SCIENTIFIC Publishing Limited

Prevention and Treatment of Natural Disasters

https://ojs.ukscip.com/index.php/ptnd/

comunication

A Mathematical Exploration of Pre-Earthquake Seismicity

Tengiz Kiria [1]* , Tamaz Chelidze [1] and Jemal KiriaL [1]

[1] M. Nodia Institute of Geophysics, Tbilisi State University, Tbilisi, Georgia, 0160,
* Correspondence: kiria8@gmail.com;

Received: 6 September 2024; **Revised**: 6 October 2024; **Accepted**: 30 October 2024; **Published**: 6 December 2024

Abstract: Understanding the dynamics of pre-earthquake seismicity is crucial for advancing earthquake forecast and risk assessment. In this paper, we embark on a mathematical exploration of pre-earthquake seismic activity, aiming to elucidate the underlying patterns and mechanisms leading up to major seismic events. Leveraging probabilistic modeling techniques, we analyze historical seismic data to identify precursory signals and assess their predictive value. Our investigation encompasses the study of foreshock activity, preceding earthquakes, shedding light on the temporal and spatial characteristics of seismic activity prior to the main shock events. Through mathematical modeling and simulation, we aim to unveil the complex interplay of factors contributing to pre-earthquake seismicity, with implications for enhancing earthquake forecasting capabilities and disaster preparedness efforts. This research contributes to the ongoing endeavor to unravel the mysteries of earthquake occurrence, ultimately striving towards a more resilient and proactive approach to seismic risk management. This study introduces a novel mathematical framework for analyzing pre-earthquake seismic activity, leveraging a 15-day foreshock window and machine learning techniques to predict seismic events. The approach addresses gaps in existing methodologies by incorporating comprehensive feature engineering and a robust random forest classification model. Additionally, we draw upon insights from prior studies, such as Kumazawa et al. (2020) and Luo et al. (2023), which emphasize the significance of spatial-temporal dynamics and natural orthogonal expansion methods in identifying seismic precursors. By integrating interdisciplinary methodologies and advanced machine learning models, this study bridges critical gaps in real-time predictive capabilities, offering a tailored approach for region-specific seismic forecasting.

1. Introduction

The study of pre-earthquake seismicity involves examining patterns and mechanisms that precede significant seismic events. By identifying precursory signals, we aim to improve earthquake forecast and enhance disaster preparedness. This research explores foreshock activity as a potential indicator of imminent major earthquakes, focusing on the temporal and spatial characteristics of seismic activity leading up to significant events.

The exploration of pre-earthquake seismicity has garnered significant interest, leading to diverse approaches in

understanding and forecasting seismic events. The study by T. Kiria and T. Chelidze emphasizes mathematical and probabilistic modeling to identify precursory signals, aiming to enhance earthquake forecasting and disaster preparedness [1]. Martinelli et al. (2023) focus on the spatial-temporal dynamics and physical parameters of seismic activity, integrating geophysical and geochemical observations for better forecast accuracy [2]. Kumazawa et al. (2020) analyze seismicity anomalies before the 2011 Tohoku- Oki earthquake, employing a two-stage stationary epidemic-type aftershock sequence model to link stress changes to seismic activity [3]. Luo et al. (2023) utilize natural orthogonal expansion on earthquake frequencies to identify pre-quake anomalies, demonstrating the method's reliability in predicting strong earthquakes [4]. Jiao and Shan (2022) introduce the Temporal Integrated Anomaly (TIA) method using remote sensing data to improve the statistical significance of pre-seismic anomalies [5]. Freund et al. (2021) review various pre- earthquake phenomena based on the peroxy defects theory, providing a comprehensive understanding of the physical and chemical processes preceding earthquakes [6]. Each study contributes uniquely to the body of knowledge, advancing the potential for effective earthquake forecast and risk management. Previous research, such as that by Martinelli et al.(2023)and Kumazawa et al.(2020),has provided valuable insights into pre-earthquake seismicity [2,3]. However,these studies often lack real-time predictive capabilities and fail to account for the variability in foreshock patterns across different regions. Our study addresses these gaps by developing a machine learning model tailored for the Southern California data. This study introduces an innovative method for analyzing pre-earth quake seismicity, focusing on temporal and spatial foreshock patterns and their predictive value.

