IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 June 2024, accepted 29 July 2024, date of publication 5 August 2024, date of current version 15 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3438556

RESEARCH ARTICLE

Predictive Modeling of Earthquakes in Los Angeles With Machine Learning and Neural Networks

CEMIL EMRE YAVAS, (Member, IEEE), LEI CHEN, (Senior Member, IEEE), CHRISTOPHER KADLEC, AND YIMING JI

Department of Information Technology, Georgia Southern University, Statesboro, GA 30460, USA

Corresponding author: Cemil Emre Yavas (cy02470@georgiasouthern.edu)

This work was suppored by the Dean's Office of the Allen E. Paulson College of Engineering and Computing at Georgia Southern University.

ABSTRACT Earthquakes pose a significant threat to urban areas, necessitating accurate forecasting models to mitigate their impact. This study focuses on earthquake forecasting in Los Angeles, a region with high seismic activity and limited research. We established a feature matrix for forecasting earthquakes within a 30-day period by analyzing the most predictive patterns from recent studies. Our model developed a subset of features capable of forecasting the highest magnitude of an earthquake. Using advanced machine learning algorithms and neural networks, our model achieved an accuracy of 69.14% in forecasting the maximum magnitude earthquake as one of the 6 categories. We aim to provide a useful guideline for future researchers.

INDEX TERMS Earthquake forecasting, machine learning, neural networks, Los Angeles, seismic activity, feature engineering, spatiotemporal analysis, predictive modeling, random forest, XGBoost, seismic energy.

I. INTRODUCTION

Earthquakes are natural disasters that can have devastating consequences, making their forecasting a crucial area of research to mitigate their impact. In our study, we delved into various papers on earthquake forecasting to compile predictive features and develop a predictive pattern matrix specifically for earthquakes in Los Angeles. By leveraging machine learning algorithms and neural networks, we aim to forecast the highest magnitude of earthquakes within a 30-day period.

Our research draws on a diverse range of studies in the field of earthquake forecasting from 1990 to 2024. Olsen et al. [1] highlighted the significant ground velocities expected during earthquakes, particularly near fault lines and in regions like the Los Angeles basin, where prolonged shaking can occur. This emphasizes the importance of understanding the dynamics of seismic events in specific geographical locations.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

In the realm of machine learning applications for earthquake forecasting, Asim et al. [2] explored the use of support vector regressors and hybrid neural networks to develop a predictive model. Their focus on seismic regions like Hindukush, Chile, and Southern California aligns with our interest in forecasting earthquakes in Los Angeles, shedding light on the relevance of considering regional seismic activity in predictive modeling.

Zhang et al. [3] proposed a precursory pattern-based feature extraction technique to enhance earthquake forecasting performance. This approach underscores the significance of extracting meaningful patterns from seismic data, which resonates with our objective of creating a predictive pattern matrix for earthquakes in Los Angeles.

The study by Bao et al. [4] on deep learning-based electromagnetic signal analysis for earthquake magnitude forecasting introduces an innovative method that combines explicit and implicit features. This approach could offer valuable insights into improving the accuracy of earthquake forecasts, aligning with our goal of enhancing forecasting precision.

Incorporating advanced neural network models, such as the graph convolutional neural network proposed by Bilal et al. [5], can significantly improve earthquake forecasting efficiency. Their emphasis on early earthquake detection using sophisticated neural network architectures underscores the potential for enhancing predictive capabilities through cutting-edge technologies.

The work of Hsu and Pratomo [6] on early peak ground acceleration forecasting using LSTM neural networks showcases the importance of leveraging models capable of capturing order dependence in seismic waves. This aligns with our approach of utilizing machine learning algorithms to forecast earthquake occurrences within a specific timeframe.

By integrating findings from a wide array of studies on earthquake forecasting, our research aims to enhance predictive modeling techniques specifically for the Los Angeles region. Through the integration of machine learning algorithms, feature extraction methods, and advanced neural network architectures, we strive to enhance the accuracy and timeliness of earthquake forecasts for improved disaster preparedness and response strategies.

II. BACKGROUND

The background section of this study provides a comprehensive overview of the scientific and technical foundations that underpin our research on earthquake forecasting in the Los Angeles region. By integrating key findings from various studies on seismic activity, fault systems, ground motion, and earthquake hazards, this section elucidates the factors influencing earthquake dynamics. It highlights significant research that has shaped our understanding of the geological conditions in Los Angeles, emphasizing the importance of detailed, site-specific data in enhancing the accuracy of seismic hazard assessments. This foundational knowledge sets the stage for the application of advanced machine learning and neural network algorithms to forecast earthquake magnitudes, ultimately contributing to improved preparedness and mitigation strategies.

The study ''Site Amplification in the Los Angeles Basin from Three-Dimensional Modeling of Ground Motion'' by Olsen [7], published in the Bulletin of the Seismological Society of America, investigates how local geological conditions in the Los Angeles Basin amplify ground motion during earthquakes. Using three-dimensional modeling techniques, the study demonstrates significant variations in ground motion amplification due to the complex subsurface structure of the basin. These findings highlight the importance of incorporating detailed, site-specific geological data into seismic hazard assessments to improve the accuracy of earthquake impact forecasts in the Los Angeles area.

The article ''Potential for a Large Earthquake Near Los Angeles Inferred from the 2014 La Habra Earthquake,'' published in Earth and Space Science, examines the implications of the 2014 La Habra earthquake for future seismic activity near Los Angeles. Donnellan et al. [8] highlight the interconnected fault systems in Los Angeles and their role in accommodating tectonic movements. By analyzing geodetic data, ground deformation, and fault interactions, the authors emphasize the concurrent movement of regional thrust, strike-slip, and oblique faults, underscoring their significance in seismic activity. The research concludes that stress changes and fault connectivity in the area could increase the likelihood of a significant seismic event. This information is vital for identifying potential earthquake triggers and patterns in the region, underscoring the importance of continuous monitoring and advanced modeling to better understand and mitigate earthquake risks in the Los Angeles area.

Hauksson's [9] research on earthquakes, faulting, and stress in the Los Angeles Basin provides valuable insights into the earthquake potential of thrust faults beneath the basin. The study suggests that underestimating the earthquake hazards in the region due to neglecting certain fault systems could have significant implications. Incorporating such findings into earthquake forecasting models is essential for enhancing the accuracy of forecasts.

The study by Shen et al. [10] on crustal deformation across the Los Angeles basin from geodetic measurements offers valuable data on the structural dynamics of the region. By aligning these findings with existing earthquake probability models, researchers can refine their forecasting algorithms and better anticipate seismic events in Los Angeles.

Loveless and Meade's [11] work on stress modulation on the San Andreas fault by fault system interactions provides insights into how stress variations influence earthquake recurrence intervals. Understanding stress dynamics along fault lines, especially those close to metropolitan Los Angeles, is crucial for refining earthquake forecasting models and assessing the likelihood of seismic events.

Incorporating insights from Romero et al. [12] on seismic hazards and water supply performance in Los Angeles can provide a holistic view of the environmental implications of earthquakes in the region. By considering factors like fault rupture, liquefaction, landslides, and site amplification, researchers can develop more comprehensive earthquake forecasting models that account for diverse hazards.

Roten et al.'s [13] study on expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity offers valuable information on how fault characteristics can influence ground motions. By simulating earthquake scenarios and forecasting ground motions in the Los Angeles Basin, researchers can refine their forecasting models and enhance the accuracy of earthquake forecasts.

Shaw and Suppe's [14] research on earthquake hazards of active blind-thrust faults under the central Los Angeles basin underscores the importance of considering newly identified fault systems in seismic risk assessments. By integrating data on these active faults into earthquake forecasting models, researchers can improve the precision of their forecasts and better prepare for potential seismic events.

Zechar and Jordan's study [15] on testing alarm-based earthquake forecasts offers valuable information on various forecasting models, including relative intensity, pattern

informatics, and the U.S. Geological Survey National Seismic Hazard Map. The article, published in Geophysical Journal International, evaluates the effectiveness of forecasting methods that issue alerts or ''alarms'' when certain seismic criteria are met, indicating an increased likelihood of an earthquake. The study involves statistical testing of these alarm-based forecasting models to determine their accuracy and reliability. Understanding the performance of these models is essential for selecting the most effective approach in forecasting earthquakes in Los Angeles. Their findings emphasize the need for rigorous testing and validation of forecasting methods to improve earthquake forecasting and ultimately enhance public safety and preparedness.

Huang et al.'s work [16] on the application of an improved Extreme Learning Machine (ELM) algorithm in earthquake casualty forecasting provides insights into the factors influencing earthquake casualties. By considering variables such as earthquake intensity, building collapse rate, and population density, researchers can enhance the accuracy of casualty forecasts, which is vital for disaster preparedness and response strategies in earthquake-prone areas like Los Angeles.

Initiatives such as the Collaboratory for the Study of Earthquake Predictability (CSEP) and the Regional Earthquake Likelihood Models Experiment (RELM) by Schorlemmer et al. [17] have paved the way for prospective earthquake forecasting efforts. Studies evaluating return periods and occurrence probabilities of maximum magnitude earthquakes by Al-Heety [18], and improved algorithms like Extreme Learning Machines (ELM) by Huang et al. [16] enhance earthquake casualty forecasts. Insights from studies on fixed recurrence and slip models by Rubinstein et al. [19] and self-organized criticality by Yang et al. [20] provide valuable perspectives on earthquake behavior forecasting and the challenges posed by complex seismic dynamics. Innovative approaches, such as the use of deep learning neural networks by Huang et al. [21] and attention mechanisms in earthquake forecasting models by Kavianpour et al. [22], offer further advancements.

While some studies, such as the work by Geller et al. [23], express skepticism about the predictability of earthquakes, highlighting the challenges in reliably forecasting the time, location, and magnitude of seismic events, these challenges continue to be addressed.

Eberhard et al.'s study [24] on a prospective earthquake forecast experiment in the western Pacific emphasizes the importance of ongoing experiments to enhance earthquake predictability models. This focus on continuous testing and refinement of forecasting models is crucial for improving the reliability of earthquake forecasts, particularly in regions with high seismic activity like Los Angeles.

Rubinstein et al.'s research [19] on fixed recurrence and slip models for earthquake behavior forecasting underscores the significance of understanding stress accumulation and release on fault lines. By integrating these models into earthquake forecasting algorithms, researchers can better anticipate seismic events and their potential impacts on regions like Los Angeles.

Tehseen et al.'s study [25] on earthquake forecasting using expert systems highlights the importance of long-term forecasts regarding the time, intensity, and location of future earthquakes. By utilizing expert systems and comprehensive data analysis, researchers can develop more robust earthquake forecasting models tailored to specific regions like Los Angeles.

Ogata's perspective [26] on earthquake forecasting research advocates for the development of statistical models of seismicity to accurately evaluate their predictive performance. By assessing the efficacy of statistical models in earthquake forecasting, researchers can enhance the reliability of forecasts and contribute to more effective disaster mitigation strategies.

Banna et al.'s work [27] on attention-based Bi-Directional Long Short-Term Memory (LSTM) networks for earthquake forecasting highlights the potential of advanced machine learning techniques in seismic forecasting. By leveraging deep learning models like LSTM networks, researchers can improve the accuracy of earthquake forecasts and enhance preparedness measures in earthquake-prone regions such as Los Angeles.

One pertinent reference is the study by Kagan [28] on the potential forecasting of earthquakes, emphasizing the role of real-time seismology in aiding relief efforts and issuing warnings of severe shaking before earthquakes occur. Understanding the feasibility of earthquake forecasting through real-time monitoring is essential for improving preparedness and response strategies in earthquake-prone regions like Los Angeles.

The research by Ma et al. [29] evaluates the largest possible earthquake magnitudes in mainland China based on extreme value theory, underscoring the significance of ground-based observations and statistical analyses in earthquake forecasting. Incorporating insights from studies on extreme value theory can help researchers refine their forecasting models and enhance the accuracy of earthquake forecasts in regions with high seismic activity.

Herrera et al.'s study [30] on long-term forecasting of strong earthquakes in various regions, including North America and South America, highlights the use of machine learning techniques to cluster earthquakes based on historical intervals with and without strong seismic events. This approach offers valuable insights into seismic patterns and can improve the predictive capabilities of earthquake forecasting models.

Michael's research [31] on testing forecasting methods for earthquake clustering versus the Poisson model stresses the importance of statistical techniques in evaluating the efficacy of earthquake forecasting methods. By comparing observed outcomes with random chance, researchers can gauge the success of different forecasting models and enhance their approaches to earthquake forecasting.

The work by Kodera et al. [32] on earthquake early warning systems for the 2016 Kumamoto earthquake in Japan provides insights into the performance evaluation of earthquake warning systems under heavy loading conditions. Understanding the effectiveness of early warning systems can guide the development of similar systems in earthquake-prone regions like Los Angeles to mitigate seismic risks.

Yuan et al.'s analysis [33] and forecasting of the SARIMA model for earthquakes in the Longmenshan Fault Zone offer a scientific basis for earthquake risk management and a practical approach to forecasting earthquake occurrence times. Leveraging advanced modeling techniques like SARIMA can enhance the accuracy of earthquake forecasts and improve disaster preparedness measures.

Hajikhodaverdikhan et al.'s study [34] on earthquake forecasting using meteorological data and particle filter-based support vector regression highlights the potential of intelligent analysis of historical meteorological datasets in earthquake forecasting. Integrating meteorological data into forecasting models can enhance the precision of earthquake forecasts and improve early warning systems.

Astuti et al.'s research [35] on investigating the characteristics of geoelectric field signals before earthquakes using adaptive STFT techniques underscores the importance of signal analysis on both normal days and the day of the earthquake for earthquake forecasting. These findings can serve as valuable input parameters for refining earthquake forecasting models and improving forecasting accuracy. Nishikawa's study [36] comparing statistical low-frequency earthquake activity models highlights the significance of quantifying and monitoring slow earthquake activity characteristics, as they may change before major earthquakes occur. Understanding these activity patterns can improve the effectiveness of earthquake forecasting models and lead to more accurate forecasts in earthquake-prone regions like Los Angeles.

The research by Nimmagadda and Dreher [37] on ontology-based data warehouse modeling and mining of earthquake data for forecasting analysis emphasizes the efficacy of data warehousing in earthquake forecasting analysis. Leveraging ontology-based approaches can enhance the efficiency and accuracy of earthquake forecasting models, aiding in disaster preparedness efforts.

Prasad et al.'s analysis [38] of earthquake magnitude detection using primary waves and secondary waves, based on the concept of an Early Earthquake Warning system, stresses the importance of analyzing ground motion through wave analysis for robust earthquake forecasting. Integrating insights from this study can enhance the reliability of earthquake forecasts in regions like Los Angeles.

Yang et al.'s research [39] on an automated regression pipeline approach for high-efficiency earthquake forecasting using LANL data highlights the complexity of data mining steps involved in earthquake forecasting. By streamlining data processing and model development through automated pipelines, researchers can improve the efficiency and accuracy of earthquake forecasting models.

Research by Zheng and Tao [40] underscores the importance of regional parameters in ground motion attenuation relationships, emphasizing the necessity of accurate geophysical data for robust earthquake forecasting. Studies like those by Hussain et al. [41] shed light on the relationship between b-values and seismic stress levels, offering insights that can aid in forecasting high-magnitude earthquakes.

Seismologists have also explored the use of diverse data sources, such as GPS data, ionospheric data, and outgoing longwave radiation, to enhance earthquake forecasting models. While Gitis et al. [42] emphasize the significance of seismological data in systematic earthquake forecasting systems, studies like those of Zhai et al. [43] delve into the detection of thermal anomalies in earthquake processes using non-seismic time series data, showcasing the multidisciplinary nature of earthquake forecasting research.

In the pursuit of advancing earthquake forecasting methodologies, researchers have also explored the potential of animal behavior as a precursor to seismic events. While traditional seismological approaches rely on instrumental data and geophysical parameters, investigations like those of Woith et al. [44] have examined the ability of animals to forecast earthquakes, highlighting the interdisciplinary nature of earthquake forecasting research.

In summary, the existing body of research provides extensive insights into the seismic activity, geological conditions, and fault dynamics in the Los Angeles region. Studies such as Olsen's investigation into ground motion amplification [7], Donnellan et al.'s analysis of fault systems [8], and Hauksson's work on regional seismicity [9] have collectively enhanced our understanding of seismic hazards. Additional research on geodetic data and ground deformation [10], as well as historical earthquake analysis [45], has been instrumental in shaping our approach. Despite these advancements, a critical gap remains in the ability to accurately forecast earthquake magnitudes with sufficient lead time to implement effective mitigation strategies. Current models often lack the integration of comprehensive, site-specific data and advanced forecasting algorithms. Our research endeavors to address this gap by employing sophisticated machine learning and neural network techniques to forecast earthquake magnitudes. We build upon the foundational work of numerous previous studies [7], [8], [9], [10], [45], striving to enhance the accuracy and reliability of seismic forecasts. We hope that our approach will contribute to improved earthquake preparedness and response efforts in the Los Angeles region, recognizing that this is a collaborative and ongoing endeavor within the scientific community.

III. DATASET

A. ORIGINAL DATASET

We used earthquake data from the Southern California Earthquake Data Center (SCEDC), maintained by the California Institute of Technology, as our dataset [46]. We selected the Los Angeles region with coordinates: Center Latitude

34.0522, Center Longitude -118.2437, Outer Radius (km) 50, between dates 2001-10-29 00:00:00 and 2024-05-27 00:00:00. We filtered for earthquakes with a magnitude above 2.0 [47].

The dataset consists of the following columns:

- **Date**: The date of the earthquake occurrence.
- **Time**: The time of the earthquake occurrence.
- **ET**: Event type identifier.
- **GT**: Geographical type identifier.
- **MAG**: Magnitude type and value.
- **M**: Magnitude value.
- **LAT**: Latitude of the earthquake epicenter.
- **LON**: Longitude of the earthquake epicenter.
- **DEPTH**: Depth of the earthquake.
- **Q**: Quality indicator.
- **EVID**: Event identifier.
- **NPH**: Number of phases used in the solution.
- **NGRM**: Number of grams.

B. DATA PREPARATION

Our dataset contains 1256 Local Magnitude (M_L), 18 Moment Magnitude (M_w), and 1 Revised Local Magnitude (M_{Lr}).

Local Magnitude (M_L) and Moment Magnitude (M_w) are two essential scales used in seismology to quantify the size and energy release of earthquakes. Local Magnitude, often referred to as Richter Magnitude, is a measure of the amplitude of seismic waves recorded on seismographs near the earthquake's epicenter. It provides a rapid assessment of an earthquake's size based on the amplitude of ground motion at a specific distance from the epicenter. Moment Magnitude, on the other hand, is a more modern and comprehensive scale that quantifies the seismic moment released during an earthquake, taking into account the fault area, slip, and rigidity. Moment Magnitude is considered a more accurate measure of an earthquake's size, especially for larger events and those occurring at greater depths. Revised Local Magnitude (M_{Lr}) is a refinement of the traditional Local Magnitude scale, aimed at improving accuracy by incorporating additional data and correction factors.

To ensure our dataset is comparable and to make better forecasts for future earthquakes, it is crucial to convert everything to Local Magnitude (M_L). Converting between Local Magnitude (M_L) and Moment Magnitude (M_w) is essential for seismic hazard assessment, earthquake monitoring, and research purposes. The conversion between M_L and M_w allows for consistency in earthquake magnitude reporting and facilitates comparisons between different seismic events. Various studies have established empirical relationships and conversion formulas to translate M_L values to M_w values based on regional characteristics, seismic data analysis, and geophysical parameters.

