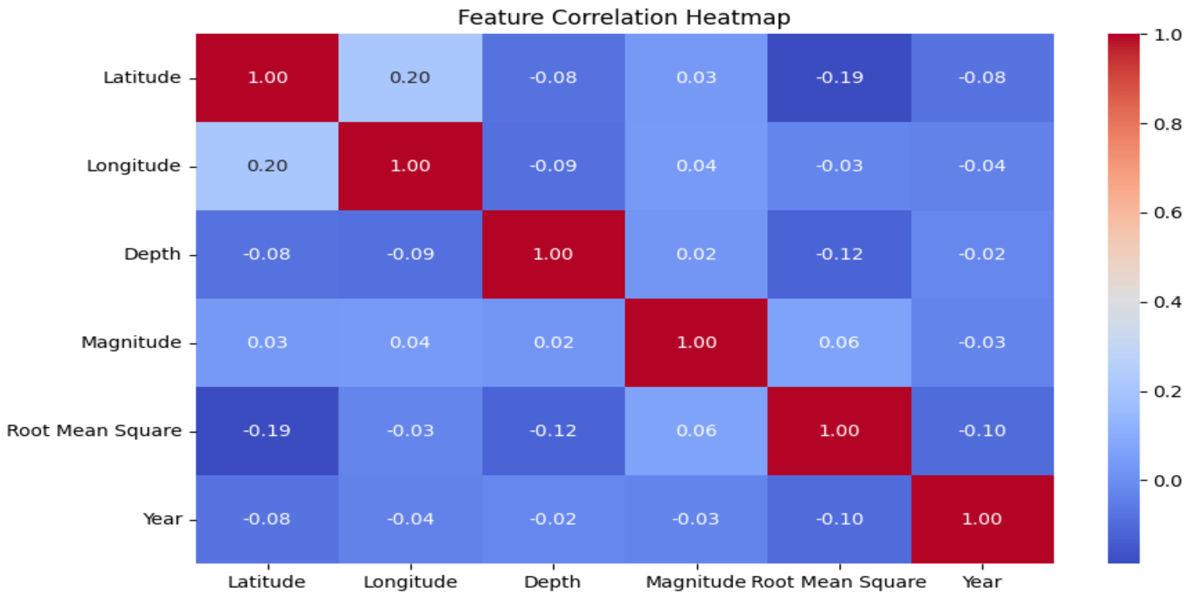

Figure 5: Feature Correlation Heatmap

Figure 6 illustrates a scatter plot showing the relationship between earthquake magnitude and depth, offering a visual analysis of how seismic energy release varies with the depth at which an earthquake occurs. Each point on the plot represents an individual seismic event, with depth on the x-axis (ranging from 0 to 700 km) and magnitude on the y-axis (ranging from approximately 5.5 to 9.1). A high density of events is observed at shallow depths (0–100 km), suggesting that most earthquakes originate near the Earth's surface. These shallow quakes span a wide range of magnitudes, including many high-magnitude events, which often result in significant surface-level damage. As depth increases, the frequency of earthquakes appears to decrease, but interestingly, high-magnitude events still occur even at depths beyond 500 km, albeit less frequently. The scatter does not suggest a strong linear relationship between depth and magnitude, aligning with the weak correlation seen in Figure 5.