2. Data and Methodology

We utilized historical seismic data from various regions of California to conduct a comprehensive analysis of pre-earthquake seismicity. The data includes information on earthquake magnitudes, dates, and locations.

Flowchart: Include steps such as data preprocessing, feature engineering, rolling window analysis, and model training.

Data Preparation: Filtering earthquakes with magnitudes greater than 3 and extracting foreshocks within a 15-day windows before each significant event. The 15-day window was selected based on empirical evidence from the prior studies (e.g., Helmstetter & Sornette, 2002) [7], which demonstrate that significant seismic precursors often cluster within this period.

Feature Engineering: Creating features such as the number of foreshocks, cumulative magnitude of foreshocks, mean and variance of foreshock magnitudes, and time intervals between foreshocks.

Statistical Analysis: Conducting correlation and regression analyses to identify patterns and relationships.

Visualization: Graphically representing foreshock timing trends and cumulative activity.

3. Results

3.1. Data Preparation and Feature Extraction

We used data from the Southern California Earthquake Data Center, namely the 2021-2024 earthquake records - https://shorturl.at/qapSr.

We filtered significant earthquakes (magnitude > 3) and extracted foreshocks within the 15- day window before each event. Key features were calculated to facilitate statistical analysis.

From the calculations performed, we derived 45,699 new records from the original database, which contained 46,558 records. For these new records, we calculated the following features based on the data from the previous 15

days: Number of Foreshocks, Cumulative Magnitude,

Mean Magnitude, Variance in Magnitude, Mean Time Interval, and Variance in Time

Intervals. These features will serve as input data for our model, with Magnitude (MAG) as the output variable, effectively preparing our dataset for regression analysis.

3.2 Date Range for Rolling Window

For each day i, identified by its date D i , the rolling window spans from:

$$D_{start} = D_i - (R + 1) \quad (1)$$

to

$$D_{end} = D_i - 1 \quad (2)$$

Where:

R is the rolling window period (15 days in this case).

D_{start} and D_{end} define the start and end dates for data included in the rolling calculations, excluding the current day Di to avoid look-ahead bias.

3.3 Number of Foreshocks

The number of foreshocks in the window is simply the count of seismic events within the defined date range:

$$N_{foreshocks} = \text{count of events where } D_{event} > D_{start} \text{ and } D_{event} \leq D_{end} \quad (3)$$

3.4 Mean Magnitude of Foreshocks

The mean magnitude of foreshocks is calculated as the average of magnitudes for all events Mevent in the window Dstart to Dend :

$$\bar{M} = \frac{1}{N_{foreshoks}} \sum_{D_{event} > D_{start} \text{ and } D_{event} \leq D_{end}} M_{event} \quad (4)$$

Where Nfores*h*oks is determined from formula (3).

3.5. Variance of Magnitude

The variance of magnitudes in the window is computed to understand the variability in earthquake sizes:

$$\sigma_M^2 = \frac{1}{N_{foreshoks}-1} \sum_{D_{event} > D_{start} \leq D_{end}} (M_{event} - \bar{M})^2 \quad (5)$$

Where M— is determined from formula (4).

3.6 Mean Time Interval Between Foreshocks

This metric calculates the average time interval between consecutive seismic events within the window:

$$\bar{T} = \frac{1}{N_{\text{forecshocks}} - 1} \sum_{k=1}^{N_{\text{foreshocks}} - 1} (t_{k+1} - t_k) \quad (6)$$

Where tk and tk+1 are the times of consecutive foreshocks.

3.7 Variance of Time Intervals

The variance of time intervals measures the spread in time intervals between seismic events:

$$\sigma_T^2 = \frac{1}{N_{foreshocks} - 1} \sum_{k=1}^{N_{foreshocks}-1} (t_{k+1} - t_k - \bar{T})^2 \quad (7)$$

where $\bar{T}$ is determined from formula (6)

We exclude the current day's data to prevent a model bias. This setup ensures that the

model's predictions are based solely on historical data, enhancing the reliability and validity of the predictive analytics. When writing about these formulas in your article, emphasize how they help to characterize and quantify seismic patterns preceding significant earthquakes

without relying on the immediate data of the day in question.

