2026 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)
16 April – 18 April 2026, Chattogram, Bangladesh

Distinguishing Mainshocks from Foreshocks Using Spatiotemporal Labeling and Interpretable Ensemble Learning

Kazi Sakib Hasan[1], Most. Afia Anjum Borsha[2], Faiyaz Bin Zaman[3], Azra Humayra Alam Prova[4], Mahir Hasan Chowdhury[5]
[1,2,3,4,5]Epoch One - AI Research Unit; Department of CSE, BRAC University
Dhaka, Postal Code: 1212, Bangladesh
Email: [1]kazi.sakib.hasan@g.bracu.ac.bd, [2]afia.anjum.borsha@g.bracu.ac.bd,
[3]faiyaz.bin.zaman@g.bracu.ac.bd, [4]azra.humayraalam.prova@g.bracu.ac.bd, [5]mahir.hasan.chowdhury@g.bracu.ac.bd

***Abstract*—Rapidly distinguishing between mainshocks and foreshocks immediately following a seismic event is critical for mitigating public panic and optimizing disaster response in seismically active regions. However, accurate real-time classification remains challenging due to the physical similarities in waveforms and the stochastic nature of earthquake sequences. This paper proposes a robust machine learning framework to classify seismic events using a historical United States Geological Survey (USGS) dataset spanning 1930 to 2018. We introduce a novel bi-directional spatiotemporal windowing algorithm to generate ground-truth labels based on a 30-day temporal window and 50 km spatial radius. Addressing the inherent class imbalance in seismic catalogs, we evaluate four gradient boosting algorithms—XGBoost, LightGBM, CatBoost, and AdaBoost, optimized via Bayesian hyperparameter tuning (Optuna). Experimental results demonstrate that the optimized XGBoost model achieves superior performance with a weighted F1-score of 0.79 and an Area Under the Curve (AUC) of 0.874. Furthermore, SHapley Additive exPlanations (SHAP) analysis reveals that while magnitude is the primary predictor, geophysical features such as focal depth and station gap play decisive roles in distinguishing isolated mainshocks from foreshock sequences. This framework provides a viable, data-driven tool for early warning systems to assess the likelihood of subsequent larger events.**

***Index Terms*—Seismic Classification, Mainshock Detection, Gradient Boosting, Spatiotemporal Labeling, SHAP, Bayesian Optimization, Disaster Management**

I. Introduction

In densely populated and seismically active developing nations, a moderate earthquake (Magnitude $M_w > 5.0$) triggers not only physical ground shaking but also widespread psychological panic. The immediate and critical question facing both civilians and disaster response agencies is binary: Is this event the mainshock, implying the worst is over, or is it a foreshock, signaling that a catastrophic event is imminent? This distinction is vital for public safety and resource allocation. If an event is identified as a mainshock, search and rescue operations can commence immediately. Conversely, if identified as a foreshock, the priority shifts to evacuation and securing infrastructure against a larger impending tremor. The consequences of misidentification are severe. For instance, during the 2016 Kumamoto earthquakes in Japan, a magnitude 6.5 event was initially perceived as the mainshock, causing many to return to vulnerable structures before a magnitude 7.3 mainshock struck two days later, exacerbating casualties [1]. Similarly, statistical studies suggest that while only a small percentage of earthquakes are foreshocks, the devastation caused by the subsequent mainshock accounts for a disproportionate amount of seismic damage [2]. Therefore, developing a rapid, data-driven system to classify the "role" of an earthquake immediately after occurrence is a pressing necessity for disaster resilience.

Distinguishing foreshocks from mainshocks in near real-time remains one of the most challenging problems in seismology. Physically, foreshocks and mainshocks are indistinguishable in their waveforms; they originate from the same tectonic processes. Traditional approaches, such as the Epidemic Type Aftershock Sequence (ETAS) model, rely heavily on stochastic processes and historical averages, which are effective for long-term probability estimation but often lack precision for individual event classification in real-time [3].

Furthermore, existing Machine Learning approaches often focus on predicting structural fragility or assessing damage, rather than the classification of the current event's role. A significant gap exists in applying modern Gradient Boosting algorithms to tabular seismic metadata to classify events explicitly as "Mainshock" or "Non-Mainshock" (foreshock/aftershock), particularly in datasets with class imbalance.