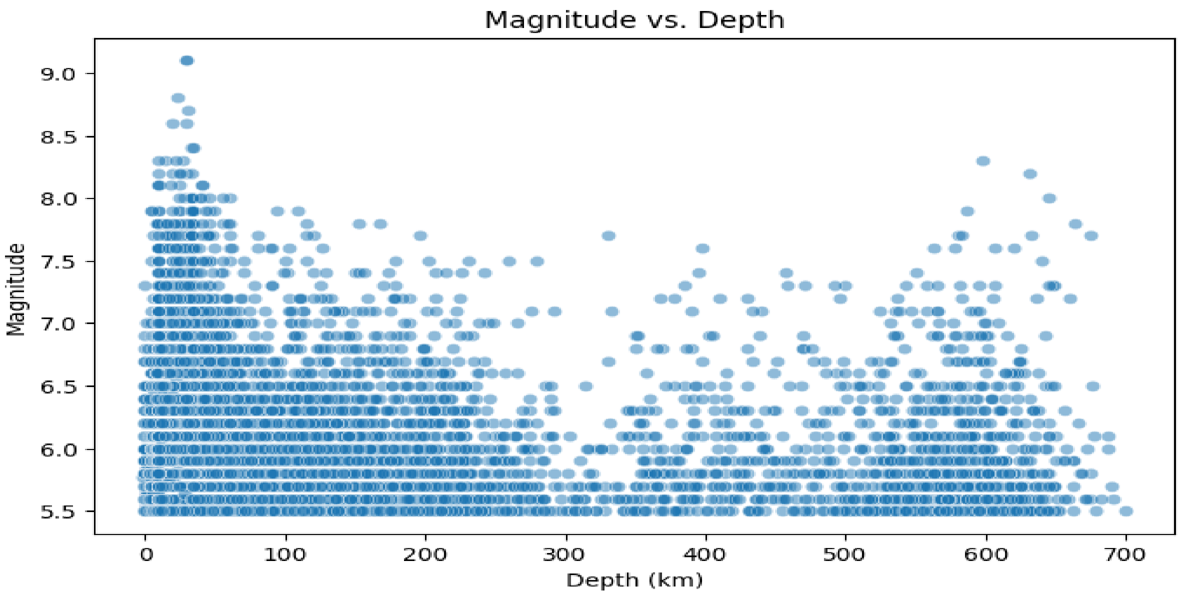

Figure 6: Magnitude vs. Depth

3.4 Data Splitting

To prepare the data set for the development of the machine learning model, the cleaned and pre-processed data was divided into input features (X) and target variables (y), which represents the output the model aims to predict. This step is critical in supervised learning tasks to enable the model to learn from historical data and generalize to unseen scenarios. Once separated, the dataset was split into training and testing subsets using the train_test_split function from the scikit-learn library. An 80-20 ratio was used, where 80% of the data was allocated for training the model (X_train, y_train) and the remaining 20% was reserved for testing its performance (X_test, y_test). This stratified split ensures that the model is exposed to a significant portion of the data during learning while preserving a representative sample for unbiased evaluation. This split helps mitigate overfitting and provides a realistic assessment of the predictive accuracy of the model. By maintaining the separation between the training and testing phases, this approach allows reliable validation of the generalisation capabilities of the model, making it a foundational step in the machine learning pipeline for earthquake prediction.

4 Design Specification

Several deep learning architectures were created for this research work. Figure 7 shows the workflow in which the deep learning models were placed.

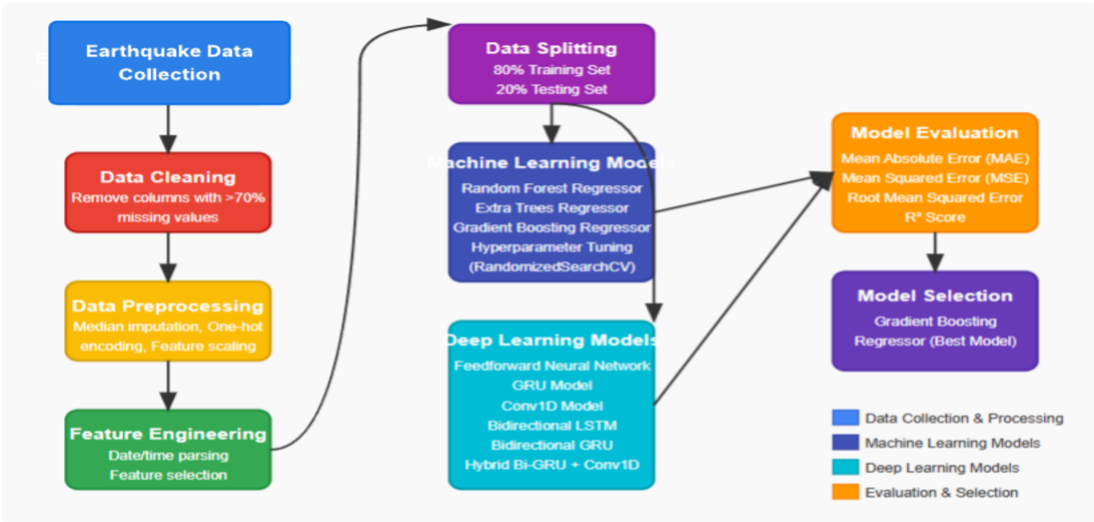

Figure 7: Workflow diagram with a range of deep learning models.