For instance, El-Aal et al. [48] presented a conversion relationship between M_w and M_L for earthquakes in Egypt, where they used specific formulas to convert M_L values to M_w values based on the earthquake's magnitude range. Similarly, Nazaruddin [49] highlighted the importance of converting different magnitude scales, including M_L into Moment Magnitude (M_w), to provide a unified magnitude scale for earthquake events. These conversion relationships are essential for creating consistent earthquake catalogs and conducting seismic hazard assessments.

Studies like Ou et al. [50] have detailed equations and methodologies for calculating Moment Magnitude (M_w) from seismic moments and other geophysical parameters. These approaches involve complex calculations based on the seismic moment release, fault characteristics, and earthquake source properties to derive accurate Moment Magnitude values. By utilizing these conversion methods, researchers can ensure standardized reporting of earthquake magnitudes and enhance the understanding of seismic events' energy release and potential impact.

According to the Southern California Earthquake Data Center (SCEDC) maintained by the California Institute of Technology [51], starting at the end of December 2015, SCSN began calculating an additional magnitude type, labeled Revised Local Magnitude (M_{Lr}), which is a revised Local Magnitude (M_L). M_{Lr} magnitudes are only calculated for events with M_L between 3.0 and 6.0 and are obtained by applying a linear adjustment to the M_L value. The adjustment is designed to bring initial magnitude values derived from M_L into closer agreement with Moment Magnitude (M_w), because M_w is expected to be the preferred magnitude type for events above magnitude 3 [51].

For most areas in southern California, M_L is systematically larger than M_w for magnitudes greater than 3.5. Consequently, the M_{Lr} adjustment is a reduction of the M_L value of up to 0.5 units (larger adjustment for larger events). M_{Lr} is calculated using the following formula [51]:

$$M_{Lr} = M_L \times 0.853 + 0.40125 \tag{1}$$

Solving for M_L, the formula becomes:

$$M_L = \frac{M_{Lr} - 0.40125}{0.853} \tag{2}$$

For the majority of earthquakes, M_L will be the preferred magnitude for events smaller than 3.5, and M_w preferred for events greater than 3.5 [51].

For our dataset from the California Earthquake Data Center (SCEDC), we converted every magnitude type to M_L using SCEDC's own formulas to ensure consistency and facilitate comparative analysis.

C. EXPLORATORY DATA ANALYSIS

To gain insights into the characteristics and patterns of earthquake data in Los Angeles, we performed an exploratory data analysis (EDA). We generated and analyzed the following graphs:

- **Distribution of Earthquake Magnitudes**: This histogram shows the frequency distribution of earthquake magnitudes. It helps in understanding the common

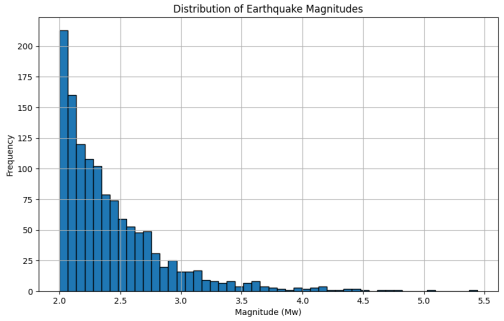


FIGURE 1. Distribution of earthquake magnitudes.

magnitude ranges of earthquakes in the region, as seen in Fig. 1.

- **Depth vs Magnitude Scatter Plot**: This scatter plot illustrates the relationship between the depth of earthquakes and their magnitudes. It helps in identifying any patterns or correlations between these two variables, as seen in Fig. 2.

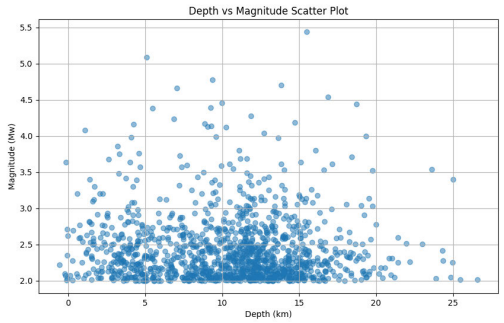


FIGURE 2. Depth vs magnitude.

- **Earthquake Count Over Time**: This line plot shows the number of earthquakes over time, aggregated monthly. It helps in identifying trends, seasonality, or any unusual activity over the analyzed period, as seen in Fig. 3.

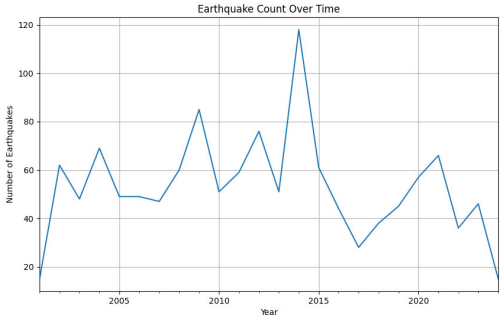


FIGURE 3. Earthquake count over time.

- **Geographic Distribution of Earthquakes**: This scatter plot maps the geographical distribution of earthquakes, with the magnitude represented by color. It helps in identifying the locations with higher seismic activity, as seen in Fig. 4.

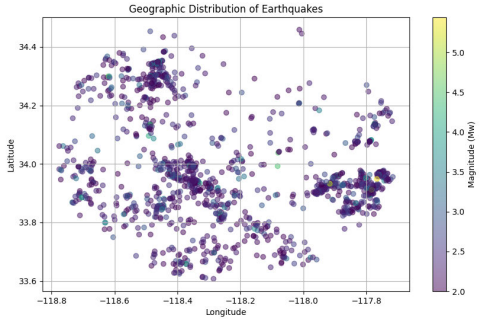


FIGURE 4. Geographic distribution of earthquakes.

IV. FEATURE ENGINEERING

To enhance the dataset and improve its forecasting power, we developed new features:

A. MAXIMUM MAGNITUDE OF NEXT SEISMIC EVENT IN THE NEXT 30 DAYS

In this feature engineering step, we calculated the maximum magnitude of the next seismic event occurring within the next thirty days for each earthquake event. This feature, named *max magnitude for the next 30 days*, helps in understanding the potential magnitude of aftershocks or subsequent earthquakes in the short term.

The distribution of the *max magnitude for the next 30 days* feature is shown in Fig. 5.

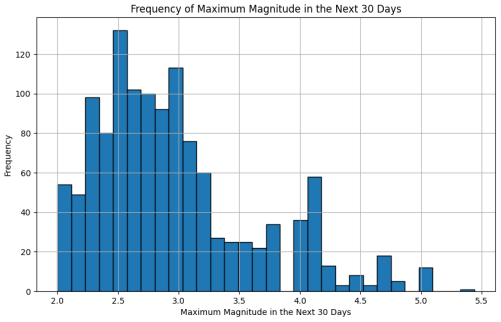


FIGURE 5. Distribution of maximum magnitude of next seismic event in the next 30 days.

To transform the earthquake forecasting problem into a classification problem, we determined target classes based on the magnitude of the most massive earthquake occurring within the next 30 days. This results in six distinct classes. The classification of the target variable is essential for applying machine learning-based models to forecast seismic activity effectively.

Previous studies indicated that an imbalanced dataset, resulting from improper classification of the target variable, can significantly diminish the performance of machine

learning models in earthquake forecasting [52], [53], [54]. To address this issue, we utilized the frequency distribution of the target variable to specify the intervals, ensuring a balanced dataset, with our target variable for ML and NN methods being the *max magnitude for the next 30 days*.

The Natural Breaks classification method, also known as Jenks optimization, was employed to determine the class boundaries. This method is specifically designed to identify natural groupings and patterns in data by minimizing variance within classes and maximizing variance between classes [55], [56].

The Natural Breaks classification method is a data clustering technique designed to determine the optimal arrangement of values into distinct classes based on the natural breaks in the data distribution [57]. This method aims to identify class boundaries that best group similar values together, minimizing the variation within each class while maximizing the differences between classes [58]. By utilizing the Jenks optimization algorithm, researchers can effectively classify data into a user-defined number of ranges, ensuring that the intervals for the target variable accurately reflect the underlying data distribution [59].

The Jenks Natural Breaks optimization method is particularly valuable in creating meaningful and interpretable classes for the target variable, allowing for a nuanced understanding of the data patterns [60]. By leveraging this classification technique, researchers can ensure that the classification thresholds are optimized to capture the inherent variability in the dataset, leading to more accurate and insightful analyses [61]. The Jenks optimization method is known for its ability to minimize within-group distances between values while maximizing the separation between different classes, resulting in well-defined and meaningful classification boundaries [62].

Moreover, the Jenks Natural Breaks classification method is based on the principle of natural grouping of data to minimize variation within classes and maximize differences between classes [56]. This approach ensures that the classification of data is optimized to reflect the inherent structure of the dataset, enhancing the interpretability and reliability of the results [54]. By employing the Jenks optimization method, researchers can effectively identify breakpoints between classes, facilitating a more nuanced and accurate classification of the target variable [53].

We applied the Natural Breaks classification method to ensure that the intervals for the target variable were determined in a way that reflected the natural distribution of the data. This resulted in a more balanced dataset, which was crucial for improving the performance and reliability of our machine learning models in earthquake forecasting.

The classes for the target variable, *max magnitude for the next 30 days*, were defined as follows: Class 1, Class 2, Class 3, Class 4, Class 5, and Class 6.

As seen in Table 1, the distribution of these classes is determined using the Natural Breaks method to ensure balanced representation across the dataset.

TABLE 1. Natural breaks classification for maximum magnitude of next seismic event in the next 30 days.

Class	Boundaries	# of Events
Class 1	0.00 - 2.38	251
Class 2	2.38 - 2.79	390
Class 3	2.79 - 3.26	344
Class 4	3.26 - 3.80	133
Class 5	3.80 - 4.46	118
Class 6	4.46 - 5.44	39

B. TIME SINCE LAST EARTHQUAKE

In this step of feature engineering, we calculated the time since the last earthquake for each event. This feature helps in understanding the temporal spacing between consecutive earthquakes, providing insights into the frequency and recurrence patterns of seismic activity in the Los Angeles region. The calculation was performed by sorting the dataset chronologically and computing the difference in time between each earthquake and the previous one. This new column is crucial for temporal analysis and can potentially reveal patterns or trends in earthquake occurrences over time.

C. CALCULATION OF THE GUTENBERG-RICHTER b-VALUE

In this subsection, we delve into the detailed methodology for calculating the Gutenberg-Richter b-value, a parameter that plays a critical role in seismology for understanding the distribution of earthquake magnitudes. The Gutenberg-Richter law is expressed as:

$$N(M) = 10^{a-bM} \tag{3}$$

where $N(M)$ is the number of events of Local Magnitude $\geq M$, a and b are constants, and b is known as the b-value [45].

1) LEAST-SQUARES METHOD

The least-squares method is employed to estimate the b-value by linearizing the Gutenberg-Richter law. By taking the logarithm of both sides of the equation, we obtain a linear relationship:

$$\log_{10} N(M) = a - bM \tag{4}$$

Given a dataset of earthquake magnitudes, we use the least-squares regression to find the best-fit line, where the slope of the line corresponds to $-b$. The equations for the least-squares estimation are derived as follows:

First, rewrite the linear form of the Gutenberg-Richter law:

$$Y = a - bM \tag{5}$$

where $Y = \log_{10} N(M)$.

The least-squares method minimizes the sum of the squared differences between the observed values and the values forecasted by the model. The slope m (which corresponds to $-b$) and the intercept c (which corresponds

to a) of the best-fit line are given by:

$$m = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2} \quad (6)$$

$$c = \frac{\sum_{i=1}^{N} y_i - m\sum_{i=1}^{N} x_i}{N} \quad (7)$$

In our case, $x_i = M_i$ and $y_i = \log_{10} N_i$. So, the slope m corresponds to $-b$:

$$-b = \frac{N\sum_{i=1}^{N} M_i \log_{10} N_i - \sum_{i=1}^{N} M_i \sum_{i=1}^{N} \log_{10} N_i}{N\sum_{i=1}^{N} M_i^2 - (\sum_{i=1}^{N} M_i)^2} \quad (8)$$

where N is the number of data points, M_i are the magnitudes, and $\log_{10} N_i$ are the logarithms of the cumulative number of events [63].

2) MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Despite the utility of the least-squares method, it is often preferable to use Maximum Likelihood Estimation (MLE) for the b-value due to its robustness, especially in dealing with infrequent large-magnitude earthquakes [64]. The MLE for the b-value is given by:

$$b = \frac{\log_{10} e}{\bar{M} - M_{\min}} \quad (9)$$

where $\bar{M}$ is the mean magnitude and $M_{\min}$ is the minimum magnitude in the dataset.

To derive the formula for the b-value using Maximum Likelihood Estimation (MLE), we start with the Gutenberg-Richter law:

$$N(M) = 10^{a-bM} \quad (10)$$

Taking the logarithm of both sides, we get:

$$\log_{10} N(M) = a - bM \quad (11)$$

To use MLE, we need to derive the likelihood function for the b-value. The likelihood function is based on the probability density function (PDF) of the earthquake magnitudes.

Step 1: Probability Density Function (PDF)

The cumulative distribution function (CDF) of magnitudes greater than or equal to M is:

$$F(M) = 10^{-b(M-M_{\min})} \quad (12)$$

where $M_{\min}$ is the minimum magnitude in the dataset.

The PDF is obtained by differentiating the CDF with respect to M:

$$f(M) = \frac{dF(M)}{dM} = -b \cdot 10^{-b(M-M_{\min})} \cdot \ln(10) \quad (13)$$

Since we are considering magnitudes $M \geq M_{\min}$, the PDF simplifies to:

$$f(M) = b \cdot 10^{-b(M-M_{\min})} \cdot \ln(10) \quad (14)$$

Step 2: Likelihood Function

Given a set of N earthquake magnitudes $\{M_1, M_2, \ldots, M_N\}$, the likelihood function $L(b)$ is the product of the individual probabilities:

$$L(b) = \prod_{i=1}^{N} f(M_i) = \prod_{i=1}^{N} \left(b \cdot 10^{-b(M_i-M_{\min})} \cdot \ln(10)\right) \quad (15)$$

Taking the natural logarithm of the likelihood function to obtain the log-likelihood function, we get:

$$\ln L(b) = \sum_{i=1}^{N} \ln\left(b \cdot 10^{-b(M_i-M_{\min})} \cdot \ln(10)\right) \quad (16)$$

$$\ln L(b) = \sum_{i=1}^{N} (\ln b + \ln\ln(10) - b(M_i - M_{\min})\ln(10)) \quad (17)$$

$$\ln L(b) = N\ln b + N\ln\ln(10) - b\ln(10)\sum_{i=1}^{N}(M_i - M_{\min}) \quad (18)$$

Step 3: Maximizing the Log-Likelihood

To find the maximum likelihood estimate of b, we differentiate the log-likelihood function with respect to b and set the derivative to zero:

$$\frac{d\ln L(b)}{db} = \frac{N}{b} - \ln(10)\sum_{i=1}^{N}(M_i - M_{\min}) = 0 \quad (19)$$

Solving for b:

$$b = \frac{N}{\ln(10)\sum_{i=1}^{N}(M_i - M_{\min})} \quad (20)$$

Since $\bar{M}$ is the mean magnitude:

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i \quad (21)$$

We can rewrite the summation as:

$$\sum_{i=1}^{N}(M_i - M_{\min}) = N(\bar{M} - M_{\min}) \quad (22)$$

Substituting this back into the equation for b:

$$b = \frac{N}{\ln(10) \cdot N(\bar{M} - M_{\min})} \quad (23)$$

Simplifying, we get:

$$b = \frac{1}{\ln(10)(\bar{M} - M_{\min})} \quad (24)$$

Since $\log_{10} e = \frac{1}{\ln(10)}$, we finally obtain:

$$b = \frac{\log_{10} e}{\bar{M} - M_{\min}} \quad (25)$$

To apply MLE, we followed these steps:

1. Calculated the mean magnitude $\bar{M}$:

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i \quad (26)$$

2. Used the MLE formula to compute the b-value:

$$b = \frac{\log_{10} e}{\bar{M} - M_{\min}} \tag{27}$$

Maximum Likelihood Estimation (MLE) is a statistical technique that offers a more robust and accurate approach to estimating the Gutenberg-Richter b-value, particularly in situations where there are limited data points or when dealing with rare, large-magnitude earthquakes [65]. MLE involves finding the parameter values that maximize the likelihood of observing the given earthquake data, taking into account the uncertainties associated with the observations. By utilizing MLE, researchers can obtain more reliable estimates of the b-value and better capture the underlying seismicity patterns in a region.

The Least-Squares Method, while simpler and easier to implement, does not match the robustness of Maximum Likelihood Estimation in handling data with varying uncertainties and complexities. It makes MLE a valuable tool for accurately estimating the Gutenberg-Richter b-value in seismicity analysis [65]. Selecting the appropriate method based on the characteristics of the earthquake data and the research objectives helps researchers obtain more precise and reliable estimates of the b-value, enhancing the understanding of seismic activity and earthquake hazard assessment.

3) ADDITION TO DATASET

We used the MLE method to calculate the b-value and added the new feature column, *b value*, to our dataset. This column represents the calculated b-value for each time window of analysis [66]. We used the fifty events that occurred prior to each event to calculate the b-value.

D. CALCULATION OF INCREMENTAL b-VALUES

We first calculated the b-values using the fifty events that occurred prior to each event, as detailed in previous studies [67], [68]. This methodology allows us to track changes in seismicity over time, which serve as forecasting features for seismic activity analysis. After obtaining the b-values, we calculated the incremental b-values by determining the differences between b-values over various time windows, specifically between events i and $i-2$, $i-2$ and $i-4$, $i-4$ and $i-6$, $i-6$ and $i-8$, and $i-8$ and $i-10$.

The study by Volant et al. [67] titled "b-Value, aseismic deformation and brittle failure within an isolated geological object: Evidences from a dome structure loaded by fluid extraction" published in Geophysical Research Letters in 1992, explores the relationship between seismic activity, aseismic deformation, and brittle failure within a geological structure subjected to fluid extraction. This study investigates the induced seismic activity and aseismic displacements resulting from gas extraction in an area previously devoid of displacement, shedding light on the impact of fluid extraction on seismicity and deformation processes [67].

The study offers a unique perspective on the relationship between fluid extraction, seismicity, and deformation, providing valuable insights that could inform the development of forecasting models incorporating incremental b-values derived from seismic data analysis over time [67].

The study by Yousefzadeh et al. [68] titled "Spatiotemporally explicit earthquake forecasting using deep neural network" published in Soil Dynamics and Earthquake Engineering in 2021, investigates the effect of spatial parameters on the performance of machine learning algorithms for forecasting the magnitude of future earthquakes in Iran. This study compares the performance of conventional methods such as Support Vector Machine (SVM), Decision Tree (DT), and Shallow Neural Network (SNN) with a contemporary Deep Neural Network (DNN) method. One of the key parameters introduced in this study is the Fault Density (FD), which, along with incremental b-values, enhances the accuracy of earthquake forecasting models.

The results showed that incremental b-values, which measure the change in seismicity over time, significantly contribute to the forecasting accuracy of earthquakes. The study highlights the importance of using both temporal and spatial parameters, including incremental b-values, in developing robust forecasting models for seismic activity [68].

We detailed the methodology for calculating the incremental b-values, which served as forecasting features for seismic activity analysis. We derived the incremental b-values from the differences in b-values calculated over various time windows.

1) b-VALUE INCREMENTS BETWEEN EVENTS i AND $i-2$

We calculated the b-value increment between events i and $i-2$ as follows:

$$\Delta b_{i,i-2} = b_i - b_{i-2} \tag{28}$$

where b_i is the b-value at event i and b_{i-2} is the b-value at event $i-2$.