To address these challenges, this paper proposes a robust Machine Learning framework for seismic event classification using historical data from 1930 to 2018. The specific contributions of this work are:

- **Novel Spatiotemporal Labeling:** We introduce a bi-directional windowing algorithm that dynamically labels historical events as Mainshocks, Foreshocks, or Aftershocks based on spatial radius (50 km) and temporal windows (30 days).
- **Comprehensive Ensemble Modeling:** We evaluate and compare four state-of-the-art boosting algorithms, XG-

979-8-3315-9694-1/25/$31.00 ©2025 IEEE

Boost, LightGBM, CatBoost, and AdaBoost, to identify the most effective architecture for tabular seismic data.
- **Optimization for Imbalanced Data:** We employ Bayesian optimization (Optuna) to tune hyperparameters specifically for maximizing the F1-score, addressing the inherent class imbalance in earthquake catalogs.
- **Interpretability:** We utilize SHAP (SHapley Additive exPlanations) to provide geophysical transparency, visualizing how features like Station Gap, Depth, and Magnitude drive the model's decision for individual predictions.

The remainder of this paper is organized as follows: Section II details the literature review; section III presents the methodology, including data preprocessing, labeling logic, and model architecture; Section IV presents the experimental results and performance analysis; and Section V concludes the paper with discussions on future scope.

II. Literature Review

Recent advances in earthquake-related machine learning research increasingly focus on damage assessment, fragility analysis, and aftershock prediction under mainshock–aftershock (MS–AS) seismic sequences. A broad category of studies explores structural vulnerability and damage forecasting using ground motion descriptors. For example, prior work employed Random Forest, Extra Trees, AdaBoost, and Gradient Boosting to predict structural damage potential under MS–AS sequences using real recorded ground motions [4]. Their findings demonstrated that Gradient Boosting consistently outperformed other ML models, and SHAP analysis revealed that critical variables influencing predicted damage depend strongly on the structural period. While this work contributed valuable insights into selecting intensity measures for seismic damage prediction, it primarily focused on structural response modelling rather than seismicity-based sequence prediction or mainshock occurrence forecasting.

Other studies took a deep-learning-based approach to cumulative structural damage, making use of CNNs and multi-input architectures to estimate damage indices at the story and global levels in RC frames [5]. The model incorporated a physics-informed loss function and achieved accurate damage-state classification. Similar progress has been made in predictive fragility modelling using ML-derived fragility surfaces [6], and active learning ensemble ML models for optimizing retrofitting decisions [7]. These works predominantly emphasize structure-centric modelling — assessing column collapse probability, interstory drift, residual displacement, and energy-based damage indices. While these methods excel at post-earthquake structural impact estimation, they do not aim to predict whether an incoming seismic sequence will contain a damaging mainshock.

Another research direction focuses on ground motion simulation and spectral prediction. Automated ML was employed to forecast aftershock response spectra and synthesize realistic accelerograms [8], and deep learning models such as CGAN and DNN were shown to outperform traditional ground motion prediction equations for Sa estimation [9]. These works are highly relevant for modelling aftershock hazard, but again the emphasis remains on predicting ground motion characteristics, not predictive classification of mainshock vs. non-mainshock seismic events.

More relevant to our research are studies that examine seismicity evolution and attempt to forecast strong earthquakes from cluster patterns. The NESTORE ML method was first applied to California seismicity to estimate the probability of a subsequent strong earthquake after a mainshock [10], later extended to Greek seismicity [11], and most recently to Japanese seismic datasets where additional algorithms were developed to mitigate class imbalance in seismic cluster classification [12]. These efforts treat the problem as supervised probabilistic classification of dangerous aftershock clusters, but they begin analysis after the mainshock has occurred. That is, they attempt to predict large secondary shaking events, not the first mainshock itself.

Similarly, hybrid ML methods have been developed for short-term earthquake magnitude forecasting using seismic indicators derived from Gutenberg–Richter statistics, cluster temporal gaps, and magnitude deficit [13]. These models successfully predict the magnitude of near-future events, but they do not classify whether a given event acts as a mainshock within its temporal–spatial sequence.

Some additional research contributes novel modelling perspectives beyond traditional damage metrics, such as energy-based damage indices using Gaussian Process modelling and CNN–LSTM attention architectures [14], or recurrent neural networks for structural fragility estimation using GRU-based surrogate models [15]. Yet these again operate within a structural damage evaluation scope rather than seismic event classification.