5 Implementation

The code developed for this report is available on the following GitHub Repository. All three research objectives are described in this section.

5.1 Random Forest Regressor

The section used Random Forest Regressor to forecast earthquake magnitude through multiple seismic measures and geographic variables. The gathered data consisted of numerical along with categorical variables including Latitude, Longitude, Type and Depth as well as Magnitude Type and Root Mean Square. One-Hot Encoding served as the method for processing categorical data by creating binary indicator variables from non-numeric attributes. The dataset split into its features X and target y so Magnitude served as the variable for prediction. The initial baseline Random Forest model containing 100 estimators received training data from X_train and y_train through the use of fit () method. The model provided predictions using the X_test data through predict () before evaluating its performance with Mean Absolute Error, Mean Squared Error and the R^2 Score. The RandomizedSearchCV method performed hyperparameter tuning operations for model enhancement. The parameter grid contained adjustments among the number of estimators along with maximum depth parameters and sample split rules. A three-fold cross-validation method through a randomised search conducted twenty times to optimise R^2 scoring metrics. These optimized hyperparameters were printed as they provide the opportunity to retrain the model and perform evaluation which leads to improved generalization and predictive power.

5.2 Feedforward Neural Network

The prediction of earthquake magnitude from scaled seismic features done on a feedforward neural network implementation that has been used TensorFlow and Keras frameworks. The scikit-learn StandardScaler function standardised the features before training through a transformation that normalised their mean to zero and their variance to one. The neural networks require this scaling process to achieve two main outcomes: uniform contribution from all features during training sessions and improved speed of convergence. The Sequential API from Keras enabled model creation with three layers including 64 neurons and ReLU activation in the first hidden layer then 32 neurons with ReLU activation in the second hidden layer before the single-neuron regression output layer terminated the network. The model input structure was created from the number of columns contained in the training dataset. ReLU served as the activation function because it effectively deals with non-linear data while preventing gradient vanishing. This FNN design system works best for continuous output estimation while learning intricate data patterns in its multiple layers. After developing the model, it could proceed to compilation and training which involved selecting appropriate functions for regression tasks.

5.3 GRU Model

The GRU-based neural network was implemented to capture the temporal and sequential dependencies within the data using TensorFlow and Keras and was used to model and predict earthquake magnitude. Gated Recurrent Units (GRUs) are a type of RNN that are specifically used for time-series or sequence-orientated data they are more computationally efficient compared to LSTMs since, while they retain long-term dependencies, they are capable of being more effective computationally. The model was created using the Sequential API and three stacked GRU layers. The GRU layer connected as the first GRU that set the return sequence to True that has 128 units, thus, it will get the whole sequence for the next GRU layer that contains 64 units. This was followed by a third GRU layer (with 32 units) which did not return sequences i.e. it passed to the next layer only its final output. A tanh activation function was used for all GRU layers since this is widely utilized in recurrent networks and can take positive and negative values. The final layers were a dense output layer with one neuron to perform the regression task of predicting the magnitude. The input shape was defined as (X_train. np.zeros (shape [1], 1) because the features have been transformed to 3D format for processing in GRU. The construction of model architecture that focuses on extracting sequential patterns that might have an impact on the prediction of seismic activity.

5.4 Convolutional 1D Model

With the right skill to determine local feature patterns throughout the duration of the input attributes, a 1D Convolutional Neural Network (Conv1D) model was constructed using TensorFlow and Keras to predict earthquake magnitude. Since Conv1D models expect input data shapes to be 3-dimensional, the model starts with a Reshape layer, converting the input shape from (X_train. shape [1],) to (X_train. shape [1], 1), considering every feature as a time step with a single channel. The initial convolutional layer utilizes 64 filters with a kernel size of 5 and implements the ReLU activation function, facilitating the ability of the model to identify more abstract representations of features. Then, a MaxPooling1D layer with pool size 2 performs down-sampling of the feature maps to reduce dimensionality and computation while retaining the most important features. A second convolutional layer with 32 filters and a kernel size of 3, helping to refine the feature extraction process, is added. Thus, its multi-dimensional feature maps are reshaped into one-dimensional tensor, and each feature map is averaged using the GlobalAveragePooling1D layer in turn to minimize the overfitting. The model has a ReLU activated dense layer with 32 neurones and finally an output layer with one neurone for regression. This architecture is suitable for capturing local dependencies in structured tabular data such as earthquake features.