From the source of the catalog presented above, we received a base information treated with formulas (1-7). See the data obtained as a result of calculations on the link.

https://ig-geophysics.grena.ge/processed_seismic_data.xlsx

3.7.1 Distribution of Foreshocks

The time intervals between foreshocks and the main event show that foreshock activity increases as the main event approaches (**Figure.1**). The distribution is determined by formula (3).

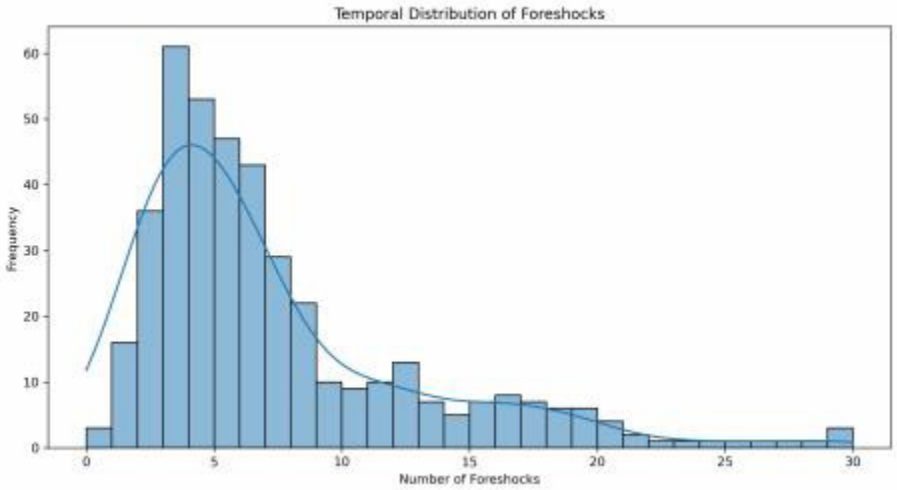


Figure 1: Distribution of Foreshocks

3.7.2 Cumulative Foreshock Activity

The cumulative magnitude of foreshocks calculated as (4) tends to accumulate more rapidly as the time to the main event decreases, indicating increasing seismic (foreshock) activity (**Figure 2**).