2) b-VALUE INCREMENTS BETWEEN EVENTS $i-2$ AND $i-4$

We calculated the b-value increment between events $i-2$ and $i-4$ as follows:

$$\Delta b_{i-2,i-4} = b_{i-2} - b_{i-4} \tag{29}$$

where b_{i-2} is the b-value at event $i-2$ and b_{i-4} is the b-value at event $i-4$.

3) b-VALUE INCREMENTS BETWEEN EVENTS $i-4$ AND $i-6$

We calculated the b-value increment between events $i-4$ and $i-6$ as follows:

$$\Delta b_{i-4,i-6} = b_{i-4} - b_{i-6} \tag{30}$$

where b_{i-4} is the b-value at event $i-4$ and b_{i-6} is the b-value at event $i-6$.

4) b-VALUE INCREMENTS BETWEEN EVENTS $i-6$ AND $i-8$

We calculated the b-value increment between events $i-6$ and $i-8$ as follows:

$$\Delta b_{i-6,i-8} = b_{i-6} - b_{i-8} \tag{31}$$

where b_{i-6} is the b-value at event $i-6$ and b_{i-8} is the b-value at event $i-8$.

5) b-VALUE INCREMENTS BETWEEN EVENTS $i-8$ AND $i-10$

We calculated the b-value increment between events $i-8$ and $i-10$ as follows:

$$\Delta b_{i-8,i-10} = b_{i-8} - b_{i-10} \tag{32}$$

where b_{i-8} is the b-value at event $i-8$ and b_{i-10} is the b-value at event $i-10$.

These incremental b-values were crucial for understanding the temporal variations in seismicity and have been shown to be effective forecasting features in recent studies [68].

E. FAULT LINE EQUATIONS AND PERPENDICULAR DISTANCE CALCULATION

The proximity to fault lines is a predictor of future seismic events. Research has shown that seismic activity and earthquake occurrences can be influenced by the distance to fault lines, with seismicity often being more prevalent in areas closer to active faults. Understanding the distance to fault lines can offer valuable insights into the potential for future seismic events and the seismic hazard level in a given region.

One study supporting the relationship between seismic activity and distance from fault lines is the research by Dieterich and Smith [69]. This study found that the number of earthquakes concerning the distance from major faults in southern California follows a power-law decay to distances of 15 km, with decay exponent values around -1.5 [69].

1) FAULT LINES

Fig. 6 illustrates the geographical distribution of all earthquakes in our dataset along with the major fault lines in the Los Angeles region. The fault lines included in the figure are the San Andreas Fault, Newport-Inglewood Fault, Whittier Fault, Puente Hills Thrust Fault, Raymond Fault, and Sierra Madre Fault Zone. Each fault line is depicted in a different color for clarity.

The earthquakes are represented by colored dots, with each color corresponding to a specific magnitude range:

- **Class 1 (0.00 - 2.38)**: Blue
- **Class 2 (2.38 - 2.79)**: Green
- **Class 3 (2.79 - 3.26)**: Yellow
- **Class 4 (3.26 - 3.80)**: Orange
- **Class 5 (3.80 - 4.46)**: Red
- **Class 6 (4.46 - 5.44)**: Red

a: SAN ANDREAS FAULT

The San Andreas Fault can be approximated by a straight line in a 2D coordinate system. Using the coordinates

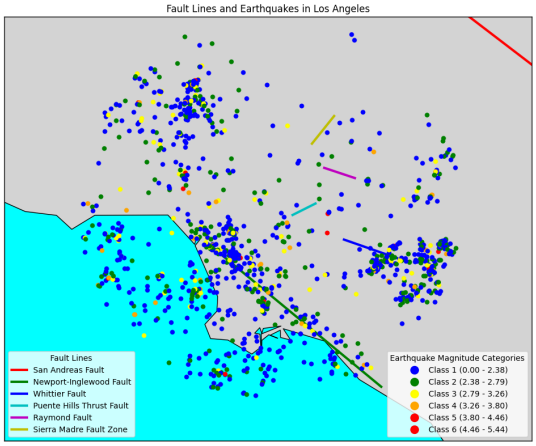


FIGURE 6. Fault lines and earthquakes in Los Angeles.

(35.768, −119.703) and (33.021, −115.354), we derive the equation of the line as follows:

Given two points (x_1, y_1) and (x_2, y_2), the line equation can be written as:

$$(y_2 - y_1)x - (x_2 - x_1)y + (x_2 y_1 - x_1 y_2) = 0 \tag{33}$$

Plugging in the values, we get:

$$-2.747x + 4.349y - 205.684 = 0 \tag{34}$$

b: NEWPORT-INGLEWOOD FAULT

The Newport-Inglewood Fault runs from the Westside of Los Angeles down through the Orange County coast. Using the coordinates (33.949, −118.395) and (33.628, −117.928), we derive the line equation as follows:

$$-0.321x - 0.467y + 209.690 = 0 \tag{35}$$

c: WHITTIER FAULT

The Whittier Fault runs from Whittier to the Puente Hills and can produce significant earthquakes. Using the coordinates (33.976, −118.034) and (33.927, −117.865), we derive the line equation as follows:

$$-0.049x - 0.169y + 198.878 = 0 \tag{36}$$

d: PUENTE HILLS THRUST FAULT

The Puente Hills Thrust Fault is located beneath the central Los Angeles basin and is considered capable of generating significant earthquakes. Using the coordinates (34.034, −118.180) and (34.061, −118.115), we derive the line equation as follows:

$$0.027x - 0.065y + 401.665 = 0 \tag{37}$$

e: RAYMOND FAULT

The Raymond Fault runs through the San Gabriel Valley and into the Los Angeles basin. Using the coordinates

(34.145, −118.090) and (34.121, −118.003), we derive the line equation as follows:

$$-0.024x - 0.087y + 296.526 = 0 \tag{38}$$

f: SIERRA MADRE FAULT ZONE

The Sierra Madre Fault Zone runs along the base of the San Gabriel Mountains and poses a risk to the northern Los Angeles area. Using the coordinates (34.202, −118.125) and (34.267, −118.062), we derive the line equation as follows:

$$0.065x - 0.063y + 403.566 = 0 \tag{39}$$

2) PERPENDICULAR DISTANCE FROM A POINT TO THE LINE

Given a point (x_0, y_0), the perpendicular distance d from the point to the line $Ax + By + C = 0$ is given by:

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} \tag{40}$$

3) IMPLEMENTATION

We have implemented this calculation in our dataset as follows:

i **San Andreas Fault**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the San Andreas Fault line using the formula:

$$d = \frac{|-2.747x_0 + 4.349y_0 - 205.684|}{\sqrt{2.747^2 + 4.349^2}} \tag{41}$$

ii **Newport-Inglewood Fault**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the Newport-Inglewood Fault line using the formula:

$$d = \frac{|-0.321x_0 - 0.467y_0 + 209.690|}{\sqrt{0.321^2 + 0.467^2}} \tag{42}$$

iii **Whittier Fault**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the Whittier Fault line using the formula:

$$d = \frac{|-0.049x_0 - 0.169y_0 + 198.878|}{\sqrt{0.049^2 + 0.169^2}} \tag{43}$$

iv **Puente Hills Thrust Fault**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the Puente Hills Thrust Fault line using the formula:

$$d = \frac{|0.027x_0 - 0.065y_0 + 401.665|}{\sqrt{0.027^2 + 0.065^2}} \tag{44}$$

v **Raymond Fault**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the Raymond Fault line using the formula:

$$d = \frac{|-0.024x_0 - 0.087y_0 + 296.526|}{\sqrt{0.024^2 + 0.087^2}} \tag{45}$$

vi **Sierra Madre Fault Zone**: For each point (x_0, y_0) in the dataset, we calculate the perpendicular distance to the Sierra Madre Fault Zone line using the formula:

$$d = \frac{|0.065x_0 - 0.063y_0 + 403.566|}{\sqrt{0.065^2 + 0.063^2}} \tag{46}$$

These formulas are applied to each data point in the dataset to compute the perpendicular distances to each of the fault lines. This allows us to analyze the proximity of earthquake events to the major fault lines in the Los Angeles region, providing insights into potential correlations between fault proximity and earthquake characteristics.

F. MAXIMUM MAGNITUDE RECORDED DURING THE LAST WEEK

The maximum magnitude recorded during the last week is a crucial feature for assessing recent seismic activity and the potential for future earthquakes. Martinsson and Törnman [70] provide insights into the relationship between induced seismic activity and production rates, depth, and size within a mining context. The study highlights that high seismic activity in a given week can increase the likelihood of elevated seismicity in the subsequent week, emphasizing the importance of monitoring and analyzing seismic events over short time intervals to assess evolving seismic activity patterns.

Bohnhoff et al. [71] discuss seismicity patterns following the Gutenberg-Richter law, indicating that a high-magnitude seismic event can be preceded by foreshocks. Monitoring the maximum magnitude recorded in a given period can offer valuable insights into the potential for larger seismic events.

Zhang et al. [3] focus on feature extraction techniques for earthquake prediction, emphasizing the importance of identifying precursory patterns in seismic data. This study highlights the significance of monitoring the maximum magnitude as a key feature for forecasting seismic events and understanding seismic activity trends.

Asim et al. [2] delve into earthquake prediction models using support vector regressor and hybrid neural networks, showcasing the capability of these methodologies in forecasting seismic events of specific magnitudes. The study underscores the importance of advanced forecasting techniques in assessing seismic hazards and the potential impact of earthquakes based on their magnitudes.

In conclusion, leveraging insights from studies emphasizing the dynamic nature of seismicity, the Gutenberg-Richter law, and advanced forecast models can enhance our understanding of seismic patterns and improve earthquake forecasts based on the maximum magnitude data recorded over specific time intervals.

This metric provides a snapshot of the largest seismic event within a short, fixed time window, which can be indicative of the stress accumulation and release in the region. This approach has been utilized in recent forecasting studies to enhance the accuracy of earthquake forecasting models.

1) CALCULATION METHODOLOGY

To calculate the maximum magnitude recorded during the last week, we followed these steps:

i Defined the Time Window: Considered a sliding window of 7 days (one week) for each event in the dataset.

ii Identified Relevant Events: For each event i, identified all seismic events that occurred within the 7 days prior to the event i.
iii Determined the Maximum Magnitude: Calculated the maximum magnitude from the identified events.

Mathematically, this can be expressed as:

$$M_{\text{max}}^{\text{last week}} = \max\{M_j \mid t_i - 7 \text{ days} \leq t_j < t_i\} \tag{47}$$

where $M_{\text{max}}^{\text{last week}}$ is the maximum magnitude recorded during the last week, M_j is the magnitude of event j, t_i is the time of event i, and t_j is the time of event j.

2) IMPLEMENTATION

In the dataset, we performed this calculation for each event, resulting in a new column that recorded the maximum magnitude observed in the week preceding each event.

G. PROBABILITY OF EVENTS WITH MAGNITUDE ≥ 5.0

The research by Zhang et al. [3] on "Precursory Pattern Based Feature Extraction Techniques for Earthquake Prediction" in IEEE Access focuses on feature extraction methods for earthquake forecasting, emphasizing the importance of identifying precursory patterns in seismic data. By incorporating the probability of high-magnitude seismic events as a feature, we can enhance the predictive capabilities of machine learning algorithms and improve the accuracy of earthquake forecasts.

The probability of events with a magnitude greater than or equal to 5.0 ($P(M \geq 5.0)$) is an important predictive feature in seismology. This probability is derived from the Gutenberg-Richter law, which describes the relationship between the magnitude and total number of earthquakes. The data with the maximum magnitude (5.44) come from an event that occurred on 2008-07-29 at 18:42:15.670Z, with a latitude of 33.9485, longitude of -117.766333, and a depth of 15.503 km. Recent studies have demonstrated the utility of this probability in improving earthquake forecasting models.

1) MATHEMATICAL FOUNDATION

The probability of an earthquake having a magnitude greater than or equal to a specific value M can be expressed as:

$$P(M \geq M_s) = 10^{-(b(M_s - M_{\text{min}}))} \tag{48}$$

where M_s is the specified magnitude threshold (in this case, 5.0), and M_{min} is the minimum magnitude in the dataset.

2) DERIVATION OF THE FORMULA

The Gutenberg-Richter law is given by:

$$\log_{10} N(M) = a - bM \tag{49}$$

where: - $N(M)$ is the cumulative number of earthquakes with magnitude greater than or equal to M. - a and b are constants.

This can be rewritten as:

$$N(M) = 10^{a-bM} \tag{50}$$

The cumulative distribution function (CDF), $F(M)$, represents the probability that an earthquake has a magnitude greater than or equal to M. Using the Gutenberg-Richter law, we get:

$$F(M) = \frac{N(M)}{N(M_{\text{min}})} \tag{51}$$

where M_{min} is the minimum magnitude in the dataset, and $N(M_{\text{min}})$ is the total number of earthquakes in the dataset.

Substituting the Gutenberg-Richter law, we have:

$$F(M) = \frac{10^{a-bM}}{10^{a-bM_{\text{min}}}} \tag{52}$$

Simplifying the expression:

$$F(M) = 10^{a-bM-(a-bM_{\text{min}})} = 10^{-b(M-M_{\text{min}})} \tag{53}$$

Thus, the probability of an earthquake having a magnitude greater than or equal to M is:

$$P(M \geq M_s) = 10^{-b(M_s - M_{\text{min}})} \tag{54}$$

where M_s is the specified magnitude threshold and M_{min} is the minimum magnitude in the dataset.

3) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Calculate the b-value: We used the maximum likelihood estimation (MLE) method.
ii Compute the Probability: We applied the derived formula to compute $P(M \geq M_s)$ for each event in the dataset.

H. GUTENBERG-RICHTER a-VALUE

The a-value in the Gutenberg-Richter law is a crucial parameter that represents the seismic activity rate in a region. It indicates the overall productivity of earthquakes and is used in conjunction with the b-value to describe the frequency-magnitude distribution of seismic events. Recent studies have highlighted the importance of accurately determining the a-value for improved seismic hazard assessment [68].

1) MATHEMATICAL FOUNDATION

The a-value can be determined from the linear relationship obtained from plotting the logarithm of the cumulative number of events against the magnitude. Given a set of magnitudes and their cumulative counts, the a-value is calculated as the intercept of the regression line on the $\log_{10} N(M)$ axis.

Mathematically, the a-value can be expressed as:

$$a = \log_{10} N + b\overline{M} \tag{55}$$

where: - $\log_{10} N$ is the logarithm of the total number of earthquakes, - $\overline{M}$ is the mean magnitude of the earthquakes.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Calculated the b-value: We used the maximum likelihood estimation (MLE) method or another appropriate method.
ii Computed the total number of earthquakes and their mean magnitude within a given time window.
iii Applied the formula to determine the a-value.

I. SUM OF THE MEAN SQUARE DEVIATION (η) FROM THE REGRESSION LINE BASED ON THE GUTENBERG-RICHTER LAW

The sum of the mean square deviation (η) from the regression line based on the Gutenberg-Richter (GR) law is a valuable metric for assessing the fit of the observed earthquake data to the GR model. This metric helps in quantifying the variability and reliability of the seismic activity forecasts. Recent studies have demonstrated that η offers valuable insights into seismic forecasts [72].

The sum of the mean square deviation (η) from the regression line based on the Gutenberg-Richter law is a valuable predictor of future seismic events. By analyzing the deviations of observed seismicity data from the regression line defined by the Gutenberg-Richter law, researchers gain insights into the consistency of seismic activity patterns and the potential for future earthquakes. The study by Fahandezhsadi & Sadi [73] on "Earthquake Magnitude Forecasting using Probabilistic Classifiers" in 2020 explores the use of the sum of the mean square deviation about the regression line as a feature for earthquake magnitude forecasting, highlighting its significance in assessing seismic activity trends and forecasting future events.

1) MATHEMATICAL FOUNDATION

The mean square deviation from the regression line is calculated to measure how well the observed data fit the GR law. For a set of observed magnitudes $\{M_i\}$ and their corresponding cumulative counts $\{N_i\}$, the deviation for each magnitude is given by:

$$d_i = \log_{10} N_i - (a - bM_i) \tag{56}$$

The mean square deviation is then:

$$\eta = \frac{1}{n} \sum_{i=1}^{n} d_i^2 \tag{57}$$

where n is the number of data points.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Calculated the a-value and b-value using the observed data.
ii Computed the deviations d_i for each observed magnitude.
iii Calculated the mean square deviation (η) using the formula provided.

These steps help in assessing the goodness-of-fit of the observed data to the Gutenberg-Richter model, providing a quantitative measure of the model's reliability [72].

J. DIFFERENCE BETWEEN THE LARGEST OBSERVED MAGNITUDE AND LARGEST EXPECTED BASED ON THE GUTENBERG-RICHTER LAW

Saichev and Sornette's study on the "Distribution of the largest aftershocks in branching models of triggered seismicity" in Physical Review E discusses Båth's law, which empirically shows an average magnitude difference of 1.2, independent of the mainshock magnitude [74]. This reference underscores the importance of evaluating the difference between observed and expected magnitudes to understand seismic activity and forecast future events accurately.

The difference between the largest observed magnitude and the largest expected magnitude based on the Gutenberg-Richter (GR) law, denoted as ΔM, is an important metric for evaluating seismic hazard. This metric helps identify regions where the observed seismicity deviates from expected patterns, which is crucial for assessing the potential for large, unexpected earthquakes. Recent studies have shown that ΔM can provide significant insights into seismic hazard assessments [75].

1) MATHEMATICAL FOUNDATION

The largest expected magnitude (M_{expected}) can be estimated using the GR law by considering the total number of events and the b-value. The Gutenberg-Richter law is given by:

$$\log_{10} N(M) = a - bM \tag{58}$$

To find the largest expected magnitude, we consider the equation when $N(M) = 1$ (i.e., the magnitude at which we expect to see one event. The cumulative number of earthquakes with a magnitude greater than or equal to the greatest magnitude is one.):

$$\log_{10}(1) = a - bM_{\text{expected}} \tag{59}$$

Since $\log_{10}(1) = 0$, the equation simplifies to:

$$0 = a - bM_{\text{expected}} \tag{60}$$

Solving for M_{expected}, we get:

$$M_{\text{expected}} = \frac{a}{b} \tag{61}$$

The difference ΔM between the largest observed magnitude (M_{observed}) and the largest expected magnitude (M_{expected}) is then calculated as:

$$\Delta M = M_{\text{observed}} - M_{\text{expected}} \tag{62}$$

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Calculated the a-value and b-value using the observed data.

ii Determined the largest observed magnitude M_{observed} in the dataset.
iii Estimated the largest expected magnitude M_{expected} using the formula $M_{\text{expected}} = \frac{a}{b}$.
iv Computed ΔM using the formula provided.

K. ELAPSED TIME (T) BETWEEN THE LAST N EVENTS

Research by Faro et al. [76] in the Journal of Personality and Social Psychology explores the influence of causal relationships on time perception and judgments of elapsed time between events. This study emphasizes the role of causal associations in shaping temporal judgments, providing insights into the forecasting value of elapsed time between seismic events.

Incorporating the elapsed time between events as a forecasting feature in machine learning algorithms can enhance the models' ability to capture temporal dependencies in seismic activity. The study by Nguyen et al. [77] highlights the informative nature of elapsed time between events, suggesting its relevance in forecasting modeling and decision-making processes.

The elapsed time (T) between the last n events is a critical metric for understanding the temporal patterns in seismic activity. This metric helps identify periods of increased or decreased seismic activity.