5.5 Bidirectional LSTM Model

A modified BiLSTM model was developed in TensorFlow and Keras to predict the magnitude of the earthquake by learning intricate temporal relationships among sequences of seismic feature data. This approach allows the model to capture patterns in the input sequences in both forward and backward directions, which is particularly useful in sequence prediction tasks, as the context from both previous and future states can enhance the performance of the predictions. The architecture consists of a single Bidirectional LSTM layer with 128 units and return_sequences=True, which makes the entire tensor pass to the next layer. Then we have another Bidirectional LSTMCELL with 64 units, also returning sequences to keep the temporal information. The third LSTM layer has 32 units and return_sequences=False, meaning it will not return output for each step of the input sequence, but rather a single condense output for the entire input sequence. The tanh activation function is used throughout all LSTM levels, typically showing good performance in sustaining stable gradients in duplicate systems. The last Dense layer with a single neuron produces the predicted earthquake magnitude. The input data into 3D (samples, time steps, features) would need to be reshaped to fit the shape of (X_train. reshape(shape [1], 1) to match the shape expected by the model. Such BiLSTM architecture is very effective for modeling sequential relationships in time-dependent geological data.

5.6 Hybrid Bidirectional GRU and Convolutional 1D Model

To predict earthquake magnitudes with better feature extraction and temporal pattern recognition, a hybrid deep learning model was created in TensorFlow and Keras, integrating both Conv1D and Bidirectional GRU (Bi-GRU) layers. We begin with a 1D Convolutional layer with 64 kernels of size 5, to get the local spatial features of the seismic input data. It applies the ReLU (Rectified Linear Unit) activation function to introduce non-linearity and allow the model to learn complex relationships. Next, we apply a MaxPooling1D layer (pool size of 2) that decreases the dimensionality of the feature maps and preserves the most salient features whilst reducing the possibility of overfitting. This output is then fed into three Bidirectional GRU layers. The first Bi-GRU layer has 128 units, followed by 64 and 32 units in the next layers. For example, the time delay neural network (TDNN) layers, which use the tanh activation function, are particularly skilled at fetching forward- and backward-looking long-range dependencies, a task essential to time-evolutionary earthquake data. Model ends with 1-unit Dense layer which is appropriate for regression output. This ability will allow the model to first abstract spatial relationships from the input data (via convolution) and then learn about those sequential dependencies (via recurrent), making it a better architecture for predicting the magnitude of an earthquake.

5.7 Extra Trees Regressor and Gradient Boosting Regressor

Scikit-learn was used to implement two strong ensemble-based machine learning models—Extra Trees Regressor (ETR) and Gradient Boosting Regressor (GBR)—to make the prediction of earthquake magnitude stronger in terms of robustness and accuracy. The Extra Trees Regressor is an ensemble learning method that generates the unpruned decision trees from the training data and takes the average of the results. It uses random sampling of features and random thresholds for splitting, making it very efficient and preventing overfitting very easily. The ETR model was initialized with 100 estimators and a fixed random state for reproducibility. X_train_scaled: The model was trained with the scaled feature set using StandardScaler, making it necessary when using a tree-based model with other models in an ensemble or comparison. The scaled test set was then used to generate predictions. In parallel, Gradient Boosting Regressor was also used, which creates models in a successive way, where each new model attempts to correct the errors of the previous model. To balance between bias and variance, we set the maximum depth of each tree, the number of estimators (set to 100), and the learning rate to 0.1. Predictions were then made using the test set after training on the scaled data. These models provide different angles by which they function—ETR reduces variance through randomization, while GBR refines prediction accuracy by limiting the error from one iteration to the next.

6 Results and Evaluation

A systematic exploration of models for this regression problem was performed. Results appear in Table 1, and are discussed here in detail.

6.1 Case Study 1: Random Forest Regressor

The Random Forest Regressor gave a Mean Absolute Error (MAE) of 0.3052, Mean Squared Error (MSE) of 0.1716 and R^2 score of 0.0629. Although it is a powerful ensemble model known for handling nonlinear relationships. It slightly underperformed in comparison to deep learning models. Its moderate MAE and low R^2 suggest that while it can capture general type of trends in the data, its predictive accuracy and variance explanation are limited which is possibly due to the temporal and complex nature of earthquake data that might benefit more from sequential models.