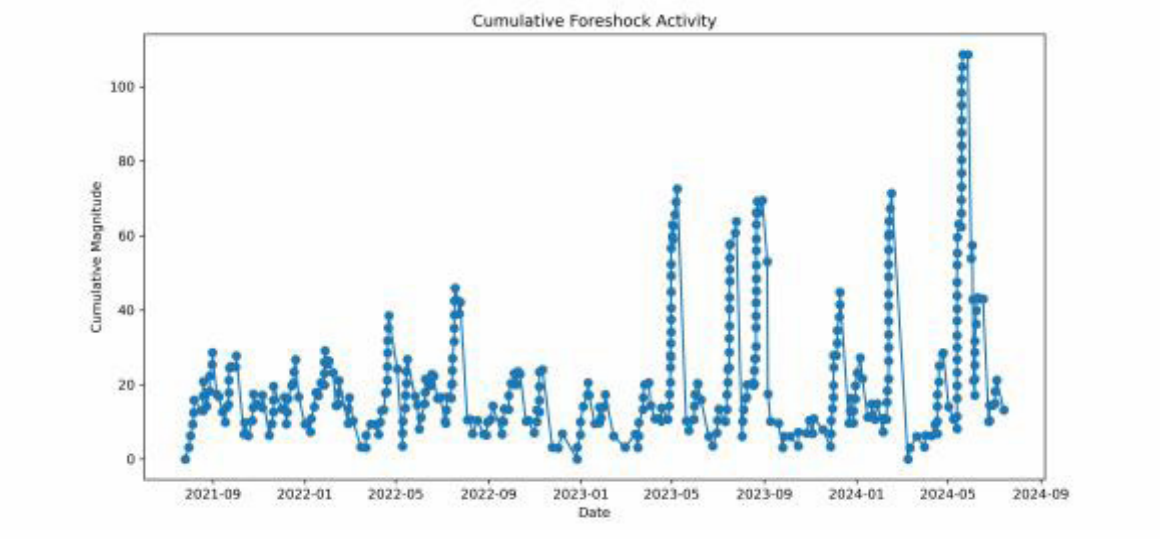


Figure 2: Cumulative Foreshock Activity — This graph clearly demonstrates that the cumulative magnitude of foreshocks increases as we approach seismic events with magnitudes greater than 3. This trend underscores the potential predictive value of monitoring cumulative foreshock magnitudes in earthquake forecasting.

3.7.3 Density Analysis

The density of foreshocks calculated as the number of foreshocks in the time interval, increases in shorter time windows before the significant event, suggesting clustering of foreshocks leading up to the main event (**Figure 3**).

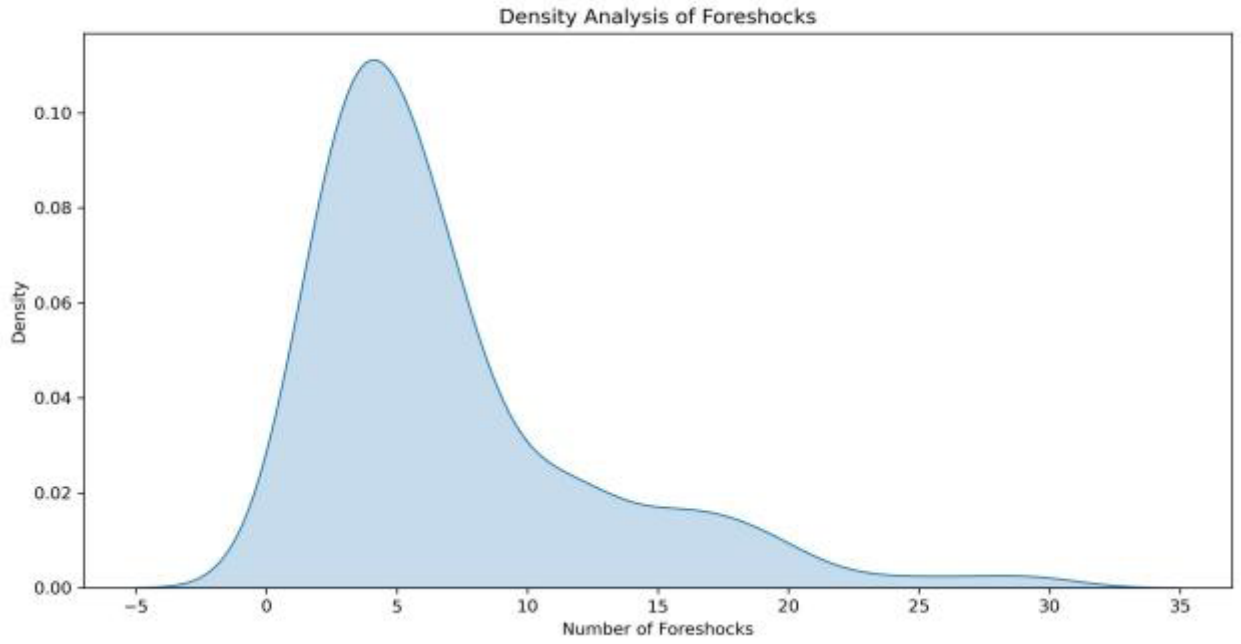


Figure 3: Density analysis of foreshocks and aftershocks

3.8. Machine learning and modeling

Here's a summary of the Machine Leaning process: we filter out all records which contain the event M > 3 from the catalog which totals 414 records. Given that the number of

precursory events' records with M < 3 is significantly larger, we plan to randomly select 4 to 5 times more events of M<3 from these records. Thus, the total number of records for

training amount to 2, 100.