1) MATHEMATICAL FOUNDATION

The elapsed time (T) is defined as the total time interval between the first and last event in a specified window of n events. For a given set of events, let t_1 be the time of the first event and t_n be the time of the n-th event. The elapsed time T is calculated as:

$$T = t_n - t_1 \tag{63}$$

where: - t_1 is the time of the first event in the window, - t_n is the time of the n-th event in the window.

This simple yet powerful metric can reveal changes in the seismic activity rate over time.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Extracted the times of the first (t_1) and last (t_n) events in the specified window of n events.
ii Calculated the elapsed time T using the formula $T = t_n - t_1$.

L. MEAN TIME BETWEEN EVENTS (μ)

The study by Salam et al. [78] on "Earthquake Prediction using Hybrid Machine Learning Techniques" in the International Journal of Advanced Computer Science and Applications includes the average time between events (μ) as one of the indicators used for earthquake forecasting. This reference highlights the importance of temporal features in predictive modeling and suggests that the mean time between events can be a valuable predictor for forecasting seismic events.

The mean time between events (μ) is an important metric for understanding the temporal distribution of seismic activity.

1) MATHEMATICAL FOUNDATION

The mean time between events (μ) is defined as the average time interval between consecutive earthquake events. For a given set of n events, let t_i be the time of the i-th event. The time interval between consecutive events is given by:

$$\Delta t_i = t_{i+1} - t_i \tag{64}$$

The mean time between events is then calculated as the average of these time intervals:

$$\mu = \frac{1}{n-1} \sum_{i=1}^{n-1} \Delta t_i \tag{65}$$

where: - t_i is the time of the i-th event, - Δt_i is the time interval between the i-th event and the $(i+1)$-th event, - n is the total number of events.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Extracted the times of the events in the specified window of n events.
ii Calculated the time intervals Δt_i between consecutive events.
iii Computed the mean time between events μ using the formula provided.

M. COEFFICIENT OF VARIATION (C)

The study by Rosenau & Oncken [79] in the Journal of Geophysical Research Atmospheres discusses the relationship between the coefficient of variation of recurrence intervals and seismic activity patterns in subduction zones. This reference highlights the importance of understanding the variability in recurrence intervals for forecasting seismic events and assessing the frequency-size distribution of earthquakes in different geological settings.

The coefficient of variation (C) is a normalized measure of the dispersion of the inter-event times in a set of earthquake occurrences. It is an important metric for understanding the variability and predictability of seismic activity. A higher coefficient of variation indicates more irregular and unpredictable seismic activity, whereas a lower coefficient suggests more regular and predictable occurrences.

1) MATHEMATICAL FOUNDATION

The coefficient of variation (C) is defined as the ratio of the standard deviation (σ) to the mean (μ) of the inter-event times. For a given set of n events, let Δt_i be the time interval between consecutive events. The mean (μ) and standard

deviation (σ) of these time intervals are given by:

$$\mu = \frac{1}{n-1}\sum_{i=1}^{n-1}\Delta t_i \tag{66}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(\Delta t_i - \mu)^2} \tag{67}$$

The coefficient of variation (C) is then calculated as:

$$C = \frac{\sigma}{\mu} \tag{68}$$

where: - μ is the mean time between events, - σ is the standard deviation of the time intervals, - n is the total number of events.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Extracted the time intervals Δt_i between consecutive events in the specified window of n events.
ii Calculated the mean time between events (μ) using the formula provided.
iii Calculated the standard deviation of the time intervals (σ).
iv Computed the coefficient of variation (C) using the formula provided.

N. THE SQUARE ROOT OF THE CUMULATIVE SEISMIC ENERGY

Salam et al. [78] conducted a study on "Earthquake Prediction using Hybrid Machine Learning Techniques" in the International Journal of Advanced Computer Science and Applications, which highlights the significance of energy-related features in earthquake forecasting models. The study includes the square root of the released energy during a specific time as one of the indicators for earthquake forecasting, supporting the notion that the rate of seismic energy release can be a valuable predictor for forecasting seismic events.

The rate of the square root of seismic energy is a metric used to quantify the energy released by seismic events. It provides a normalized measure of seismic activity by considering the energy release rate, which is important for understanding the dynamics of earthquake processes and assessing seismic hazards [68], [80].

1) MATHEMATICAL FOUNDATION

The seismic energy (E) released by an earthquake can be estimated using its magnitude (M) through the following relationship:

$$E = 10^{1.5M+4.8} \tag{69}$$

where: - E is the seismic energy in joules, - M is the magnitude of the earthquake.

To calculate the square root of the cumulative seismic energy ($\sqrt{E_\Sigma}$), we first compute the seismic energy for each event using the formula $10^{1.5M_i+4.8}$. Then, we sum these values over a specified window of $n = 50$ events and take the square root of the sum. The expression is given by:

$$\sqrt{E_\Sigma} = \sqrt{\sum_{i=1}^{50} 10^{1.5M_i+4.8}} \tag{70}$$

where: - M_i is the magnitude of the i-th event in the window.

2) IMPLEMENTATION

We implemented this calculation in our dataset as follows:

i Extracted the magnitudes M_i of the events in the specified window of $n = 50$ events.
ii Computed the seismic energy $E_i = 10^{1.5M_i+4.8}$ for each event.
iii Summed the seismic energy values.
iv Computed the square root of the cumulative seismic energy ($\sqrt{E_\Sigma}$) using the formula provided.

O. MEAN MAGNITUDE OF THE LAST N EVENTS (M_{mean})

The mean magnitude (M_{mean}) of the last n events is a straightforward yet powerful metric for characterizing the average size of recent earthquakes. This metric is crucial for understanding the general trend in seismic activity and for making short-term predictions about future seismic events [81], [82].

In their study on "Earthquake Prediction using Hybrid Machine Learning Techniques," Salam et al. [78] utilized the average magnitude of N events (M_{mean}) as an indicator for earthquake forecasting. This research highlights the significance of magnitude-related features in earthquake forecasting models and suggests that the mean magnitude of recent events can be a valuable predictor for forecasting seismic events.

1) MATHEMATICAL FOUNDATION

The mean magnitude (M_{mean}) is calculated as the arithmetic mean of the magnitudes of the last n events. For a given set of n events, let M_i be the magnitude of the i-th event. The mean magnitude is then given by:

$$M_{\text{mean}} = \frac{1}{n}\sum_{i=1}^{n} M_i \tag{71}$$

where: - M_i is the magnitude of the i-th event, - n is the total number of events considered.

2) IMPLEMENTATION

We implement this calculation in our dataset as follows:

i Extract the magnitudes M_i of the events in the specified window of n events.
ii Sum the magnitudes of these n events.
iii Compute the mean magnitude (M_{mean}) using the formula provided.

V. METHODOLOGY

A. EARTHQUAKE FORECASTING USING MACHINE LEARNING AND NEURAL NETWORKS

In this study, we evaluated various machine learning (ML) algorithms and neural network (NN) models to forecast the class of earthquake magnitudes within the next 30 days. The models were trained and tested using a dataset of seismic events with features scaled for optimal performance. We utilized a dataset from the Southern California Earthquake Data Center (SCEDC) for the Los Angeles region, covering earthquakes from 2001 to 2024. The dataset includes a range of features engineered to enhance predictive power. The features used in our analysis are summarized in Table 2. By integrating these features, we aim to develop robust machine learning models capable of forecasting the class of future seismic events in the Los Angeles region.

For repeatability, we used a **random state** of 15 for each of the methodologies. The random state ensures that the results are reproducible by setting the seed for the random number generator used in the algorithms. Also, our test sample was 20% throughout the project. Our methodology is shown in Fig. 7.

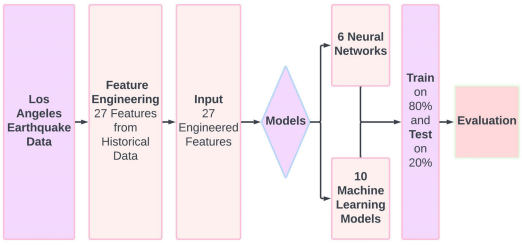


FIGURE 7. Methodology for earthquake forecasting.

Our Machine Learning (ML) algorithms included Logistic Regression, Decision Trees, Random Forest, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Naive Bayes, AdaBoost, XGBoost, and LightGBM. These models are based on statistical methods and mathematical algorithms to make predictions based on input data.

Our Neural Networks (NN) included Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), Gated Recurrent Units (GRU), and Transformer Models. These models are inspired by the structure and function of the human brain, using layers of interconnected nodes (neurons) to learn from data.

1) XGBoost

We applied XGBoost to forecast earthquake categories, and the resulting accuracy is shown in Table 3. XGBoost, which stands for Extreme Gradient Boosting, is a machine learning algorithm that has gained popularity across various fields due to its efficiency and effectiveness in handling complex datasets. In the field of seismology, XGBoost shows promise in earthquake forecasting and analysis. Seismologists are exploring the application of machine learning techniques like XGBoost to enhance their ability to forecast seismic events accurately and efficiently [83]. By utilizing XGBoost, researchers have developed models capable of forecasting earthquakes by analyzing various features and patterns in seismic data [83]. This approach signifies a shift towards more advanced and data-driven methodologies in seismology, aiming to improve the accuracy and timeliness of earthquake forecasts.

TABLE 2. Our input variables for earthquake forecasting.

No.	Variable	Explanation
1	latitude	Latitude of the earthquake epicenter
2	longitude	Longitude of the earthquake epicenter
3	depth	Depth of the earthquake
4	mag	Magnitude of the earthquake
5	time since last earthq.	Time elapsed since the last earthquake
6	b value	Gutenberg-Richter b-value for the magnitude
7	$\Delta b_{i,i-2}$	Incremental b-value between events i and $i-2$
8	$\Delta b_{i-2,i-4}$	Incremental b-value between events $i-2$ and $i-4$
9	$\Delta b_{i-4,i-6}$	Incremental b-value between events $i-4$ and $i-6$
10	$\Delta b_{i-6,i-8}$	Incremental b-value between events $i-6$ and $i-8$
11	$\Delta b_{i-8,i-10}$	Incremental b-value between events $i-8$ and $i-10$
12	Dist. to San Andreas Fault	Distance to the San Andreas Fault
13	Dist. to Newport Inglewood Fault	Distance to the Newport-Inglewood Fault
14	Dist. to Whittier Fault	Distance to the Whittier Fault
15	Dist. to Puente Hills Thrust Fault	Distance to the Puente Hills Thrust Fault
16	Dist. to Raymond Fault	Distance to the Raymond Fault
17	Dist. to Sierra Madre Fault Zone	Distance to the Sierra Madre Fault
18	$M_{\max}^{\text{last week}}$	Maximum magnitude recorded during the last week
19	$P(M \geq 5)$	Probability of an earthquake with magnitude ≥ 5.0
20	a	Gutenberg-Richter a-value
21	η	Sum of the mean square deviation from the regression line based on the Gutenberg-Richter law
22	ΔM	Difference between the largest observed magnitude and the largest expected magnitude based on the Gutenberg-Richter law
23	T	Elapsed time between the last n events
24	μ	Mean time between events
25	C	Coefficient of variation of the inter-event times
26	$\sqrt{E_\Sigma}$	Square Root of the Cumulative Seismic Energy
27	M_{mean}	Mean magnitude of the last n events

In the realm of earthquake forecasting, XGBoost has been used alongside other artificial intelligence models to evaluate earthquake spatial probability, particularly in regions like the Arabian Peninsula [84]. The integration of XGBoost with explainable artificial intelligence (XAI) models shows promising results, emphasizing the importance of including additional factors such as seismic gaps and tectonic contacts to enhance forecast accuracy [84]. This fusion of advanced machine learning techniques with traditional seismic analysis methods demonstrates a multidimensional approach to earthquake forecasting, highlighting the significance of comprehensive data analysis in seismology.

XGBoost has been integrated into earthquake monitoring and early warning systems to provide real-time alerts before significant ground shaking occurs [85]. By employing broadband P waveform data and XGBoost algorithms, researchers have developed systems capable of issuing earthquake warnings several seconds prior to the onset of a seismic event [85]. This proactive approach to earthquake forecasting underscores the potential of machine learning algorithms like XGBoost in improving seismic monitoring and disaster mitigation efforts.

XGBoost has been incorporated into comprehensive earthquake forecasting models, combining neural networks and other machine learning classifiers to analyze seismic data and forecast earthquake impacts [86].

2) RANDOM FOREST

We utilized a Random Forest model with 100 estimators to forecast earthquake categories, and the resulting accuracy is presented in Table 3. Random Forest is a machine learning algorithm that has been successfully applied in seismology for earthquake forecasting and analysis. Researchers have utilized Random Forest to develop models that can forecast earthquakes by analyzing seismic data and identifying patterns that precede seismic activity [87]. This approach represents a significant advancement in earthquake forecasting methodologies, emphasizing data-driven techniques to enhance the reliability of seismic forecasts.

In seismology, Random Forest has been used to distinguish seismic waveforms, allowing researchers to differentiate between earthquake signals and background noise effectively [88]. By training Random Forest classifiers with a substantial dataset of earthquake and noise waveforms, researchers have created models capable of automatically extracting features and classifying seismic events with high accuracy [88]. This application of Random Forest underscores its potential in improving earthquake early warning systems by facilitating rapid and precise identification of seismic events.

Random Forest has been employed in detecting and classifying seismic signals related to various geological phenomena, such as landslides and glacial earthquakes [87], [89]. By utilizing Random Forest classifiers, researchers have automated the process of identifying and categorizing seismic events, leading to more efficient monitoring and analysis of geological activities [87], [89]. This automated approach not only enhances event recognition speed but also improves the overall comprehension of seismic processes in geologically active regions.

In the field of earthquake forecasting, Random Forest shows promise in forecasting the magnitude and occurrence of seismic events [89]. Studies indicate that Random Forest models can effectively forecast earthquake magnitudes in specific regions, providing valuable insights for disaster preparedness and risk mitigation strategies [89]. This predictive capability highlights the potential of Random Forest in supporting decision-making processes related to earthquake response and mitigation efforts.

3) LIGHTGBM

We applied LightGBM using the following parameters: force_col_wise: True, min_split_gain: 0.5, min_child_samples: 20, num_leaves: 64, and max_depth: 6 to forecast earthquake categories. The resulting accuracy is shown in Table 3.

LightGBM, a tree-based boosting algorithm, has been utilized in earthquake forecasting within seismology. Researchers successfully used LightGBM to develop models capable of forecasting earthquake magnitudes and mapping seismic vulnerability by leveraging artificial intelligence techniques [90]. By utilizing historical strong motion data from databases like NGA-west2, LightGBM models demonstrated the ability to swiftly and accurately replicate the distribution of strong motion near earthquake epicenters [90]. This application of LightGBM represents a significant advancement in earthquake forecasting methodologies, showcasing the algorithm's efficiency in handling seismic data and enhancing predictive capabilities in seismology.

LightGBM is acknowledged for its efficiency in data processing and memory consumption reduction, making it a valuable tool for analyzing seismic data in large sample applications [91]. The algorithm's capacity to enhance processing speed while maintaining accuracy is particularly advantageous in seismology, where timely analysis of seismic events is critical for effective earthquake forecasting and risk assessment. By leveraging LightGBM's capabilities, researchers can streamline data processing tasks and improve the efficiency of earthquake forecasting models.

In the realm of earthquake forecasting, LightGBM plays a crucial role in forecasting seismic events and evaluating seismic vulnerability in earthquake-prone regions. Through the integration of LightGBM into predictive models, researchers have been able to analyze seismological parameters and forecast the areas impacted by earthquake-induced landslides using sophisticated data processing techniques [92]. This approach underscores the algorithm's versatility in handling complex seismic datasets and providing valuable insights into earthquake impacts, aiding in disaster preparedness and risk mitigation efforts.

The incorporation of LightGBM into earthquake forecasting models enables researchers to enhance the accuracy of seismic forecasts and deepen the understanding of seismic processes. By integrating LightGBM into comprehensive earthquake forecasting frameworks, seismologists can leverage the algorithm's capabilities to analyze seismic data, identify seismic patterns, and forecast earthquake magnitudes with greater precision [2].

4) GRADIENT BOOSTING MACHINES (GBM)

We applied Gradient Boosting Machines (GBM) to forecast earthquake categories, and the resulting accuracy is shown in Table 3. Gradient Boosting Machines (GBM) are an ensemble learning technique developed by Jerome Friedman. GBM is composed of weak learners, typically regression trees, that are boosted by adding weak learners using a functional gradient descent to minimize the loss function of the entire ensemble [93]. In seismology, GBM is utilized to enhance earthquake forecasting models by optimizing the loss function and improving the accuracy of seismic forecasts.

5) MULTILAYER PERCEPTRON (MLP)

We employed a Multilayer Perceptron (MLP) with the following parameters to forecast earthquake categories, and the resulting accuracy is detailed in Table 3: max_iter: 1000, learning_rate_init: 0.001, hidden_layer_sizes: (100, 100). In seismology, the Multilayer Perceptron (MLP) neural network model has been utilized to forecast earthquake magnitudes and assess seismic events accurately. Researchers have employed MLP to develop models capable of forecasting the magnitude of earthquakes, providing valuable insights into seismic activity [94]. By leveraging the capabilities of MLP, seismologists can analyze seismic data and forecast earthquake magnitudes with enhanced precision, contributing to more effective disaster preparedness and risk mitigation strategies in earthquake-prone regions.

The application of MLP in seismology enables researchers to forecast the magnitude of earthquakes using neural network models with multiple hidden layers [94]. By training MLP models with seismic data, researchers can extract patterns and features that aid in forecasting earthquake magnitudes, thereby improving the accuracy of seismic event forecasts. This approach highlights the effectiveness of MLP in handling complex seismic datasets and enhancing the understanding of seismic processes in seismology.

MLP is utilized to forecast the occurrence of seismic events and assess earthquake magnitudes based on historical seismic data [95]. By employing MLP neural networks with backpropagation learning algorithms, researchers can analyze seismic patterns and forecast the magnitude of earthquakes accurately. This utilization of MLP in earthquake forecasting models demonstrates the algorithm's effectiveness in handling seismic data and enhancing the reliability of seismic forecasts in seismology.

MLP is applied in seismology to create earthquake forecasting models that utilize artificial neural networks to forecast seismic events [96].

6) DECISION TREES

We employed Decision Trees to forecast earthquake categories, and the resulting accuracy is detailed in Table 3.

Decision Trees are a widely used machine learning algorithm in seismology for earthquake forecasting and analysis. They are structured as tree-like models where each internal node represents a feature or attribute, each branch signifies a decision rule, and each leaf node indicates the outcome or prediction [97]. In seismology, Decision Trees have been effectively utilized to analyze seismic data, forecast earthquake magnitudes, evaluate seismic vulnerability, and categorize seismic events based on various parameters.

Researchers have applied Decision Trees in seismology to forecast earthquake magnitudes and assess seismic vulnerability by creating models and forecasting seismic events using a tree structure [97]. By developing Decision Trees based on seismic data, researchers can identify patterns and relationships that assist in forecasting earthquake magnitudes and comprehending seismic processes. This methodology showcases the efficacy of Decision Trees in managing intricate seismic datasets and enhancing the precision of earthquake forecasts in seismology.

Decision Trees have been utilized to assess parameters influencing earthquake damage and simulate earthquake damage distributions in seismically active regions [98]. Through the application of Decision Tree techniques, researchers holistically evaluate earthquake damages, considering both structural and non-structural factors to accurately forecast and model earthquake damage distributions. This use of Decision Trees underscores their adaptability in analyzing seismic data and forecasting the impact of seismic events on structures and infrastructure.