Most existing studies treat the mainshock as a given event and analyze what happens afterward — predicting aftershocks, assessing damage, or estimating structural fragility. In contrast, our research focuses on predicting whether an incoming seismic event constitutes a mainshock before its role in the sequence is known. Rather than analyzing structural mechanics or waveform spectra, we emphasize spatiotemporal clustering of seismic events and classification of mainshocks versus smaller foreshock/aftershock events. This places our work in a novel problem category, complementary to, but distinct from structural impact models and aftershock hazard forecasting.

Furthermore, the class imbalance between mainshocks and non-mainshocks, and the sparse representation of high-magnitude initiating shocks, presents a challenge largely unaddressed in prior literature. Existing methods (e.g., REPENESE for imbalance handling) do not directly solve this for mainshock classification. Our research addresses this gap, building a methodological framework for identifying precursor patterns and predicting the onset of sequence-initiation mainshocks.

III. Methodology

The proposed framework for mainshock detection consists of three primary stages: (1) Data Preprocessing and Seismic Labeling, (2) Feature Engineering and Imputation, and (3)

Gradient Boosting-based Classification with Bayesian Optimization. The pipeline is illustrated in Figure 1.

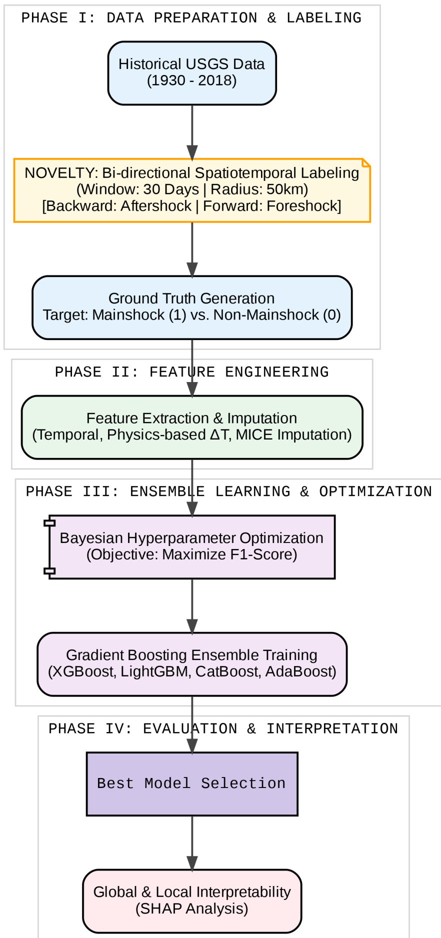

Fig. 1: Proposed methodology pipeline for Mainshock Detection, illustrating the bi-directional labeling process, feature engineering, and gradient boosting optimization framework.

A. Dataset Collection and Preprocessing

The earthquake dataset used in this study was obtained from Kaggle, based on USGS earthquake records spanning 1930 to 2018. Initially, irrelevant or incomplete columns such as `id`, `updated`, `place`, `type`, `horizontalError`, `depthError`, `magError`, `magNst`, `status`, `locationSource`, and `magSource` were removed. Magnitude values were normalized to moment magnitude scale (M_w) using standard conversion formulas for different magnitude types (`ml`, `mb`) to ensure uniformity across events [16].

$$M_w = \begin{cases} 0.85 M_L + 0.15, & \text{if type is } M_L \\ 0.85 m_b + 0.33, & \text{if type is } m_b \\ M_{\text{original}}, & \text{otherwise} \end{cases}$$

B. Spatiotemporal Sequence Labeling

A critical challenge in earthquake prediction is the lack of explicit "mainshock" labels in raw catalogs. We implemented a **Bi-directional Spatiotemporal Windowing** algorithm to generate ground-truth labels.

Let E_i be an earthquake event defined by tuple $(t_i, \text{lat}_i, \text{lon}_i, m_i)$. We define a spatiotemporal neighborhood using a time window $T = 30$ days and a spatial radius $R = 50\,\text{km}$. An event E_i is classified as a Mainshock ($y_i = 1$) if and only if it is the maximum magnitude event within its spatiotemporal neighborhood. The labeling logic proceeds in two steps:

1) **Backward Pass (Aftershock Detection):** We verify if any event E_j exists in the past window $[t_i - T, t_i)$ within distance R such that $m_j > m_i$. If such an event exists, E_i is labeled an aftershock.
2) **Forward Pass (Foreshock Detection):** We verify if any event E_k exists in the future window $(t_i, t_i + T]$ within distance R such that $m_k > m_i$. If such an event exists, E_i is labeled a foreshock.