Since we are trying to guess stronger magnitudes, we converted the output column to binary format (if MAG>=3, 1, MAG<3, 0). The final output column in the machine learning

model will contain 0s and 1s.

See the training data on this link: https://ig-geophysics.grena.ge/Training_data.xlsx input and output variables:

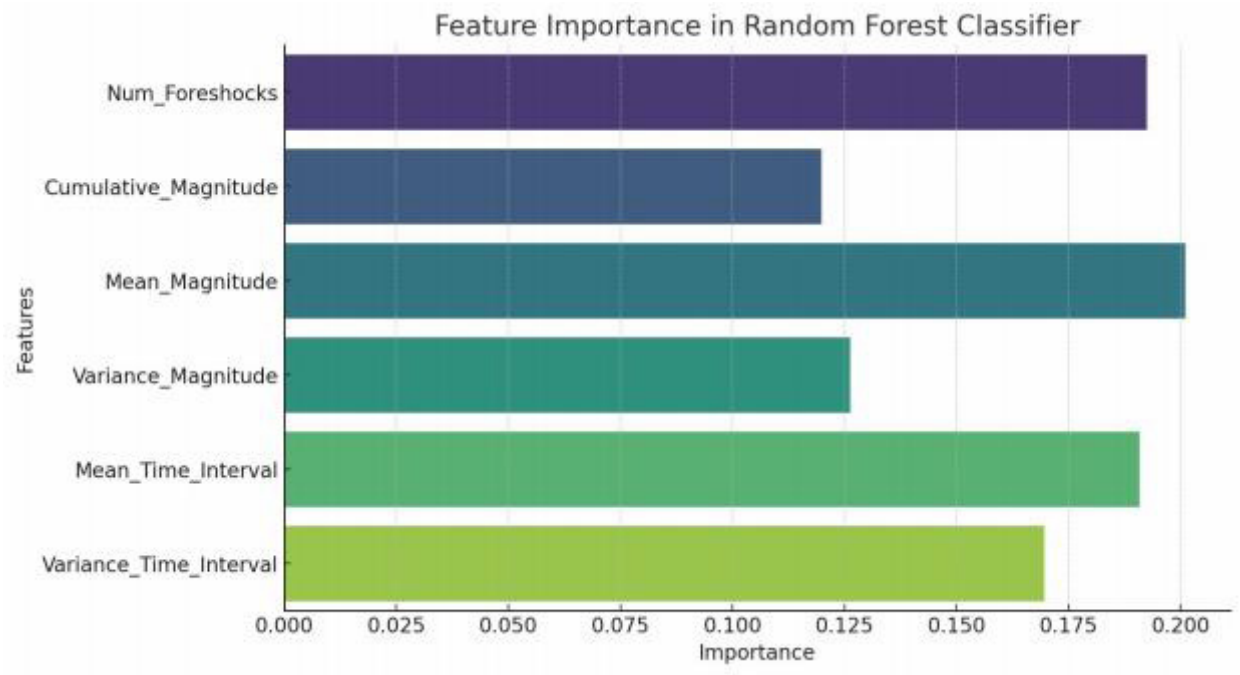
Input variables include Num_Foreshocks, Cumulative_Magnitude, Mean_Magnitude,Variance_Magnitude, Mean_Time_Interval, and Variance_Time_Interval. The output variable is binary, indicating whether the event is a major (MAG ≥ 3) or minor (MAG < 3) earthquake.

1. Data Splitting: The dataset was divided into training and testing sets, using a 60/40 split. This means 60% of the data was used for training the model, and 40% was