In the realm of earthquake forecasting, Decision Trees have been employed to classify seismic events, differentiate between various types of seismic signals, and forecast the likelihood of earthquakes based on historical seismic data [99]. By leveraging Decision Trees, researchers establish models that aid in decision-making during seismic events, enhance earthquake emergency response strategies, and refine earthquake forecasting methodologies. This utilization of Decision Trees demonstrates their effectiveness in analyzing seismic data and supporting decision-making processes in seismology.

Decision Trees have been integrated with other machine learning algorithms to forecast earthquake occurrences, evaluate the seismic performance of structures, and enhance disaster planning and response strategies [100].

7) SUPPORT VECTOR MACHINES (SVM)

We used Support Vector Machines (SVM) to forecast earthquake categories, and the accuracy results are shown in Table 3.

Support Vector Machines (SVM) have been utilized in seismology for earthquake forecasting and analysis. Researchers have employed SVM as a machine learning tool to enhance earthquake forecasting models and improve the accuracy of seismic event forecasts [101]. By leveraging the capabilities of SVM, seismologists analyze seismic data, classify seismic events, and forecast earthquake occurrences with greater precision, contributing to more effective disaster management strategies and risk mitigation efforts in earthquake-prone regions.

In the context of seismology, SVM is used to classify seismic signals, differentiate between various types of seismic events, and forecast the likelihood of earthquakes based on historical seismic data [101]. By utilizing SVM algorithms, researchers develop models that aid in decision-making during seismic events, enhance earthquake emergency response strategies, and refine earthquake forecasting methodologies. This application of SVM highlights its effectiveness in analyzing seismic data and supporting decision-making processes in seismology.

SVM has been combined with other machine learning algorithms to forecast earthquake occurrences, assess seismic vulnerability, and improve disaster planning and response strategies [102].

8) K-NEAREST NEIGHBORS (K-NN)

We applied k-Nearest Neighbors (k-NN) to forecast earthquake categories, and the accuracy results are shown in Table 3. In seismology, the k-Nearest Neighbors (k-NN) algorithm has been utilized as a valuable tool for earthquake forecasting and analysis. The k-NN algorithm is a popular non-parametric method used for classification and regression, making it suitable for handling seismic data and forecasting seismic events [103]. By leveraging the k-NN algorithm, seismologists analyze seismic patterns, classify seismic events, and forecast earthquake occurrences with enhanced accuracy, contributing to more effective disaster management strategies and risk mitigation efforts in earthquake-prone regions.

Researchers have employed the k-NN algorithm in seismology to classify seismic signals, differentiate between various types of seismic events, and forecast the likelihood of earthquakes based on historical seismic data [104]. By applying the k-NN algorithm, researchers develop models that aid in decision-making during seismic events, improve earthquake emergency response strategies, and refine earthquake forecasting methodologies. This utilization of the k-NN algorithm demonstrates its effectiveness in analyzing seismic data and supporting decision-making processes in seismology [105].

9) AdaBoost

We utilized AdaBoost with the SAMME algorithm to forecast earthquake categories. The accuracy results are displayed in Table 3. In seismology, the AdaBoost machine learning algorithm has enhanced earthquake prediction models and improved the accuracy of seismic event forecasts. AdaBoost, which stands for Adaptive Boosting, is a boosting algorithm that combines multiple weak learners to create a strong predictive model. Researchers have employed AdaBoost in seismology to analyze seismic data, predict earthquake occurrences, and assess seismic vulnerability effectively [106].

A study introduced a novel earthquake prediction framework based on the classical AdaBoost machine learning algorithm, incorporating satellite remote sensing products like infrared and hyperspectral gases to detect earthquake perturbations [107]. By integrating AdaBoost within the framework of inverse boosting pruning trees (IBPT), the researchers achieved promising forecasting results in the retrospective validation of global earthquake cases, demonstrating the algorithm's effectiveness in earthquake prediction [107].

AdaBoost has been integrated into earthquake prediction models to evaluate seismic vulnerability and forecast seismic ground motions. The seismic vulnerability of Reinforced Concrete (RC) structures under single and multiple seismic events was forecasted using various machine learning algorithms, including the AdaBoost Regressor [108]. This incorporation of AdaBoost into seismic vulnerability assessment models underscores its usefulness in analyzing seismic data and improving the prediction of earthquake impacts on structures.

AdaBoost has been applied in earthquake prediction systems that merge earthquake indicators with genetic programming to enhance prediction accuracy. An earthquake prediction system that utilizes AdaBoost alongside earthquake forecasting indicators has led to improved results in earthquake forecasting [109].

10) CONVOLUTIONAL NEURAL NETWORKS (CNN)

We utilized Convolutional Neural Networks (CNN) to forecast earthquake categories. The accuracy results are displayed in Table 3. Our architecture is as follows: Input layer with shape (number of features, 1); Conv1D layer with 32 filters, kernel size of 3, and 'relu' activation; Flatten layer; Dense layer with 64 units and 'relu' activation; and Dense layer with Softmax activation for output. In seismology, Convolutional Neural Networks (CNN) have proven to be valuable tools for earthquake prediction and analysis. CNNs, a type of deep neural network that incorporates convolution calculations and has a deep structure, are well-suited for handling seismic data and forecasting seismic events [110]. Researchers have successfully utilized CNNs in seismology to analyze seismic patterns, classify seismic events, and forecast earthquake occurrences with increased accuracy, contributing to more effective disaster management strategies and risk mitigation efforts in earthquake-prone regions.

A study demonstrated the development of a CNN model capable of detecting and classifying seismic body wave phases across various circumstances, highlighting the

effectiveness of CNNs in seismic phase detection [111]. Through the application of CNNs, researchers can automate the process of identifying seismic phases, leading to improved seismic event classification and analysis. This utilization of CNNs emphasizes their value in enhancing seismic data processing and interpretation in seismology.

CNNs have been integrated into earthquake prediction models to assess seismic vulnerability and forecast seismic ground motions. By incorporating CNNs into seismic vulnerability assessment frameworks, researchers can enhance the accuracy of seismic impact predictions on structures and infrastructure. This integration showcases the effectiveness of CNNs in analyzing seismic data and improving the prediction of earthquake impacts, thereby supporting disaster preparedness and risk mitigation strategies in seismology.

CNNs have been employed in the classification of seismic events based on waveform data, showcasing their ability to process complex seismic signals and accurately classify seismic events [112].

11) RECURRENT NEURAL NETWORKS (RNN)

We employed Recurrent Neural Networks (RNN) to forecast earthquake categories, and the accuracy results are shown in Table 3. Our architecture includes the following layers: an input layer with shape (number of features, 1); an LSTM layer with 64 units and return sequences set to True; a flatten layer; a dense layer with 64 units and 'relu' activation; and a dense layer with Softmax activation for output. Recurrent Neural Networks (RNN) have become a valuable tool in seismology for earthquake forecasting and analysis. RNNs, a type of neural network incorporating feedback loops, are well-suited for handling seismic time-series data due to their ability to capture temporal dependencies in sequential data [113]. Researchers have successfully applied RNNs in seismology to model postseismic deformation, classify seismic events, and forecast earthquake occurrences with improved accuracy, contributing to more effective disaster management strategies and risk mitigation efforts in earthquake-prone regions.

A study introduced a machine-learning approach using RNNs to characterize the postseismic deformation of the 2011 Tohoku-Oki Earthquake based on time-series data, demonstrating the effectiveness of RNNs in accurately modeling observed seismic phenomena [113]. By leveraging the capabilities of RNNs, researchers can analyze seismic data over time and forecast the evolution of seismic events, providing valuable insights into the dynamics of seismic processes in seismology.

RNNs have been utilized in earthquake detection systems to analyze seismic array data and detect seismic events efficiently. A study focused on developing a graph-partitioning based CNN for earthquake detection using a seismic array showcased the effectiveness of RNNs in processing large-scale seismic network data sets and improving earthquake detection techniques [114]. This application of RNNs highlights their ability to handle complex spatiotemporal data and enhance earthquake detection capabilities in seismology.

RNNs have been integrated into seismic event classification models to analyze seismic waveforms and classify seismic events accurately. By leveraging RNNs for seismic event classification, researchers can extract features from seismic signals and categorize seismic events based on their characteristics, leading to more precise earthquake forecasts and assessments in seismology [115].

12) LONG SHORT-TERM MEMORY NETWORKS (LSTM)

We used Long Short-Term Memory Networks (LSTM) to forecast earthquake categories, and the accuracy results are presented in Table 3. Our architecture includes the following layers: an input layer with a shape of (number of features, 1); an LSTM layer with 64 units; a dense layer with 64 units and 'relu' activation; and a dense layer with Softmax activation for output.

Long Short-Term Memory Networks (LSTM) are a type of neural network architecture that is particularly well-suited for sequential data analysis due to their ability to retain information over long periods. In the context of seismology, LSTM networks have been increasingly utilized for earthquake forecasting. These networks excel in capturing the temporal dependencies present in seismic data, making them valuable tools for forecasting seismic events. Studies such as those by Hsu et al. [116], Cao et al. [117], and Abri and Artuner [118] have demonstrated the effectiveness of LSTM networks in forecasting various seismic parameters like peak ground acceleration (PGA) and earthquake occurrences.

In the realm of seismology, the forecasting of earthquakes has long been a challenging and critical endeavor due to its implications for public safety and disaster mitigation. Researchers have explored various approaches to improve earthquake forecasting accuracy, with a focus on leveraging advanced technologies like deep learning and neural networks. The study by Dias and Papa [96] highlights the application of neural networks, specifically multilayer perceptron models, for probabilistic earthquake forecasting, showcasing the potential of machine learning techniques in seismic event forecasting.

The integration of attention mechanisms with LSTM networks, as demonstrated in the work by Banna et al. [27], has shown promising results in enhancing earthquake forecasting accuracy. By incorporating attention mechanisms, which allow the model to focus on relevant parts of the input sequence, the LSTM network can better capture subtle patterns in seismic data, leading to improved forecasting capabilities.

13) GATED RECURRENT UNITS (GRU)

We implemented Gated Recurrent Units (GRU) to forecast earthquake categories, and the accuracy results are detailed in Table 3. Our architecture comprises: an input layer with a shape of (number of features, 1); a GRU layer with 64 units; a dense layer with 64 units and 'relu'

activation; and a dense layer with Softmax activation for output. Gated Recurrent Units (GRU) are a type of neural network architecture designed to efficiently model sequential data, similar to LSTM networks. In seismology, GRUs have become valuable for earthquake forecasting due to their streamlined architecture with fewer parameters, making them computationally efficient for certain applications in earthquake forecasting. Studies by Dias and Papa [96] and Wang et al. [119] have explored the use of neural networks, including GRUs, in earthquake forecasting, demonstrating the potential of these models in capturing complex temporal patterns in seismic data.

Seismologists increasingly utilize advanced machine learning techniques, such as GRUs, to enhance the accuracy and reliability of earthquake forecasting models. By employing GRU networks, researchers can analyze seismic data sequences effectively and extract meaningful patterns for more precise seismic event forecasts. Akter [120] utilized an Evidential Reasoning Approach to forecast earthquakes based on specific signs and patterns, showcasing the versatility of neural network models like GRUs in seismic hazard assessment.

Integrating GRUs with additional data sources, such as GPS data and outgoing longwave radiation, shows promise in improving earthquake forecasting accuracy. While Gitis et al. [42] stress the importance of using artificial neural networks for earthquake forecasting, studies like that of Zhai et al. [43] demonstrate the effectiveness of combining GRU models with time series forecasting techniques to detect thermal anomalies in earthquake processes, highlighting the interdisciplinary approach required in modern seismology research.

In earthquake forecasting, evaluating seismic parameters and their spatial variations is crucial for developing robust forecasting models. Research by Hussain et al. [41] on the spatial variation of b-values and their relationship with fault blocks suggests the potential of using such parameters alongside GRU networks to forecast high-magnitude earthquakes. Additionally, studies like that of Marc et al. [92] focus on forecasting the area affected by earthquake-induced landsliding based on seismological parameters, illustrating the practical applications of integrating GRU models with geophysical data for hazard assessment.

Analyzing earthquake catalogs and historical seismicity patterns provides valuable insights for refining earthquake forecasting models. Investigations such as those by Chouliaras [121] on the earthquake catalog of the National Observatory of Athens and Alabi et al. [122] on seismicity patterns in Southern Africa emphasize the importance of leveraging historical seismic data to enhance the performance of GRU-based forecasting models.

14) TRANSFORMER MODELS

We utilized Transformer Models to forecast earthquake categories, and the accuracy results are presented in Table 3. Our architecture includes: an input layer with a shape of (number of features, 1); a dense layer with 64 units and 'relu' activation; and a dense layer with Softmax activation for output. Transformer models have become a valuable tool in various seismological applications, including earthquake forecasting and seismic event analysis. In the context of seismology, transformer models are utilized for tasks such as earthquake detection, phase picking, earthquake source characterization, and early warning systems [123]. These models have demonstrated their effectiveness in processing large volumes of seismic data efficiently and capturing complex temporal patterns present in seismic signals.

Seismologists increasingly turn to machine learning techniques, including transformer models, to enhance earthquake forecasting accuracy and improve seismic event forecasting. The ability of transformer models to handle sequential data and learn dependencies across different time steps makes them well-suited for analyzing seismic signals and extracting meaningful features for earthquake forecasting. The application of transformer models in seismology has shown promising results in enhancing the understanding of seismic events and improving the reliability of earthquake forecasts.

Transformer models have been instrumental in separating earthquake signals from ambient noise in seismograms, contributing to more accurate earthquake detection and analysis [123].

15) LOGISTIC REGRESSION

We applied Logistic Regression to forecast earthquake categories, and the accuracy results are displayed in Table 3.

Logistic regression is a statistical method commonly used in various fields, including seismology, to analyze the relationship between a binary outcome and one or more predictor variables. In the context of seismology, logistic regression has been applied to forecast and assess different aspects related to earthquakes. For instance, Jessee et al. [124] developed a global empirical model for assessing seismically induced landslides using logistic regression to understand the distribution of earthquake-triggered landslides based on factors like ground shaking, topographic slope, and land cover type. This study highlights the utility of logistic regression in modeling the impact of earthquakes on the occurrence of landslides.

Logistic regression has been utilized in earthquake forecasting studies, although it is noted that traditional models based on physical principles and statistical seismology laws have limitations in forecasting large earthquakes [125]. While logistic regression has been used in earthquake forecasting models, it is essential to acknowledge the challenges in accurately forecasting significant seismic events solely based on empirical laws and physical principles.

In the specific context of seismically induced damage patterns, Rawat et al. [126] employed logistic regression to investigate seismic hazard assessment by considering

site-specific parameters such as lithology, proximity to fault lines, soil texture, and groundwater. This application demonstrates how logistic regression can be used to understand the factors influencing seismic damage patterns and assess earthquake risks in different geological settings.

Logistic regression has been applied in studies focusing on earthquake-induced landslides. Vilder et al. [127] used a logistic regression model to correlate earthquake-induced landslide inventories with various topographic, geological, and seismological parameters to determine the factors contributing to coseismic landslides.

16) NAIVE BAYES

We used Naive Bayes to forecast earthquake categories, and the accuracy results are shown in Table 3. One study by Fahandezhsadi and Sadi [73] focused on earthquake magnitude forecasting using probabilistic classifiers, including Naive Bayes. The research aimed to enhance the accuracy of Naive Bayes by relaxing its strong conditional independence assumption, indicating an interest in exploring the potential of Naive Bayes in seismic event forecasting. This study suggests that Naive Bayes, when adapted and optimized for seismic data, could potentially contribute to earthquake forecasting efforts.

In a broader context of seismic event discrimination, a study by Elkhouly [128] employed multiple machine learning techniques, including Naive Bayes, to distinguish between nuclear explosions and natural earthquakes. While the primary focus was on seismic discrimination, the inclusion of Naive Bayes in the machine learning models underscores its versatility and potential applicability in seismic data analysis. This research highlights the adaptability of Naive Bayes in complex seismic event classification tasks.

A study by Murwantara et al. [129] comparing machine learning algorithms for earthquake forecasting in Indonesia evaluated Naive Bayes alongside other methods like multinomial logistic regression and support vector machine. The research aimed to assess the performance of these algorithms in medium-to-long-term earthquake forecasting using historical data, indicating the consideration of Naive Bayes as a potential tool for seismic forecasting. This study suggests that Naive Bayes can be part of a comprehensive approach to earthquake forecasting when combined with other predictive models.

B. ACCURACY COMPARISON

After applying machine learning and neural networks, we calculated the accuracies of each model. These results, along with their statistical significance, are presented in Table 3. According to this table, eight models demonstrate statistically significant accuracies: XGBoost, Random Forest, LightGBM, Gradient Boosting Machines, Multilayer Perceptron, Decision Trees, Support Vector Machines, and k-Nearest Neighbors.

TABLE 3. Statistical significance of different models.

#	Model	Acc.	Sign.
1	XGBoost	0.5597	Yes
2	Random Forest	0.5350	Yes
3	LightGBM	0.5350	Yes
4	Gradient Boosting Machines	0.5185	Yes
5	Multilayer Perceptron	0.4650	Yes
6	Decision Trees	0.4280	Yes
7	Support Vector Machines	0.4115	Yes
8	k-Nearest Neighbors	0.3951	Yes
9	AdaBoost	0.3704	No
10	Convolutional Neural Networks	0.3539	No
11	Recurrent Neural Networks	0.3498	No
12	Gated Recurrent Units	0.3498	No
13	Long Short-Term Mem. Networks	0.3457	No
14	Transformer Models	0.3333	No
15	Logistic Regression	0.2798	No
16	Naive Bayes	0.2593	No

Baseline accuracy in machine learning models refers to the minimum level of accuracy that a model should achieve to be considered better than random guessing or a simplistic approach. It serves as a reference point for evaluating the performance of a model [130].

To assess the statistical significance of accuracy for machine learning (ML) models using confidence interval calculations, researchers can rely on established methodologies and techniques. Confidence intervals are crucial in quantifying the uncertainty associated with estimated parameter values derived from a sample [131].

1) BASELINE ACCURACY

The baseline accuracy is the accuracy achieved by always predicting the most frequent class. In our dataset, the baseline accuracy is:

$$\text{Baseline Accuracy} = 31.60\% \approx 0.3160 \tag{72}$$

2) Z-SCORE CALCULATION

The Z-score for a 95% confidence interval is the value that leaves 2.5% in each tail of the standard normal distribution. This is found using the cumulative distribution function (CDF):

$$Z = \text{CDF}^{-1}(0.975) \approx 1.96 \tag{73}$$

3) CONFIDENCE INTERVAL CALCULATION

For a 95% confidence interval, we use the Z-score corresponding to the desired confidence level (for 95%, $Z \approx 1.96$). The confidence interval (CI) is calculated as:

$$CI = p \pm Z \times SE \tag{74}$$

4) CRITICAL ACCURACY DETERMINATION

The critical accuracy, which is the upper bound of the confidence interval, is given by:

$$\text{Critical Accuracy} = p + Z \times SE \tag{75}$$

5) RESULTS

Given our dataset:

$$p = 0.3160 \tag{76}$$

$$n = (0.2 \times \text{Data Size}) \tag{77}$$

The 95% confidence interval for the baseline accuracy is:

$$CI \approx (0.2576, 0.3745) \tag{78}$$

Any model accuracy significantly above 37.45% would be considered statistically significant compared to the baseline accuracy at a 95% confidence level. The results are shown in Fig 8.