6.2 Case Study 2: Feedforward Neural Network

The Feedforward Neural Network has been achieved an MAE of 0.2848 and MSE of 0.2033 with a Root Mean Square Error (RMSE) of 0.4509 and a negative R^2 of -0.1101. This shows that the model performed worse than a simple average predictor. While the network was able to reduce the Mean Absolute Error (MAE) slightly compared to Random Forest, the high Mean Squared Error (MSE) and negative R-squared value reveal limitations in its ability to generalize. The static structure of feedforward networks may struggle with the dynamic and temporal dependencies present in seismic data.

6.3 Case Study 3: GRU Model

The GRU (Gated Recurrent Unit) model improved performance with an MAE of 0.2849, MSE of 0.1572, RMSE of 0.3965 and an R^2 score of 0.1414. GRUs are good for sequential data, such as earthquakes, as they maintain memory of past inputs. This model has shown better generalisation and prediction accuracy compared to feedforward and Random Forest models by showing its performance in capturing temporal dependencies which are very important for predicting events that unfold over time such as earthquakes.

6.4 Case Study 4: Convolutional 1D Model

The Conv1D model has achieved an MAE of 0.2931, MSE of 0.1622, RMSE of 0.4028, and R^2 of 0.1140. By extracting local features and patterns using convolutional layers this model has captured important trends from the data. While slightly less accurate than GRU, Conv1D still outperformed traditional models. Its performance confirms that spatial pattern recognition from sequential data can enhance predictions, although combining it with temporal models might give even better results.

6.5 Case Study 5: Bidirectional LSTM Model

The BiLSTM model showed good performance with an MAE of 0.2803, MSE of 0.1544, RMSE of 0.3929, and an R^2 score of 0.1571. Using long-term dependencies in both forward and backward time sequences, complex temporal patterns have been captured. This bidirectional flow enabled the model to predict with higher accuracy and generalization than previous approaches which is very useful in earthquake forecasting, where past and future context plays an important role.

6.6 Case Study 6: Bidirectional GRU Model

Using the BiLSTM, the Bidirectional GRU model achieved the same MAE of 0.2803, MSE of 0.1544, RMSE of 0.3929, and R^2 of 0.1571. It benefits from fewer parameters and faster training than LSTM, while still effectively capturing sequence information. This model's performance illustrates GRU's suitability for temporal earthquake data, offering a balance between accuracy and computational efficiency without sacrificing much in predictive strength.

6.7 Case Study 7: Hybrid Bidirectional GRU and Conv1D Model

The hybrid model combining Bidirectional GRU and Conv1D recorded an MAE of 0.2819, MSE of 0.1532, RMSE of 0.3914, and R^2 of 0.1633. This integration merges the temporal sensitivity of GRU with the extraction of local features from convolutional layers, resulting in improved accuracy. The model captures both sequential and spatial characteristics of the data, which are vital for understanding and predicting complex earthquake patterns, making it one of the most effective models in this study.

6.8 Case Study 8: Extra Trees Regressor and Gradient Boosting Regressor

Extra Trees Regressor produced an MAE of 0.2955 and R^2 of 0.0704, while Gradient Boosting Regressor outperformed it with an MAE of 0.2817 and R^2 of 0.1734. GBR's iterative error correction and ability to handle non-linearity make it a strong performer. Although tree-based models do not take advantage of sequential context, GBR's superior performance highlights its efficiency in capturing feature interactions, positioning it among the best-performing models for this earthquake prediction task.

6.9 Case Study 9: Support Vector Regression (SVR) Model

To further expand the predictive capabilities of the study which is an SVR model and that was implemented. SVR is well-suited for regression tasks involving non-linear relationships and can handle high-dimensional data effectively through the use of kernel functions. The following outlines the methodology, preprocessing, modeling pipeline, and evaluation metrics used in this approach.

6.9.1 Data Preprocessing for SVR

Similar to the previous models, preprocessing was a critical step for ensuring SVR performance:

- **Missing Value Handling**: Columns with over 70% missing data were removed. The remaining missing types of values in numeric columns have been imputed with their respective median values while categorical values have been imputed with the mode.
- **Feature Selection**: Only numerical features were retained for the SVR pipeline.
- **Target Transformation**:Since SVR benefits from normally distributed target values, a Box-Cox transformation was applied to the Magnitude column after ensuring all values were positive.