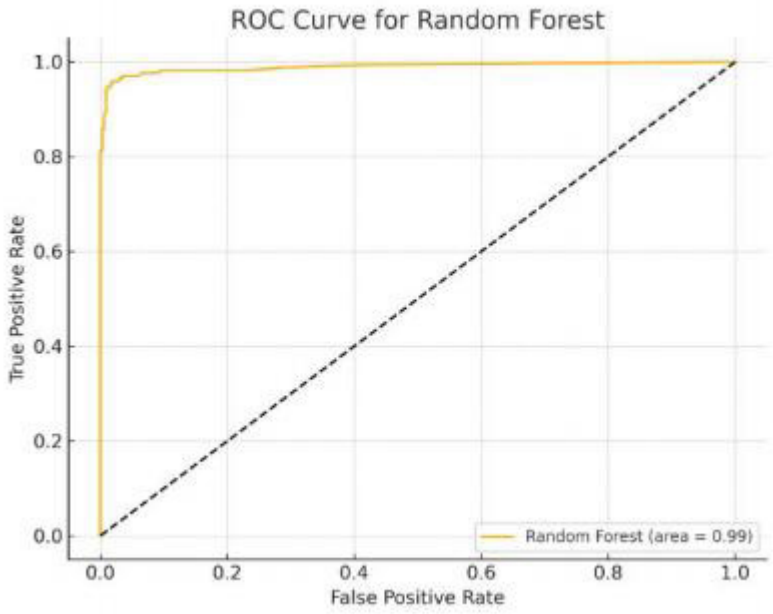
reserved for testing. The 60/40 split was chosen to ensure sufficient data for testing, given the smaller dataset size compared to traditional studies: this split ratio is widely

2. Model Training: The Random Forest classifier was trained using only the training data. This involves the model learning to make predictions based solely on the information (features and labels) provided in the training set.

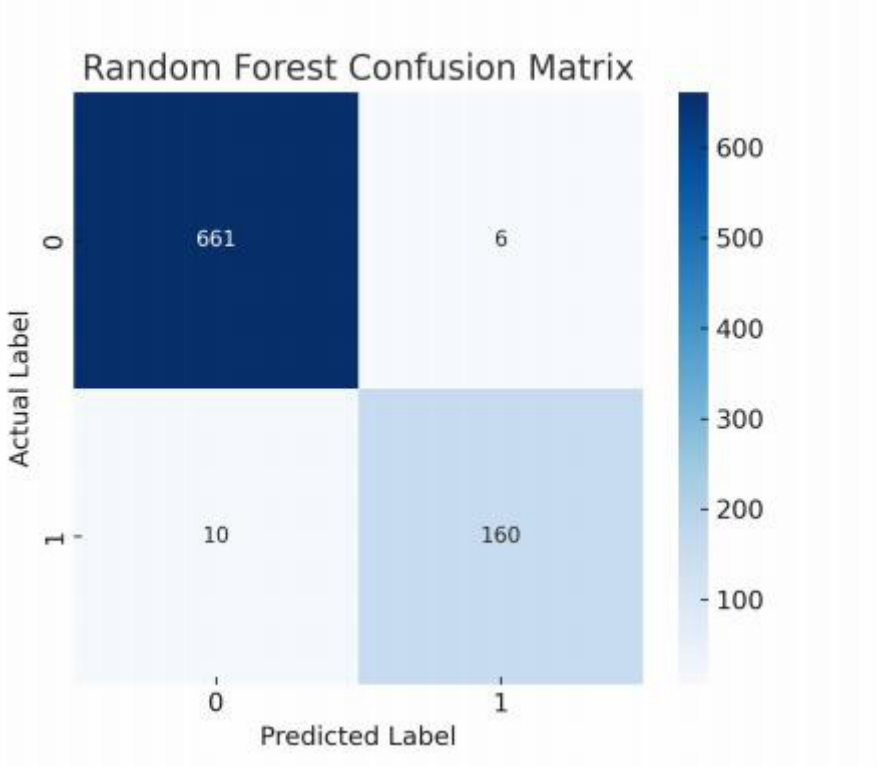
3. Model Evaluation: After training, the model's performance was evaluated on the test the set. Metrics such as the Matthews Correlation Coefficient, ROC curve, and confusion matrix were computed using the predictions made on this test set.



(a)



(b)



(c)

Figure 4: a) Feature importance in Random Forest Classifier; b) ROC curve for Random Forest model; 3) Random Forest Confusion Matrix

Figure 4 includes a detailed visualization of feature importance, the ROC curve, and the confusion matrix.