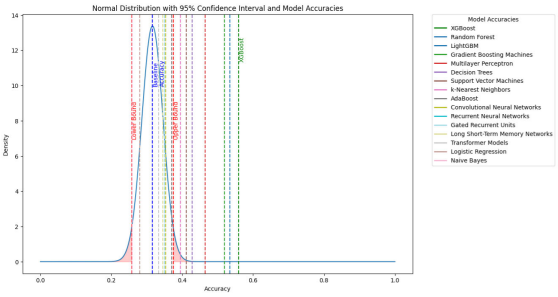


FIGURE 8. Normal distribution with 95% confidence interval for baseline accuracy.

VI. HIGHEST ACCURACY: XGBoost

Since the XGBoost ML Algorithm achieved the highest accuracy and captured complex patterns in the earthquake data better than other models, leading to higher forecasting accuracy, we will continue our analysis with XGBoost. XGBoost, which stands for Extreme Gradient Boosting, is a machine learning algorithm known for its efficiency and effectiveness across various domains. The algorithm optimizes by boosting weak learners, typically decision trees, into a strong learner through gradient descent optimization. This process involves iteratively adding new models to correct errors made by existing models [132].

Mathematically, XGBoost can be represented as an optimization problem aiming to minimize the sum of the loss function and a regularization term. This formulation enables XGBoost to handle complex datasets efficiently while preventing overfitting by penalizing overly complex models [133]. XGBoost incorporates features from the random forest algorithm, continuously reduces residuals to decrease overfitting, uses standardized regularization terms to mitigate overfitting, and allows for parallel calculations to enhance forecast efficiency and precision [134].

Studies comparing XGBoost with other machine learning algorithms have shown its superior forecasting performance, especially in risk forecasting models, where it outperformed random forest, support vector machine, and k-nearest neighbor algorithms [135]. XGBoost has been successfully applied in finance, accounting, business, and audit, demonstrating its versatility and robustness [136].

XGBoost's ability to handle imbalanced datasets effectively makes it suitable for tasks like intrusion detection, where data distribution is skewed [137]. Its performance in handling complex datasets and providing interpretable insights has proven valuable in healthcare settings for forecasting complications and unplanned readmissions [138].

In terms of optimization, XGBoost utilizes gradient boosting to enhance its performance by iteratively minimizing the loss function using gradient descent. This iterative process allows XGBoost to construct a strong ensemble model from weak learners, leading to highly accurate forecasts [132].

XGBoost's efficiency with large-scale datasets is attributed to its implementation of weighted quantile sketch for proposal calculation, a sparsity-aware algorithm for parallel tree learning, and a cache-aware block structure for out-of-core tree learning. These features contribute to the algorithm's scalability and performance, enabling it to process massive amounts of data efficiently [139].

A. MATHEMATICAL BACKGROUND OF XGBoost

XGBoost, short for eXtreme Gradient Boosting, is an efficient and scalable implementation of gradient boosting that leverages decision trees as base learners. The core idea of gradient boosting is to build an ensemble of weak learners, typically decision trees, in a sequential manner, where each new tree attempts to correct the errors made by the previous ones. The mathematical foundation of XGBoost involves the following key components:

1) OBJECTIVE FUNCTION

The objective function in XGBoost consists of two parts: the loss function and the regularization term. The loss function measures how well the model fits the training data, while the regularization term penalizes model complexity to prevent overfitting. The objective function $\mathcal{L}$ for XGBoost can be expressed as:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{79}$$

where l is the loss function (e.g., mean squared error for regression, log loss for classification), $\hat{y}_i$ is the predicted value for the i-th instance, Ω is the regularization term, f_k represents the k-th decision tree, and K is the number of trees.

2) ADDITIVE TRAINING

XGBoost builds trees additively, meaning that new trees are added to the model to correct the residuals (errors) of the existing trees. At each step t, the prediction $\hat{y}_i^{(t)}$ is updated by adding the prediction from the new tree $f_t(x_i)$:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{80}$$

The new tree f_t is trained to minimize the objective function.

3) GRADIENT DESCENT AND TREE CONSTRUCTION

The trees in XGBoost are constructed using gradient descent to optimize the objective function. For a given iteration t, the objective function can be approximated using a second-order Taylor expansion:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{81}$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ are the first and second-order gradients of the loss function with respect to the prediction. The regularization term for tree t can be defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \tag{82}$$

where T is the number of leaves in the tree, w_j is the weight of leaf j, and γ and λ are regularization parameters.

4) SPLIT FINDING AND LEAF WEIGHT CALCULATION

To construct the tree, XGBoost evaluates all possible splits for each feature and selects the split that maximizes the gain in the objective function. The gain from a split j is calculated as:

$$\text{Gain}(j) = \frac{1}{2} \left[\frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \frac{G_L^2}{H_L + \lambda} - \frac{G_R^2}{H_R + \lambda} \right] - \gamma \tag{83}$$

where G_L and G_R are the sums of the first-order gradients and H_L and H_R are the sums of the second-order gradients for the left and right child nodes, respectively.

Once the best split is found, the weights for the leaves are updated to minimize the objective function:

$$w_j = -\frac{G_j}{H_j + \lambda} \tag{84}$$

where G_j and H_j are the sums of the first and second-order gradients for leaf j.

5) REGULARIZATION AND PRUNING

XGBoost includes regularization terms to prevent overfitting. The parameters γ and λ control the complexity of the model by penalizing the number of leaves and the weights of the leaves, respectively. XGBoost also performs tree pruning by removing splits that do not improve the objective function sufficiently, using a minimum loss reduction parameter.

B. GETTING THE HIGHEST ACCURACY SUBSET WITH THE INFORMATION GAIN METHOD FOR XGBoost

As seen in Table 4, we calculated the Information Gain (IG) values for each feature to assess their importance in forecasting the earthquake class.

The Information Gain method is a fundamental technique used in machine learning algorithms to improve the efficiency and accuracy of models by selecting the most relevant features for training [140]. This method works by assessing the information provided by each feature to the model, thereby enhancing the model's performance [141]. Metrics such as the Gini index and information gain are calculated from the feature vector to construct decision trees more effectively, resulting in improved classification and forecasting outcomes [141]. Various studies have applied the Information Gain method to optimize performance and operational efficiency by reducing the number of features in machine learning models [142].

The Information Gain method has been combined with other techniques to enhance its effectiveness in different applications. For example, the fusion of Information Gain and Recursive Feature Elimination (IG-RFE) with Support Vector Machines (SVM) has been suggested to enhance the stability and performance of feature selection methods, particularly in gene expression data analysis. This integration showcases the versatility of the Information Gain method across different domains and its contribution to optimizing machine learning processes [143].

The Information Gain method has been utilized in studies related to cybersecurity and malware detection. In the assessment of features for identifying cyber-phishing attacks, methods like ANOVA, X$\hat{2}$, and Information Gain have been employed to evaluate the relevance of different features in detecting phishing attempts [144]. Similarly, in the realm of Android malware detection, feature selection based on the Genetic Algorithm combined with Information Gain has been used to boost the accuracy and efficiency of machine learning models in recognizing malicious software [142].

TABLE 4. Feature importance scores.

No.	Feature	Importance
1	ΔM	0.066928
2	$M_{\text{max}}^{\text{last week}}$	0.062145
3	$\sqrt{E_\Sigma}$	0.050640
4	T	0.050327
5	b value	0.047221
6	C	0.046383
7	η	0.045447
8	longitude	0.044914
9	Dist. to San Andreas Fault	0.041380
10	time since last earthquake	0.041129
11	μ	0.039281
12	depth	0.037938
13	$\Delta b_{i-4,i-6}$	0.037527
14	$\Delta b_{i-6,i-8}$	0.033738
15	latitude	0.033622
16	$\Delta b_{i-8,i-10}$	0.032063
17	Dist. to Whittier Fault	0.031791
18	Dist. to Newport Inglewood Fault	0.029016
19	$\Delta b_{i-2,i-4}$	0.028531
20	Dist. to Sierra Madre Fault	0.028284
21	M_{mean}	0.027988
22	$\Delta b_{i,i-2}$	0.027489
23	Dist. to Puente Hills Thrust Fault	0.025059
24	Dist. to Raymond Fault	0.024859
25	mag	0.023619
26	a value	0.023321
27	$P(M \geq 5)$	0.019360

The Information Gain method has been valuable in addressing challenges such as overfitting and underfitting in machine learning algorithms. By leveraging information-theoretic insights, researchers have gained a new perspective on algorithm capacity and performance issues, leading to the development of more robust and generalizable models [145].

In machine learning, Information Gain (IG) is used to measure the reduction in uncertainty or entropy when a dataset is split based on a feature. It is a critical metric in decision trees and ensemble methods like XGBoost, helping to select the most informative features for building the model.

Entropy is a measure of the impurity or randomness in the data. For a dataset D with classes $C_1, C_2, \ldots, C_k$, the entropy $H(D)$ is defined as:

$$H(D) = -\sum_{i=1}^{k} p(C_i) \log_2 p(C_i) \tag{85}$$

where $p(C_i)$ is the proportion of examples in D that belong to class C_i.

Information Gain measures the reduction in entropy after the dataset is split based on a feature. For a feature A with possible values $\{a_1, a_2, \ldots, a_v\}$, the Information Gain $IG(D, A)$ is calculated as:

$$IG(D, A) = H(D) - \sum_{j=1}^{v} \frac{|D_j|}{|D|} H(D_j) \tag{86}$$

where: - $H(D)$ is the entropy of the original dataset D. - D_j is the subset of D where feature A has value a_j. - $|D_j|$ is the number of examples in D_j. - $|D|$ is the total number of examples in D. - $H(D_j)$ is the entropy of subset D_j.

We then built subsets of features based on their IG values, starting with the most important feature and incrementally adding one feature at a time. For each subset, we trained an XGBoost model and evaluated its accuracy.

The accuracy of the model was calculated for each subset of features, and the results are documented in Table 5, showing the feature set and the corresponding accuracy:

The highest accuracy of 0.650206 was achieved with the 7 Variable Subset, as seen in Table 6.

TABLE 5. Accuracy results for incremental feature subsets with the information gain method for XGBoost.

Features	Accuracy
1 Variable Subset	0.316872
2 Variable Subset	0.522634
3 Variable Subset	0.580247
4 Variable Subset	0.637860
5 Variable Subset	0.637860
6 Variable Subset	0.641975
7 Variable Subset	0.650206
8 Variable Subset	0.604938
9 Variable Subset	0.596708
10 Variable Subset	0.592593
11 Variable Subset	0.576132
12 Variable Subset	0.584362
13 Variable Subset	0.563786
14 Variable Subset	0.580247
15 Variable Subset	0.588477
16 Variable Subset	0.580247
17 Variable Subset	0.576132
18 Variable Subset	0.592593
19 Variable Subset	0.580247
20 Variable Subset	0.576132
21 Variable Subset	0.576132
22 Variable Subset	0.588477
23 Variable Subset	0.584362
24 Variable Subset	0.596708
25 Variable Subset	0.584362
26 Variable Subset	0.580247

TABLE 6. Selected features for earthquake forecasting.

No.	Var	Explanation
1	ΔM	Difference between the largest observed magnitude and the largest expected magnitude based on the Gutenberg-Richter law
2	$M_{\max}^{\text{last week}}$	Maximum magnitude recorded during the last week
3	$\sqrt{E_\Sigma}$	Square Root of the Cumulative Seismic Energy
4	T	Elapsed time between the last n events
5	b value	Gutenberg-Richter b-value for the magnitude
6	C	Coefficient of variation of the inter-event times
7	η	Sum of the mean square deviation from the regression line based on the Gutenberg-Richter law

C. GETTING THE HIGHEST ACCURACY SUBSET THROUGH ITERATION OF ALL COMBINATIONS OF SUBSETS

For this part of the study, high-performance computing (HPC) systems of Georgia Southern University were utilized. The highest accuracy subset, determined using the Information Gain method with XGBoost, was found to be 65% in the previous section. We wondered if there is a subset with a higher accuracy, which when trained with 80% of the data, can forecast with a higher accuracy the 20% test data. For this, we examined all the subsets of the 27 features.

Mathematically, the total number of possible subsets of n features is given by the power set, which is 2^n. Excluding the empty subset, the number of non-empty subsets is calculated as:

$$\sum_{r=1}^{n} \binom{n}{r} = 2^n - 1 \tag{87}$$

where $\binom{n}{r}$ is the binomial coefficient representing the number of ways to choose r features from n. With 27 features, the total number of non-empty subsets is:

$$2^{27} - 1 = 134,217,727 \tag{88}$$

Given the large number of subsets, the tasks were divided into 100 jobs on the Georgia Southern University's Talon servers to facilitate efficient computation. By iterating

through all these possible combinations of feature subsets, the subset yielding the highest accuracy was identified.

The subset of features yielding the highest accuracy, which is 69.14% for forecasting the maximum class of earthquake within a 30-day period, is shown in Table 7.

TABLE 7. Subset of features yielding the highest accuracy.

No.	Variable	Explanation
1	$\Delta b_{i-2,i-4}$	Incremental b-value between events $i-2$ and $i-4$
2	$\Delta b_{i-4,i-6}$	Incremental b-value between events $i-4$ and $i-6$
3	$\Delta b_{i-6,i-8}$	Incremental b-value between events $i-6$ and $i-8$
4	$M_{\text{max}}^{\text{last week}}$	Maximum magnitude recorded during the last week
5	a	Gutenberg-Richter a-value
6	μ	Mean time between events
7	C	Coefficient of variation of the inter-event times
8	$\sqrt{E_\Sigma}$	Square Root of the Cumulative Seismic Energy
9	M_{mean}	Mean magnitude of the last n events

VII. CONCLUSION

In our study, we have analyzed numerous references related to earthquake forecasting. Our research involved the development of a predictive pattern matrix, integrating machine learning algorithms and neural networks to forecast earthquakes. By feature engineering 27 diverse predictive features from historical earthquake records, we have achieved an accuracy of 69.14% with 9 features for the Los Angeles region. This result underscores the importance of combining advanced computational techniques with thorough data analysis, offering a promising direction for future research and application in earthquake forecasting. Our ability to correctly forecast the category of earthquakes across six different categories within a 30-day period is significant. One interesting aspect was that closeness to fault lines did not play an important role in forecasting the maximum magnitude class in the Los Angeles region. Achieving this level of accuracy is crucial for improving disaster preparedness and response strategies in Los Angeles, a region prone to seismic activity. Our approach provides a comprehensive and precise method for earthquake forecasting, contributing valuable insights to the field of seismology.

REFERENCES

[1] K. B. Olsen, R. J. Archuleta, and J. R. Matarese, "Three-dimensional simulation of a magnitude 7.75 earthquake on the San Andreas fault," *Science*, vol. 270, no. 5242, pp. 1628–1632, Dec. 1995, doi: 10.1126/science.270.5242.1628.

[2] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks," *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0199004, doi: 10.1371/journal.pone.0199004.

[3] L. Zhang, L. Si, H. Yang, Y. Hu, and J. Qiu, "Precursory pattern based feature extraction techniques for earthquake prediction," *IEEE Access*, vol. 7, pp. 30991–31001, 2019, doi: 10.1109/ACCESS.2019.2902224.

[4] V. Skrickij, E. Šabanovič, D. Shi, S. Ricci, L. Rizzetto, and G. Bureika, "Visual measurement system for wheel–rail lateral position evaluation," *Sensors*, vol. 21, no. 4, p. 1297, Feb. 2021, doi: 10.3390/s21041297.

[5] M. A. Bilal, Y. Ji, Y. Wang, M. P. Akhter, and M. Yaqub, "Early earthquake detection using batch normalization graph convolutional neural network (BNGCNN)," *Appl. Sci.*, vol. 12, no. 15, p. 7548, Jul. 2022, doi: 10.3390/app12157548.

[6] S. Kobayashi, K. Takemura, and F. Koga, "Apparent diffusion coefficient value as a biomarker for detecting muscle-invasive and high-grade bladder cancer: A systematic review," *Appl. Sci.*, vol. 12, no. 3, p. 1278, Jan. 2022, doi: 10.3390/app12031278.

[7] K. B. Olsen, "Site amplification in the Los Angeles basin from three-dimensional modeling of ground motion," *Bull. Seismological Soc. Amer.*, vol. 90, no. 6B, pp. S77–S94, Dec. 2000, doi: 10.1785/0120000506.

[8] A. Donnellan, L. Grant Ludwig, J. W. Parker, J. B. Rundle, J. Wang, M. Pierce, G. Blewitt, and S. Hensley, "Potential for a large earthquake near Los Angeles inferred from the 2014 La Habra earthquake," *Earth Space Sci.*, vol. 2, no. 9, pp. 378–385, Sep. 2015, doi: 10.1002/2015ea000113.

[9] E. Hauksson, "Earthquakes, faulting, and stress in the Los Angeles basin," *J. Geophys. Res., Solid Earth*, vol. 95, no. B10, pp. 15365–15394, Sep. 1990, doi: 10.1029/jb095ib10p15365.

[10] Z. Shen, D. D. Jackson, and B. X. Ge, "Crustal deformation across and beyond the Los Angeles basin from geodetic measurements," *J. Geophys. Res., Solid Earth*, vol. 101, no. B12, pp. 27957–27980, Dec. 1996, doi: 10.1029/96jb02544.

[11] J. P. Loveless and B. J. Meade, "Stress modulation on the San Andreas fault by interseismic fault system interactions," *Geology*, vol. 39, no. 11, pp. 1035–1038, Nov. 2011, doi: 10.1130/g32215.1.

[12] N. Romero, T. D. O'Rourke, L. K. Nozick, and C. A. Davis, "Seismic hazards and water supply performance," *J. Earthq. Eng.*, vol. 14, no. 7, pp. 1022–1043, Sep. 2010, doi: 10.1080/13632460903527989.

[13] D. Roten, K. B. Olsen, S. M. Day, Y. Cui, and D. Fäh, "Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity," *Geophys. Res. Lett.*, vol. 41, no. 8, pp. 2769–2777, Apr. 2014, doi: 10.1002/2014gl059411.

[14] J. H. Shaw and J. Suppe, "Earthquake hazards of active blind-thrust faults under the Central Los Angeles basin, California," *J. Geophys. Res., Solid Earth*, vol. 101, no. B4, pp. 8623–8642, Apr. 1996, doi: 10.1029/95jb03453.

[15] J. D. Zechar and T. H. Jordan, "Testing alarm-based earthquake predictions," *Geophys. J. Int.*, vol. 172, no. 2, pp. 715–724, Feb. 2008, doi: 10.1111/j.1365-246X.2007.03676.x.

[16] X. Huang, M. Luo, and H. Jin, "Application of improved ELM algorithm in the prediction of earthquake casualties," *PLoS ONE*, vol. 15, no. 6, Jun. 2020, Art. no. e0235236, doi: 10.1371/journal.pone.0235236.

[17] D. Schorlemmer, J. D. Zechar, M. J. Werner, E. H. Field, D. D. Jackson, and T. H. Jordan, "First results of the regional earthquake likelihood models experiment," *Pure Appl. Geophys.*, vol. 167, nos. 8–9, pp. 859–876, Aug. 2010, doi: 10.1007/s00024-010-0081-5.

[18] E. A. Al-Heety, H. F. Rafea, and O. J. Mohammad, "Evaluation of return period and occurrence probability of the maximum magnitude earthquakes in Iraq and surroundings," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 1300, no. 1, Feb. 2024, Art. no. 012001, doi: 10.1088/1755-1315/1300/1/012001.

[19] J. L. Rubinstein, W. L. Ellsworth, K. H. Chen, and N. Uchida, "Fixed recurrence and slip models better predict earthquake behavior than the time- and slip-predictable models: 1. Repeating earthquakes," *J. Geophys. Res., Solid Earth*, vol. 117, no. B2, Feb. 2012, Art. no. B02306, doi: 10.1029/2011jb008724.