6.9.2 SVR Pipeline Construction

A machine learning pipeline was created using scikit-learn's Pipeline class. It consisted of:

- **StandardScaler**: Normalized feature values to have zero mean and unit variance.
- **SVR Model**: Used with default RBF kernel. Parameters were fine-tuned using empirical observation

6.10 Discussion

This section provides an in-depth interpretation of the experimental results presented in the previous section and addresses the core research questions posed in this study. The overall aim was to explore the effectiveness of various machine learning (ML) and deep learning (DL) models in predicting earthquake magnitudes using historical seismic data. While deep learning models have been widely recognized for their ability to model complex relationships in data, the results from this study present a more nuanced picture.

Table 1: Comparison Table of all Models

Model	**MAE**	**MSE**	**RMSE**	**R^2 Score**
Random Forest Regressor	0.3052	0.1716	N/A	0.0629
Feedforward Neural Network	0.2848	0.2033	0.4509	-0.1101
GRU Model	0.2849	0.1572	0.3965	0.1414
Conv1D Model	0.2931	0.1622	0.4028	0.1140
Bidirectional LSTM	0.2803	0.1544	0.3929	0.1571
Bidirectional GRU	0.2803	0.1544	0.3929	0.1571
Bi-GRU + Conv1D Hybrid	0.2819	0.1532	0.3914	0.1633
Extra Trees Regressor	0.2955	0.1702	0.4126	0.0704
Gradient Boosting Regressor	0.2817	0.1514	0.3891	0.1734
SVR	0.2676	0.8713	0.4048	0.1315

6.10.1 Analysis of model performance (Research Question One)

From the comparative analysis of multiple models, it is evident that the Gradient Boosting Regressor (GBR) achieved the best overall performance. With the lowest MAE (0.2817), lowest MSE (0.1514), and highest R^2 score (0.1734), GBR proved to be the most accurate and reliable model for the given dataset. Interestingly, this ensemble-based machine learning approach outperformed more complex deep learning models like Bidirectional LSTM, GRU, and hybrid architectures. This highlights that in some real-world applications with limited or non-sequential features, traditional ML models can still outperform DL architectures.

6.10.2 Does tuning improved Random Forest for earthquake prediction (Research Question Two)

Hyperparameter tuning using RandomizedSearchCV significantly improved the Random Forest Regressor's performance in terms of generalization. However, even after tuning, the Random Forest achieved an R^2 score of only 0.0629, which is lower than both deep learning models like GRU (0.1414) and other ensemble methods like GBR (0.1734). This suggests that while tuning can enhance performance, it cannot compensate for structural limitations of the model when dealing with complex nonlinear patterns in earthquake data.

6.10.3 Can a hybrid GRU and Conv1D model for Earthquake prediction (Research Question Three)

The hybrid Bidirectional GRU + Conv1D model performed competitively, with an MAE of 0.2819, MSE of 0.1532, and R^2 score of 0.1633. While it did not surpass GBR, it did outperform all other individual DL models tested, such as GRU, Conv1D, and BiLSTM. This confirms the hypothesis that combining sequential and spatial feature extraction improves performance. However, it also indicates that the data may not contain strong enough temporal signals to fully utilize the potential of sequential DL models.

6.11 Limitations

The main shortcomings of this study are that we did not have sequential, temporal data, therefore GRU and LSTM are not the best layers. Furthermore, the data set does not include some important geophysical parameters, such as tectonic plate limits or seismic wave properties, that would be able to improve the working power of the models. In addition, the feature engineering was light. Power transformations and discretisations could have been attempted with more time. Another interesting area for further insights could have been the application of model interpretability techniques. These include the use of more complex multi-origin data, deep learning algorithms for a better learning sequence, explainable AI (XAI) to enhance the model's explainability and deploying real-time models for actual earthquake prediction.

7 Conclusion and Future Work

7.1 Conclusion

In this research, various deep learning and machine learning models were examined and evaluated to predict the magnitudes of tectonic earthquakes. These ranged from classical ensemble regressors to sophisticated neural architectures. After thorough experimentation, the Gradient Boosting Regressor was found to be the best performer overall, yielding the highest prediction accuracy as well as the most reliable across essential evaluation parameters. In contrast, the Feedforward neural network model alone exhibited the weakest performance with its limited ability to discern intricate patterns present in the dataset. RandomizedSearchCV (Hyperparameter tuning) was next implemented, smartly testing multiple parameter combinations automatically for optimal performance of the Random Forest model (number of trees, depth of trees, sample splits, etc). This method increased the efficacy of the model, but it was still not as effective as the best-performing model. In general, the results of this study show that methodology based on boosting is likely better suited to the problem of earthquake magnitude on this data set, and that tuning and model selection play an important role in yielding more reliable results.