3.9 Confusion Matrix Analysis

The Confusion Matrix for our model presents a clear picture of its classification accuracy:

True Positives (TP): The majority of positive cases were correctly identified, indicating high sensitivity.

True Negatives (TN): Similarly, the model successfully recognized most of the negative cases, demonstrating high specificity.

False Positives (FP): There were very few instances where negative cases were incorrectly labeled as positive.

False Negatives (FN): The model rarely missed identifying positive cases.

The high number of TP and TN relative to the low FP and FN underscores the model's effectiveness in accurately classifying both classes, which is crucial for reliable predictions in practical applications.

3.10 Overall Performance

The Matthews Correlation Coefficient (MCC) for our Random Forest model stands at 0.94, reinforcing the model's high-quality performance across class predictions. This coefficient, along with the ROC AUC, confirms the model's strength and reliability in operational scenarios, ensuring that it can be confidently deployed for real-world applications.

4. Discussion

The findings reveal that foreshock activity intensifies as the main event approaches, both in terms of frequency and cumulative magnitude. This suggests that monitoring foreshock patterns can be crucial for earthquake prediction. Further, our analysis of foreshock density in various time windows highlights the importance of short-term monitoring data for anticipating significant seismic events.

While the model achieved high predictive accuracy (MCC = 0.94), it is limited by the regional specificity of the dataset. Future studies should test this methodology on the data from other seismic regions to ensure generalizability.

The increasing density of foreshocks observed in this study aligns with the findings of Jiao and Shan (2022) [5], underscoring the importance of short-term seismic monitoring.

5. Conclusion

We investigated the earthquake catalog of Southern California from 2021 to 2024. The research revealed that earthquakes with a magnitude greater than 3 are preceded by aftershocks of varying intensity and power. The mathematical processing of these aftershocks provided us with anomalies in foreshock activities for predicting strong earthquakes, calculated using formulas (1–7). We utilized these values to create a machine learning model for forecasting events of magnitude M>3. Using machine learning, we obtained an evaluation of the predictive model on test data, with a Matthews correlation coefficient of 0.94. The validity of the mentioned algorithm can be tested for other regions.

This study contributes to the understanding of seismic activity, preceding earthquakes by identifying key patterns in foreshocks that precede major earthquakes. The information obtained from this research can form the basis for earthquake prediction models and enhance disaster preparedness.

This study demonstrates the predictive potential of foreshock patterns using machine learning. However, the approach is constrained by the quality and regional specificity of the seismic catalog. Future research should explore integrating geochemical and geophysical data to enhance predictive capabilities.

Funding

Not appilcable.

Institutional Review Board Statement

Not appilcable.

Informed Consent Statement

Not appilcable.

References

1. Kiria, T.; Chelidze, T.; Melikadze, G.; Jimsheladze, T.; Kobzev, G. Earthquake forecast by imbalance machine learning using geophysical predictors. Annals Geophys. 2023, 66, SE636.
2. Martinelli, G.; Fu, Y.; Li, Y.; Vallianatos, F. Editorial: Pre-earthquake observations and methods for earthquake forecasting and seismic hazard reduction. Front. Earth Sci. 2023, 11, 1150414.
3. Kumazawa, T.; Ogata, Y.; Toda, S. Wide-area seismicity anomalies before the 2011 Tohoku–Oki earthquake. Geophys. J. Int. 2020, 223, 1304–1312.
4. Luo, G.; Ding, F.; Ma, H.; Yang, M. Pre-quake frequency characteristics of Ms ≥7.0 earthquakes in mainland China. Front. Earth Sci. 2023, 10, 992858.
5. Jiao, Z.; Shan, X. Pre-Seismic Temporal Integrated Anomalies from Multiparametric Remote Sensing Data. Remote Sens. 2022, 14, 2343;
6. Freund, F.; Ouillon, G.; Scoville, J.; Sornette, D. Earthquake precursors in the light of peroxy defects theory: Critical review of systematic observations. Eur. Phys. J. Spec. Top. 2021, 230, 7–46.
7. Helmstetter, A.; Sornette, D. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. J. Geophys. Res: Solid Earth 2002, 107, ESE 10-1–ESE 10-21.