[20] X. Yang, S. Du, and J. Ma, "Do earthquakes exhibit self-organized criticality?" *Phys. Rev. Lett.*, vol. 92, no. 22, Jun. 2004, Art. no. 228501, doi: 10.1103/physrevlett.92.228501.

[21] J. Huang, X. Wang, Y. Zhao, C. Xin, and H. Xiang, "Large earthquake magnitude prediction in Taiwan based on deep learning neural network," *Neural Netw. World*, vol. 28, no. 2, pp. 149–160, 2018, doi: 10.14311/nnw.2018.28.009.

[22] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM model with attention mechanism for earthquake prediction," *J. Supercomput.*, vol. 79, no. 17, pp. 19194–19226, Nov. 2023, doi: 10.1007/s11227-023-05369-y.

[23] R. J. Geller, D. D. Jackson, Y. Y. Kagan, and F. Mulargia, "Earthquakes cannot be predicted," *Science*, vol. 275, no. 5306, p. 1616, Mar. 1997, doi: 10.1126/science.275.5306.1616.

[24] D. A. J. Eberhard, J. D. Zechar, and S. Wiemer, "A prospective earthquake forecast experiment in the Western Pacific," *Geophys. J. Int.*, vol. 190, no. 3, pp. 1579–1592, Sep. 2012, doi: 10.1111/j.1365-246X.2012.05548.x.

[25] R. Tehseen, M. S. Farooq, and A. Abid, "Earthquake prediction using expert systems: A systematic mapping study," *Sustainability*, vol. 12, no. 6, p. 2420, Mar. 2020, doi: 10.3390/su12062420.

[26] Y. Ogata, "A prospect of earthquake prediction research," *Stat. Sci.*, vol. 28, no. 4, pp. 521–541, Nov. 2013, doi: 10.1214/13-sts439.

[27] Md. H. A. Banna, T. Ghosh, Md. J. A. Nahian, K. A. Taher, M. S. Kaiser, M. Mahmud, M. S. Hossain, and K. Andersson, "Attention-based bi-directional long-short term memory network for earthquake prediction," *IEEE Access*, vol. 9, pp. 56589–56603, 2021, doi: 10.1109/ACCESS.2021.3071400.

[28] Y. Y. Kagan, "Are earthquakes predictable?" *Geophys. J. Int.*, vol. 131, no. 3, pp. 505–525, Dec. 1997, doi: 10.1111/j.1365-246X.1997.tb06595.x.

[29] N. Ma, Y. Bai, and S. Meng, "Return period evaluation of the largest possible earthquake magnitudes in mainland China based on extreme value theory," *Sensors*, vol. 21, no. 10, p. 3519, May 2021, doi: 10.3390/s21103519.

[30] V. M. V. Herrera, E. A. Rossello, M. J. Orgeira, L. Arioni, W. Soon, G. Velasco, L. R. de la Cruz, E. Zúñiga, and C. Vera, "Long-term forecasting of strong earthquakes in North America, South America, Japan, Southern China and northern India with machine learning," *Frontiers Earth Sci.*, vol. 10, Jun. 2022, Art. no. 905792, doi: 10.3389/feart.2022.905792.

[31] A. J. Michael, "Testing prediction methods: Earthquake clustering versus the Poisson model," *Geophys. Res. Lett.*, vol. 24, no. 15, pp. 1891–1894, Aug. 1997, doi: 10.1029/97gl01928.

[32] Y. Kodera, J. Saitou, N. Hayashimoto, S. Adachi, M. Morimoto, Y. Nishimae, and M. Hoshiba, "Earthquake early warning for the 2016 Kumamoto earthquake: Performance evaluation of the current system and the next-generation methods of the Japan meteorological agency," *Earth, Planets Space*, vol. 68, no. 1, pp. 1–14, Dec. 2016, doi: 10.1186/s40623-016-0567-1.

[33] X. Yuan, H. Dan, Y. Qiuyin, Z. Wenjun, Y. Jing, and R. Min, *Analysis and Prediction of the SARIMA Model for a Time Interval of Earthquakes in the Longmenshan Fault Zone*. London, U.K.: IntechOpen, 2023, doi: 10.5772/intechopen.109174.

[34] P. Hajikhodaverdikhan, M. Nazari, M. Mohsenizadeh, S. Shamshirband, and K.-W. Chau, "Earthquake prediction with meteorological data by particle filter-based support vector regression," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 679–688, Jan. 2018, doi: 10.1080/19942060.2018.1512010.

[35] W. Astuti, W. Sediono, R. Akmeliawati, A. M. Aibinu, and M. J. E. Salami, "Investigation of the characteristics of geoelectric field signals prior to earthquakes using adaptive STFT techniques," *Natural Hazards Earth Syst. Sci.*, vol. 13, no. 6, pp. 1679–1686, Jun. 2013, doi: 10.5194/nhess-13-1679-2013.

[36] T. Nishikawa, "Comparison of statistical low-frequency earthquake activity models," *Earth, Planets Space*, vol. 76, no. 1, pp. 1–12, Apr. 2024, doi: 10.1186/s40623-024-02007-6.

[37] S. L. Nimmagadda and H. Dreher, "Ontology based data warehouse modeling and mining of earthquake data: Prediction analysis along eurasian-australian continental plates," in *Proc. 5th IEEE Int. Conf. Ind. Informat.*, Jul. 2007, pp. 597–602, doi: 10.1109/INDIN.2007.4384825.

[38] L. V. Narasimha Prasad, P. S. Murthy, and C. K. Kumar Reddy, "Analysis of magnitude for earthquake detection using primary waves and secondary waves," in *Proc. Int. Conf. Human Comput. Interact. (ICHCI)*, Aug. 2013, pp. 1–6, doi: 10.1109/ICHCI-IEEE.2013.6887820.

[39] F. Yang, M. Kefalas, M. Koch, A. V. Kononova, Y. Qiao, and T. Bäck, "Auto-REP: An automated regression pipeline approach for high-efficiency earthquake prediction using LANL data," in *Proc. 14th Int. Conf. Comput. Autom. Eng. (ICCAE)*, Mar. 2022, pp. 127–134, doi: 10.1109/ICCAE55086.2022.9762437.

[40] X. Zheng and Z. Tao, "Preliminary evaluation of crustal medium parameters in Western China," *E3S Web Conf.*, vol. 406, p. 01003, Oct. 2023, doi: 10.1051/e3sconf/202340601003.

[41] H. Hussain, Z. Shuangxi, M. Usman, and M. Abid, "Spatial variation of b-values and their relationship with the fault blocks in the western part of the Tibetan Plateau and its surrounding areas," *Entropy*, vol. 22, no. 9, p. 1016, Sep. 2020, doi: 10.3390/e22091016.

[42] V. Gitis, A. Derendyaev, and K. Petrov, "Analyzing the performance of GPS data for earthquake prediction," *Remote Sens.*, vol. 13, no. 9, p. 1842, May 2021, doi: 10.3390/rs13091842.

[43] D. Zhai, X. Zhang, and P. Xiong, "Detecting thermal anomalies of earthquake process within outgoing longwave radiation using time series forecasting models," *Ann. Geophys.*, vol. 63, no. 5, p. PA548, Nov. 2020, doi: 10.4401/ag-8057.

[44] H. Woith, G. M. Petersen, S. Hainzl, and T. Dahm, "Review: Can animals predict earthquakes?" *Bull. Seismological Soc. Amer.*, vol. 108, no. 3A, pp. 1031–1045, Jun. 2018, doi: 10.1785/0120170313.

[45] H. Kanamori and E. E. Brodsky, "The physics of earthquakes," *Phys. Today*, vol. 54, no. 6, pp. 34–40, Jun. 2001, doi: 10.1063/1.1387590.

[46] SCEDC. (2024). *Southern California Earthquake Data Center*. Accessed: Jun. 16, 2024. [Online]. Available: https://scedc.caltech.edu

[47] SCEDC. (2024). *Radius Search Tool*. Accessed: Jun. 16, 2024. [Online]. Available: https://service.scedc.caltech.edu/eq-catalogs/radius.php

[48] A. E.-A.-K. A. El-Aal, H. E. AbdelHafiez, H. Saadalla, and M. S. Soliman, "A homogenous moment magnitude and local magnitude scaling relation for earthquakes in Egypt," *NRIAG J. Astron. Geophys.*, vol. 9, no. 1, pp. 532–538, Jan. 2020, doi: 10.1080/20909977.2020.1794445.

[49] D. A. Nazaruddin and M. A. Md, "Southern Thailand's seismicity and crustal deformation: Relations to regional neotectonics," *Res. Square*, vol. 1, pp. 1–23, Jan. 2024, doi: 10.21203/rs.3.rs-3847651/v1.

[50] Q. Ou, G. Kulikova, J. Yu, A. Elliott, B. Parsons, and R. Walker, "Magnitude of the 1920 Haiyuan earthquake reestimated using seismological and geomorphological methods," *J. Geophys. Res., Solid Earth*, vol. 125, no. 8, pp. 1–28, Aug. 2020, doi: 10.1029/2019jb019244.

[51] (2024). *SCEDC Change History*. Accessed: Jun. 16, 2024. [Online]. Available: https://scedc.caltech.edu/eq-catalogs/change-history.html

[52] S. L. Ghose, T. A. Yap, A. Q. Byrne, H. Sulaeman, E. B. Rosenblum, A. Chan-Alvarado, S. Chaukulkar, E. Greenbaum, M. S. Koo, M. T. Kouete, K. Lutz, D. McAloose, A. J. Moyer, E. Parra, D. M. Portik, H. Rockney, A. G. Zink, D. C. Blackburn, and V. T. Vredenburg, "Continent-wide recent emergence of a global pathogen in African amphibians," *Frontiers Conservation Sci.*, vol. 4, Mar. 2023, Art. no. 1069490, doi: 10.3389/fcosc.2023.1069490.

[53] T. Li, R. Liu, and W. Qi, "Regional heterogeneity of migrant rent affordability stress in urban China: A comparison between skilled and unskilled migrants at prefecture level and above," *Sustainability*, vol. 11, no. 21, p. 5920, Oct. 2019, doi: 10.3390/su11215920.

[54] M. L. Lima, A. Romanelli, and H. E. Massone, "Decision support model for assessing aquifer pollution hazard and prioritizing groundwater resources management in the wet Pampa plain, Argentina," *Environ. Monitor. Assessment*, vol. 185, no. 6, pp. 5125–5139, Jun. 2013, doi: 10.1007/s10661-012-2930-4.

[55] W. Muttitanon, "Clustering analysis influenza disease to identify spatio-temporal spread pattern in Thailand," Dept. Fac. Eng., Mahidol Univ., Thailand, Tech. Rep. 5, 2021, doi: 10.52939/ijg.v17i5.2015.

[56] R. Sinha, S. Singh, K. Mishra, and S. Swarnkar, "Channel morphodynamics and sediment budget of the lower ganga river using a hydrogeomorphological approach," *Earth Surf. Processes Landforms*, vol. 48, no. 1, pp. 14–33, Jan. 2023, doi: 10.1002/esp.5325.

[57] R. Fernández-Álvarez and R. Fernández-Nava, "Adaptive co-management of urban forests: Monitoring reforestation programs in Mexico city," *Polibotánica*, vol. 49, pp. 243–258, Jan. 2020, doi: 10.18387/polibotanica.49.15.

[58] A. Mustafa, A. Van Rompaey, M. Cools, I. Saadi, and J. Teller, "Addressing the determinants of built-up expansion and densification processes at the regional scale," *Urban Stud.*, vol. 55, no. 15, pp. 3279–3298, Nov. 2018, doi: 10.1177/0042098017749176.

[59] B. V. Lancellotti, K. L. Underwood, J. N. Perdrial, A. W. Schroth, E. D. Roy, and C. E. Adair, "Complex drivers of riparian soil oxygen variability revealed using self-organizing maps," *Water Resour. Res.*, vol. 59, no. 6, pp. 1–18, Jun. 2023, doi: 10.1029/2022wr034022.

[60] P. Babuna, X. Yang, and D. Bian, "Water use inequality and efficiency assessments in the Yangtze river economic delta of China," *Water*, vol. 12, no. 6, p. 1709, Jun. 2020, doi: 10.3390/w12061709.

[61] Q. Xia, X. Zhang, Y. Hu, W. Tian, W. Miao, B. Wu, Y. Lai, J. Meng, Z. Fan, C. Zhang, L. Xin, J. Miao, Q. Wu, M. Jiao, L. Shan, N. Wang, B. Shi, and Y. Li, "The superposition effects of air pollution on government health expenditure in China—Spatial evidence from GeoDetector," *BMC Public Health*, vol. 22, no. 1, p. 1411, Dec. 2022, doi: 10.1186/s12889-022-13702-y.

[62] C. Vazquez, R. G. M. de Goede, G. W. Korthals, M. Rutgers, A. J. Schouten, and R. Creamer, "The effects of increasing land use intensity on soil nematodes: A turn towards specialism," *Funct. Ecol.*, vol. 33, no. 10, pp. 2003–2016, Oct. 2019, doi: 10.1111/1365-2435.13417.
[63] D. Schorlemmer and S. Wiemer, "Microseismicity data forecast rupture area," *Nature*, vol. 434, no. 7037, pp. 1086–1086, Apr. 2005, doi: 10.1038/4341086a.
[64] Y. Ogata and K. Katsura, "Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues," *Geophys. J. Int.*, vol. 113, no. 3, pp. 727–738, Jun. 1993, doi: 10.1111/j.1365-246x.1993.tb04663.x.
[65] M. Taroni, G. Vocalelli, and A. De Polis, "Gutenberg–Richter b-value time series forecasting: A weighted likelihood approach," *Forecasting*, vol. 3, no. 3, pp. 561–569, Aug. 2021, doi: 10.3390/forecast3030035.
[66] T. Utsu, Y. Ogata, R. S, and Matsu'ura, "The centenary of the Omori formula for a decay law of aftershock activity," *J. Phys. Earth*, vol. 43, no. 1, pp. 1–33, 1995, doi: 10.4294/jpe1952.43.1.
[67] P. Volant, J. Grasso, J. Chatelain, and M. Frogneux, "B-value, aseismic deformation and brittle failure within an isolated geological object: Evidences from a dome structure loaded by fluid extraction," *Geophys. Res. Lett.*, vol. 19, no. 11, pp. 1149–1152, Jun. 1992, doi: 10.1029/92gl01074.
[68] M. Yousefzadeh, S. A. Hosseini, and M. Farnaghi, "Spatiotemporally explicit earthquake prediction using deep neural network," *Soil Dyn. Earthq. Eng.*, vol. 144, May 2021, Art. no. 106663, doi: 10.1016/j.soildyn.2021.106663.
[69] J. H. Dieterich and D. E. Smith, "Nonplanar faults: Mechanics of slip and off-fault damage," *Pure Appl. Geophys.*, vol. 166, nos. 10–11, pp. 1799–1815, Oct. 2009, doi: 10.1007/s00024-009-0517-y.
[70] J. Martinsson and W. Törnman, "Modelling the dynamic relationship between mining induced seismic activity and production rates, depth and size: A mine-wide hierarchical model," *Pure Appl. Geophys.*, vol. 177, no. 6, pp. 2619–2639, Jun. 2020, doi: 10.1007/s00024-019-02378-y.
[71] M. Bohnhoff, P. Malin, J. T. Heege, J.-P. Deflandre, and C. Sicking, "Suggested best practice for seismic monitoring and characterization of non-conventional reservoirs," *1st Break*, vol. 36, no. 2, pp. 59–64, Feb. 2018, doi: 10.3997/1365-2397.n0070.
[72] D. L. Turcotte and J. B. Rundle, "Self-organized complexity in the physical, biological, and social sciences," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 1, pp. 2463–2465, Feb. 2002, doi: 10.1073/pnas.012579399.
[73] M. Fahandezhsadi and H. F. Sadi, "Earthquake magnitude prediction using probabilistic classifiers," *Res. Square*, vol. 1, pp. 1–11, Jun. 2020, doi: 10.21203/rs.3.rs-36094/v1.
[74] A. Saichev and D. Sornette, "Distribution of the largest aftershocks in branching models of triggered seismicity: Theory of the universal Båth law," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 5, May 2005, Art. no. 056127, doi: 10.1103/physreve.71.056127.
[75] Y. Radzyner, M. Galun, and B. Nadler, "A statistical approach to estimate seismic monitoring stations' biases and error levels," *Bull. Seismological Soc. Amer.*, vol. 113, no. 6, pp. 2596–2614, Dec. 2023, doi: 10.1785/0120230009.
[76] D. Faro, A. L. McGill, and R. Hastie, "Naïve theories of causal force and compression of elapsed time judgments," *J. Personality Social Psychol.*, vol. 98, no. 5, pp. 683–701, 2010, doi: 10.1037/a0019261.
[77] A. Nguyen, S. Chatterjee, S. Weinzierl, L. Schwinn, M. Matzner, and B. Eskofier, "Time matters: Time-aware LSTMs for predictive business process monitoring," in *Proc. Process Mining Workshops*. Cham, Switzerland: Springer, 2021, pp. 112–123, doi: 10.1007/978-3-030-72693-5_9.
[78] M. A. Salam, L. Ibrahim, and D. S. Abdelminaam, "Earthquake prediction using hybrid machine learning techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 5, pp. 654–665, 2021, doi: 10.14569/ijacsa.2021.0120578.
[79] M. Rosenau and O. Oncken, "Fore-arc deformation controls frequency-size distribution of megathrust earthquakes in subduction zones," *J. Geophys. Res., Solid Earth*, vol. 114, no. B10, pp. 1–12, Oct. 2009, doi: 10.1029/2009jb006359.
[80] D. Muhammad, I. Ahmad, M. I. Khalil, W. Khalil, and M. O. Ahmad, "A generalized deep learning approach to seismic activity prediction," *Appl. Sci.*, vol. 13, no. 3, p. 1598, Jan. 2023, doi: 10.3390/app13031598.
[81] G. Teng, J. W. Baker, and D. J. Wald, "Evaluation of intensity prediction equations (IPEs) for small-magnitude earthquakes," *Bull. Seismological Soc. Amer.*, vol. 112, no. 1, pp. 316–330, Feb. 2022, doi: 10.1785/0120210150.
[82] K. M. Asim, S. S. Moustafa, I. A. Niaz, E. A. Elawadi, T. Iqbal, and F. Martínez-Álvarez, "Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus," *Soil Dyn. Earthq. Eng.*, vol. 130, May 2020, Art. no. 105932, doi: 10.1016/j.soildyn.2019.105932.
[83] M. N. Brykov, I. Petryshynets, C. I. Pruncu, V. G. Efremenko, D. Y. Pimenov, K. Giasin, S. A. Sylenko, and S. Wojciechowski, "Machine learning modelling and feature engineering in seismology experiment," *Sensors*, vol. 20, no. 15, p. 4228, Jul. 2020, doi: 10.3390/s20154228.
[84] R. Jena, A. Shanableh, R. Al-Ruzouq, B. Pradhan, M. B. A. Gibril, M. A. Khalil, O. Ghorbanzadeh, G. P. Ganapathy, and P. Ghamisi, "Explainable artificial intelligence (XAI) model for earthquake spatial probability assessment in Arabian peninsula," *Remote Sens.*, vol. 15, no. 9, p. 2248, Apr. 2023, doi: 10.3390/rs15092248.
[85] S. Tsuboi, M. Saito, and M. Kikuchi, "Real-time earthquake warning by using broadband P waveform," *Geophys. Res. Lett.*, vol. 29, no. 24, p. 40, Dec. 2002, doi: 10.1029/2002gl016101.
[86] P. Chittora, T. Chakrabarti, P. Debnath, A. Gupta, P. Chakrabarti, S. P. Praveen, M. Margala, and A. A. Elngar, "Experimental analysis of earthquake prediction using machine learning classifiers, curve fitting, and neural modeling," *Res. Square*, vol. 1, pp. 1–26, Sep. 2022, doi: 10.21203/rs.3.rs-1896823/v2.
[87] C. Hibert, D. Michéa, F. Provost, J.-P. Malet, and M. Geertsema, "Exploration of continuous seismic recordings with a machine learning approach to document 20 yr of landslide activity in Alaska," *Geophys. J. Int.*, vol. 219, no. 2, pp. 1138–1147, Nov. 2019, doi: 10.1093/gji/ggz354.
[88] C. Groult, C. Hibert, J.-P. Malet, and F. Provost, "Identifying landslides from massive seismic data and machine learning: The case of the European Alps," in *Proc. 25th EGU General Assembly*, 2023, p. 1, doi: 10.5194/egusphere-egu23-7062.
[89] A. Jufriansah, A. Khusnani, S. Saputra, and D. S. Wahab, "Forecasting the magnitude category based on the flores sea earthquake," *Jurnal RESTI Rekayasa Sistem dan Teknologi Informasi*, vol. 7, no. 6, pp. 1439–1447, Dec. 2023, doi: 10.29207/resti.v7i6.5495.
[90] S. Biswas, D. Kumar, and U. K. Bera, "Prediction of earthquake magnitude and seismic vulnerability mapping using artificial intelligence techniques: A case study of Turkey," *Res. Square*, vol. 1, pp. 1–54, Jan. 2023, doi: 10.21203/rs.3.rs-2863887/v1.
[91] W. Liu, S. Liu, S. G. Hassan, Y. Cao, L. Xu, D. Feng, L. Cao, W. Chen, Y. Chen, J. Guo, T. Liu, and H. Zhang, "A novel hybrid model to predict dissolved oxygen for efficient water quality in intensive aquaculture," *IEEE Access*, vol. 11, pp. 29162–29174, 2023, doi: 10.1109/ACCESS.2023.3260089.
[92] O. Marc, P. Meunier, and N. Hovius, "Prediction of the area affected by earthquake-induced landsliding based on seismological parameters," *Natural Hazards Earth Syst. Sci.*, vol. 17, no. 7, pp. 1159–1175, Jul. 2017, doi: 10.5194/nhess-17-1159-2017.
[93] A. Merghadi, B. Abderrahmane, and D. Tien Bui, "Landslide susceptibility assessment at Mila basin (Algeria): A comparative assessment of prediction capability of advanced machine learning methods," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 7, p. 268, Jul. 2018, doi: 10.3390/ijgi7070268.
[94] J. Mahmoudi, M. A. Arjomand, M. Rezaei, and M. H. Mohammadi, "Predicting the earthquake magnitude using the multilayer perceptron neural network with two hidden layers," *Civil Eng. J.*, vol. 2, no. 1, pp. 1–12, Jan. 2016, doi: 10.28991/cej-2016-00000008.
[95] P. Lara, Q. Bletery, J. Ampuero, A. Inza, and H. Tavera, "Earthquake early warning starting from 3 s of records on a single station with machine learning," *J. Geophys. Res., Solid Earth*, vol. 128, no. 11, pp. 1–19, Nov. 2023, doi: 10.1029/2023jb026575.
[96] V. H. A. Dias and A. R. R. Papa, "Application of neural networks in probabilistic forecasting of earthquakes in the southern California region," *Int. J. Geosci.*, vol. 9, no. 6, pp. 397–413, 2018, doi: 10.4236/ijg.2018.96025.
[97] Y.-J. Chuo, "Earthquake shake detecting by data mining from social network platforms," *Appl. Sci.*, vol. 10, no. 3, p. 812, Jan. 2020, doi: 10.3390/app10030812.
[98] B. Şengezer, A. Ansal, and Ö. Bilen, "Evaluation of parameters affecting earthquake damage by decision tree techniques," *Natural Hazards*, vol. 47, no. 3, pp. 547–568, Dec. 2008, doi: 10.1007/s11069-008-9238-2.
[99] M. Cin and Ş. A. Değirmençay, "Decision-making of middle school students during an earthquake," *Rev. Int. Geographical Educ. Online*, vol. 8, no. 3, pp. 556–570, Dec. 2018, doi: 10.33403/rigeo.505271.