7.2 Future Works

Future work in this study can explore incorporating real-time seismic data and additional geophysical features such as tectonic plate boundaries, seismic waveforms, and soil compositions to improve model accuracy. Advanced feature engineering techniques and explainable AI (XAI) can enhance interpretability, enabling better trust and transparency in predictions. Integrating ensemble deep learning models or attention-based architectures like Transformers may further capture complex spatial-temporal dependencies. Additionally, deploying the predictive framework in a real-time environment using IoT-enabled sensor networks can support early warning systems. Cross-regional generalizability and scalability to other seismic zones should also be validated through transfer learning and domain adaptation techniques.

References

Al Banna, M. H., Ghosh, T., Al Nahian, M. J., Taher, K. A., Kaiser, M. S., Mahmud, M., Hossain, M. S. and Andersson, K. (2021). Attention-based bi-directional long-short term memory network for earthquake prediction, *IEEE Access* **9**: 56589–56603.
URL: *http://dx.doi.org/10.1109/ACCESS.2021.3071400*

Almasoudi, F. M. (2023). Enhancing power grid resilience through real-time fault detection and remediation using advanced hybrid machine learning models, *Sustainability* **15**(10): 8348.
URL: *https://doi.org/10.3390/su15108348*

Anbarasan, M., Muthu, B., Sivaparthipan, C., Sundarasekar, R., Kadry, S., Krishnamoorthy, S., Dasel, A. A. et al. (2020). Detection of flood disaster system based on iot, big data and convolutional deep neural network, *Computer Communications* **150**: 150–157.
URL: *https://doi.org/10.1016/j.comcom.2019.11.022*

Bhardwaj, A., Sam, L. and Martin-Torres, F. (2021). The challenges and possibilities of earthquake predictions using non-seismic precursors, *The European Physical Journal Special Topics* **230**(1): 367–380.
URL: *http://dx.doi.org/10.1140/epjst/e2020-000257-3*

Esposito, M., Palma, L., Belli, A., Sabbatini, L. and Pierleoni, P. (2022). Recent advances in internet of things solutions for early warning systems: A review, *Sensors* **22**(6): 2124.
URL: *https://doi.org/10.3390/s22062124*

Guha, S., Jana, R. K. and Sanyal, M. K. (2022). Artificial neural network approaches for disaster management: A literature review, *International Journal of Disaster Risk Reduction* **81**: 103276.
URL: *https://doi.org/10.1016/j.ijdrr.2022.103276*

Jena, R., Naik, S. P., Pradhan, B., Beydoun, G., Park, H.-J. and Alamri, A. (2021). Earthquake vulnerability assessment for the indian subcontinent using the long short-term memory model (lstm), *International Journal of Disaster Risk Reduction* **66**: 102642.
URL: *https://doi.org/10.1016/j.ijdrr.2021.102642*

Malebary, S. J. (2023). Early fire detection using long short-term memory-based instance segmentation and internet of things for disaster management, *Sensors* **23**(22): 9043.
URL: *https://doi.org/10.3390/s23229043*

Martinelli, G. (2020). Previous, current, and future trends in research into earthquake precursors in geofluids, *Geosciences* **10**(5): 189.
URL: *http://dx.doi.org/10.3390/geosciences10050189*

Munawar, H. S., Ullah, F., Qayyum, S., Khan, S. I. and Mojtahedi, M. (2021). Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection, *Sustainability* **13**(14): 7547.
URL: *https://doi.org/10.3390/su13147547*

Pi, Y., Nath, N. D. and Behzadan, A. H. (2020). Convolutional neural networks for object detection in aerial imagery for disaster response and recovery, *Advanced Engineering Informatics* **43**: 101009.
URL: *https://doi.org/10.1016/j.aei.2019.101009*

Tang, J., Yang, R., Dai, Q., Yuan, G. and Mao, Y. (2023). Research on feature extraction of meteorological disaster emergency response capability based on an rnn autoencoder, *Applied Sciences* **13**(8): 5153.
URL: *https://doi.org/10.3390/app13085153*