[100] B. Asgarian and B. Ordoubadi, "Probabilistic evaluation of seismic performance for a steel moment frame using damage indices," in *Proc. 6th Int. Conf. Comput. Methods Structural Dyn. Earthq. Eng. (COMPDYN)*. Athens, Greece: Research School of Civil Engineering National Technical University of Athens, 2017, pp. 1779–1796, doi: 10.7712/120117.5528.17007.

[101] W. Liu, K. He, Q. Gao, and C. Y. Liu, "Application of EMD-based SVD and SVM to coal-gangue interface detection," *J. Appl. Math.*, vol. 2014, pp. 1–6, Apr. 2014, doi: 10.1155/2014/283606.

[102] H.-M. Zhang, S. Zhou, C. Xu, and Y.-R. Zhang, "A real-time automatic method for target locating under unknown wall characteristics in through-wall imaging," *Prog. Electromagn. Res. M*, vol. 89, pp. 189–197, 2020, doi: 10.2528/pierm19111101.

[103] M. Azadkia, "Optimal choice of k for k-nearest neighbor regression," 2019, *arXiv:1909.05495*.

[104] R. E. Yancey, B. Xin, and N. Matloff, "Modernizing k-nearest neighbors," *Stat*, vol. 10, no. 1, p. e335, Dec. 2021, doi: 10.1002/sta4.335.

[105] T. Setiyorini and R. T. Asmono, "Penerapan metode K-nearest neighbor Dan Gini index Pada Klasifikasi Kinerja Siswa," *Jurnal Techno Nusa Mandiri*, vol. 16, no. 2, pp. 121–126, Sep. 2019, doi: 10.33480/techno.v16i2.747.

[106] J. Huang, T. C. Hales, R. Huang, N. Ju, Q. Li, and Y. Huang, "A hybrid machine-learning model to estimate potential debris-flow volumes," *Geomorphology*, vol. 367, Oct. 2020, Art. no. 107333, doi: 10.1016/j.geomorph.2020.107333.

[107] P. Xiong, D. Marchetti, A. De Santis, X. Zhang, and X. Shen, "SafeNet: SwArm for earthquake perturbations identification using deep learning networks," *Remote Sens.*, vol. 13, no. 24, p. 5033, Dec. 2021, doi: 10.3390/rs13245033.

[108] K. Morfidis and K. Kostinakis, "Special issue on application of artificial neural networks for seismic design and assessment," *Appl. Sci.*, vol. 12, no. 12, p. 6192, Jun. 2022, doi: 10.3390/app12126192.

[109] H. Qu, T. Feng, Y. Zhang, and Y. Wang, "Ensemble learning with stochastic configuration network for noisy optical fiber vibration signal recognition," *Sensors*, vol. 19, no. 15, p. 3293, Jul. 2019, doi: 10.3390/s19153293.

[110] S. Yang, A. Jin, and Y. Xu, "Recognition of oil and gas reservoir space based on deep learning," *E3S Web Conf.*, vol. 267, p. 01038, Jun. 2021, doi: 10.1051/e3sconf/202126701038.

[111] Z. E. Ross, M. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learning," *Bull. Seismological Soc. Amer.*, vol. 108, no. 5A, pp. 2894–2901, Oct. 2018, doi: 10.1785/0120180080.

[112] L. Yue, J. Qu, S. Zhou, B. Qu, Y. Zhang, and Q. Xu, "Seismic event classification based on a two-step convolutional neural network," *J. Seismol.*, vol. 27, no. 3, pp. 527–535, Jun. 2023, doi: 10.1007/s10950-023-10153-9.

[113] N. Yamaga and Y. Mitsui, "Machine learning approach to characterize the postseismic deformation of the 2011 Tohoku-Oki earthquake based on recurrent neural network," *Geophys. Res. Lett.*, vol. 46, no. 21, pp. 11886–11892, Nov. 2019, doi: 10.1029/2019gl084578.

[114] J. Chen and N. S. Chaudhari, "Bidirectional segmented-memory recurrent neural network for protein secondary structure prediction," *Soft Comput.*, vol. 10, no. 4, pp. 315–324, Feb. 2006, doi: 10.1007/s00500-005-0489-5.

[115] M. Albaba, A. Qassab, and A. Yilmaz, "Human activity recognition and classification using of convolutional neural networks and recurrent neural networks," *Int. J. Appl. Math. Electron. Comput.*, vol. 8, no. 4, pp. 185–189, Dec. 2020, doi: 10.18100/ijamec.803105.

[116] T. Y. Hsu and A. Pratomo, "Early peak ground acceleration prediction for on-site earthquake early warning using LSTM neural network," *Frontiers Earth Sci.*, vol. 10, pp. 1–17, Jul. 2022, doi: 10.3389/feart.2022.911947.

[117] C. Cao, X. Wu, L. Yang, Q. Zhang, X. Wang, D. A. Yuen, and G. Luo, "Long short-term memory networks for pattern recognition of synthetical complete earthquake catalog," *Sustainability*, vol. 13, no. 9, p. 4905, Apr. 2021, doi: 10.3390/su13094905.

[118] R. Abri and H. Artuner, "LSTM-based deep learning methods for prediction of earthquakes using ionospheric data," *Gazi Univ. J. Sci.*, vol. 35, no. 4, pp. 1417–1431, Dec. 2022, doi: 10.35378/gujs.950387.

[119] W. Wang, G.-F. Wu, and X.-Y. Song, "The application of neural networks to comprehensive prediction by seismology prediction method," *Acta Seismologica Sinica*, vol. 13, no. 2, pp. 210–215, Mar. 2000, doi: 10.1007/s11589-000-0012-0.

[120] S. Akter, M. J. A. Patwary, and S. Hossain, "Earthquake prediction by using evidential reasoning approach," *Int. J. Res. Eng. Technol.*, vol. 4, no. 12, pp. 149–151, Dec. 2015, doi: 10.15623/ijret.2015.0412028.

[121] G. Chouliaras, "Investigating the earthquake catalog of the national observatory of Athens," *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 3, pp. 905–912, Jun. 2009, doi: 10.5194/nhess-9-905-2009.

[122] A. A. Alabi, O. D. Akinyemi, and A. Adewale, "Seismicity pattern in Southern Africa from 1986 to 2009," *Earth Sci. Res.*, vol. 2, no. 2, p. 1, Aug. 1986.

[123] J. Yin, M. Denolle, and B. He, "A multi-task encoder–decoder to separating earthquake and ambient 1 noise signal in seismograms," *Authorea*, vol. 1, pp. 1–35, Jul. 2022, doi: 10.1002/essoar.10510129.1.

[124] M. A. N. Jessee, M. W. Hamburger, K. Allstadt, D. J. Wald, S. M. Robeson, H. Tanyas, M. Hearne, and E. M. Thompson, "A global empirical model for near-real-time assessment of seismically induced landslides," *J. Geophys. Res., Earth Surf.*, vol. 123, no. 8, pp. 1835–1859, Aug. 2018, doi: 10.1029/2017jf004494.

[125] A. Mignan and M. Broccardo, "Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations," *Seismological Res. Lett.*, vol. 91, no. 4, pp. 2330–2342, Jul. 2020, doi: 10.1785/0220200021.

[126] A. Rawat, R. S. Chatterjee, D. Kumar, H. Kumar, and S. Suman, "Can site specific parameters help to identify seismically induced damage pattern: An assessment," *Res. Square*, vol. 1, pp. 1–15, Mar. 2023, doi: 10.21203/rs.3.rs-2721236/v1.

[127] S. de Vilder, C. Massey, B. Lukovic, T. Taig, and R. Morgenstern, "What drives landslide risk? Disaggregating risk analyses, an example from the Franz Josef Glacier and Fox Glacier valleys, New Zealand," *Natural Hazards Earth Syst. Sci.*, vol. 22, no. 7, pp. 2289–2316, Jul. 2022, doi: 10.5194/nhess-22-2289-2022.

[128] S. H. Elkhouly and G. Ali, "Seismic discrimination between nuclear explosions and natural earthquakes using multi-machine learning techniques," *Pure Appl. Geophys.*, vol. 1, pp. 1–12, Apr. 2024, doi: 10.1007/s00024-024-03463-7.

[129] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," *TELKOMNIKA Telecommunication Comput. Electron. Control*, vol. 18, no. 3, p. 1331, Jun. 2020, doi: 10.12928/telkomnika.v18i3.14756.

[130] J. H. Kim, J. Kim, G. Lee, and J. Park, "Machine learning-based models for accident prediction at a Korean container port," *Sustainability*, vol. 13, no. 16, p. 9137, Aug. 2021, doi: 10.3390/su13169137.

[131] M. Puth, M. Neuhäuser, and G. D. Ruxton, "On the variety of methods for calculating confidence intervals by bootstrapping," *J. Animal Ecol.*, vol. 84, no. 4, pp. 892–897, Jul. 2015, doi: 10.1111/1365-2656.12382.

[132] T. Chen and C. Guestrin, *XGBoost*. New York, NY, USA: ACM, 2016, doi: 10.1145/2939672.2939785.

[133] H. Sharma, H. Harsora, and B. Ogunleye, "An optimal house price prediction algorithm: XGBoost," *Analytics*, vol. 3, no. 1, pp. 30–45, Jan. 2024, doi: 10.3390/analytics3010003.

[134] J.-J. Liu and J.-C. Liu, "Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization," *Geofluids*, vol. 2022, no. 1, 2022, Art. no. 2263329, doi: 10.1155/2022/2263329.

[135] Y. Liu, S. Zhao, W. Du, Z. Tian, H. Chi, C. Chao, and W. Shen, "Applying interpretable machine learning algorithms to predict risk factors for permanent stoma in patients after TME," *Frontiers Surg.*, vol. 10, pp. 1–10, Mar. 2023, doi: 10.3389/fsurg.2023.1125875.

[136] X. Li, "Financial fraud: Identifying corporate tax report fraud under the XGBoost algorithm," *Eur. Alliance Innov.*, vol. 10, pp. 1–7, May 2023, doi: 10.4108/eetsis.v10i3.3033.

[137] V. B. Lahari, T. Amrutha, S. N. B. T. Reddy, P. D. Leela, and Y. V. Narayana, "A machine learning approach for intrusion detection," *EPRA Int. J. Res. Develop. (IJRD)*, vol. 8, no. 10, pp. 282–289, Nov. 2023, doi: 10.36713/epra14771.

[138] S. K. Devana, A. A. Shah, C. Lee, A. R. Jensen, E. Cheung, M. van der Schaar, and N. F. SooHoo, "Development of a machine learning algorithm for prediction of complications and unplanned readmission following primary anatomic total shoulder replacements," *J. Shoulder Elbow Arthroplasty*, vol. 6, Apr. 2022, Art. no. 247154922210754, doi: 10.1177/24715492221075444.

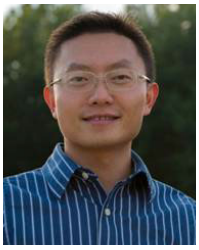
[139] K. Gu, J. Wang, H. Qian, and X. Su, "Study on intelligent diagnosis of rotor fault causes with the PSO-XGBoost algorithm," *Math. Problems Eng.*, vol. 2021, pp. 1–17, Apr. 2021, doi: 10.1155/2021/9963146.

[140] M. Ashraf, K. Le, and X. Huang, "Information gain and adaptive neuro-fuzzy inference system for breast cancer diagnoses," in *Proc. 5th Int. Conf. Comput. Sci. Converg. Inf. Technol.*, Nov. 2010, pp. 911–915, doi: 10.1109/ICCIT.2010.5711189.

[141] M. M. H. Joy, M. Hasan, A. S. M. Miah, A. Ahmed, S. A. Tohfa, M. F. I. Bhuaiyan, A. Zannat, and M. M. Rashid, "Multiclass MI-task classification using logistic regression and filter bank common spatial patterns," in *Computing Science, Communication and Security*. Singapore: Springer, 2020, doi: 10.1007/978-981-15-6648-6_13.
[142] J. Lee, H. Jang, S. Ha, and Y. Yoon, "Android malware detection using machine learning with feature selection based on the genetic algorithm," *Mathematics*, vol. 9, no. 21, p. 2813, Nov. 2021, doi: 10.3390/math9212813.
[143] E. K. Shaveta Tatwani, "A stable SVM-RFE feature selection method for gene expression data," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 2110–2115, 2019.
[144] Y.-H. Chen and J.-L. Chen, "AI@ntiPhish—Machine learning mechanisms for cyber-phishing attack," *IEICE Trans. Inf. Syst.*, vol. 102, no. 5, pp. 878–887, 2019, doi: 10.1587/transinf.2018nti0001.
[145] D. Bashir, G. D. Montañez, S. Sehra, P. S. Segura, and J. Lauw, "An information-theoretic perspective on overfitting and underfitting," in *AI 2020: Advances in Artificial Intelligence*. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-64984-5_27.

CEMIL EMRE YAVAS (Member, IEEE) received the B.Sc. degree in industrial engineering from Bosphorus University and the M.B.A. degree from Istanbul Bilgi University.

He is an Industrial Engineer with over 30 years of global experience in the IT industry, having been actively involved since the beginning of the internet era. He founded Turkline, a leading digital services company, where he managed a diverse team of over 100 professionals and led numerous award-winning projects. He excels in full-stack software development, with specific expertise in mobile device development and home automation systems. He serves as the manager for prestigious IT investment trusts, overseeing projects primarily focused on enterprise resource planning (ERP) systems. His leadership in this area has significantly contributed to the strategic enhancement of business processes and operational efficiency for numerous organizations. He has garnered over 50 international awards for web design and development, including prestigious accolades, such as W3 Awards, Interactive Media Awards (IMA), and Golden Spider Awards.



LEI CHEN (Senior Member, IEEE) received the Ph.D. degree in computer science and software engineering from Auburn University. He is currently a tenured Professor and the Graduate Program Director with Georgia Southern University. He has authored over 100 peer-reviewed scholarly works, including notable publications in the IEEE INTERNET OF THINGS JOURNAL and *Mobile Networks and Applications* (Springer). His book *Wireless Network Security: Theories and Practices* (Springer), has been widely recognized in the field. His contributions extend to serving as an editor for high-impact journals and chairing numerous international conferences, with his work supported by U.S. National Security Agency (NSA) and the National Science Foundation (NSF).

CHRISTOPHER KADLEC received the Ph.D. degree in information systems from the Terry College of Business, University of Georgia.

In 2007, he joined the Department of Information Technology, Georgia Southern University, where he is currently a Professor. He worked in IT support prior to his graduate work with Georgia Southern University and the University of Mississippi. He has also worked in IT support for more than 20 years and has taught in the field for more than 15 years. He is an advocate for students and the field of information technology. To help both of these, he is working on articulation agreements between universities and community colleges, ensuring that students can get the latest skills and the best jobs while satisfying the needs of the business community.

YIMING JI received the B.S. degree and first M.S. degree in aerodynamics, the second M.S. degree in computer science, and the Ph.D. degree in computer science. Since 2019, he has been a tenured Professor and the Chair of the Department of Information Technology at Georgia Southern University. Formerly the Director of the Computational Science Program at USC Beaufort, he has secured over $12 million in grant funding and consistently received external research funding, including NSF NeTS and NSF EPSCOR RII awards. He has taught over 100 class sections covering 22 subjects. He was named a "Rising Star" by the University of South Carolina System in 2010. He received the South Carolina Governor's Award for Excellence in Scientific Research in 2016.