Events that are neither foreshocks nor aftershocks (i.e., they are the local maxima of energy release) are labeled as Mainshocks.

To optimize neighbor queries on the spherical Earth surface, geodetic coordinates (latitude ϕ, longitude λ) were converted into 3D Cartesian vectors for indexing via a KD-Tree. This transformation reduces the computational complexity of distance calculations from computationally expensive Haversine formulas to efficient Euclidean metrics: $x = cos(\phi)cos(\lambda), y = cos(\phi)sin(\lambda), z = sin(\phi)$

C. Temporal Feature Engineering and Imputation

We extracted temporal features to capture cyclic seismic patterns, decomposing timestamps into month, day, day of week, hour, minute, and second. Additionally, a physics-informed feature, inter-event time ΔT, was calculated representing the time elapsed since the immediately preceding recorded seismic event globally. Real-world seismic sensor data frequently suffers from missing values. To handle missing values, we employed a Multivariate Imputation by Chained Equations (MICE) using an Iterative Imputer. It models each feature with missing values as a function of other features. We utilized a Bayesian Ridge estimator with 10 iterations to preserve the statistical distribution of the data.

D. Exploratory Data Analysis

Prior to modeling, exploratory data analysis (EDA) was conducted to understand the structure and characteristics of the earthquake dataset. Two key aspects were visualized: class distribution and feature separability. The dataset exhibits a moderate imbalance between mainshock and non-mainshock events. Figure 2 shows a bar plot of the number of mainshock (class 1) and non-mainshock (class 0) events. This imbalance informed the choice of evaluation metrics, prompting the use of the F1-score to balance precision and recall.

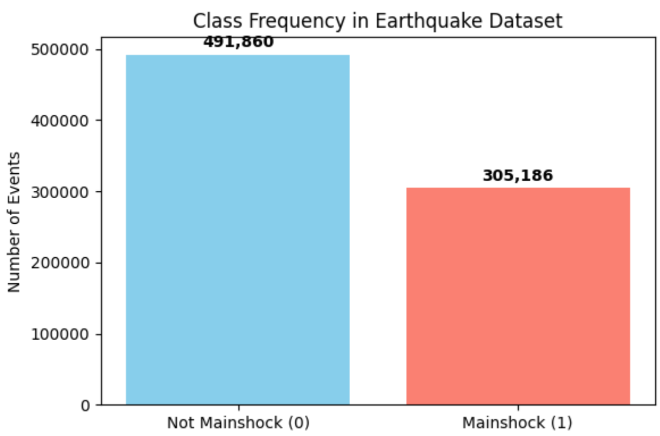

Fig. 2: Frequency Count of Mainshock and Non-Mainshock Events in the Dataset

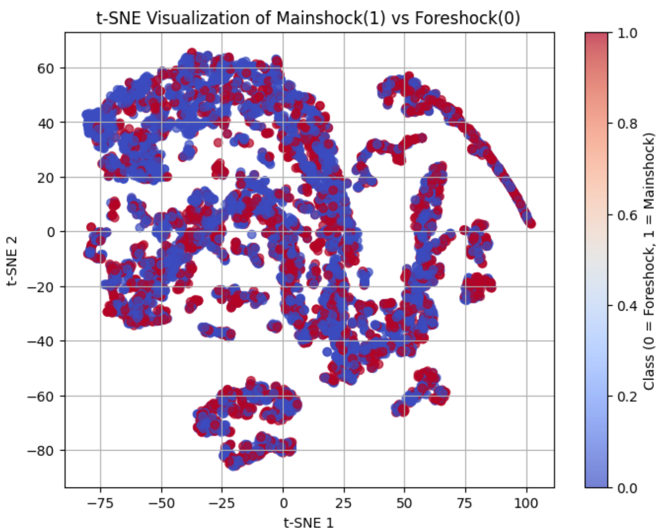

Fig. 3: Two-Dimensional t-SNE Embedding Showing Class Distribution

To assess the separability of mainshock and foreshock events in the feature space, we applied t-distributed Stochastic Neighbor Embedding (t-SNE), a non-linear dimensionality reduction technique that preserves local structure and relationships in high-dimensional data. Figure 3 shows a two-dimensional t-SNE projection of a random subset of the dataset, with points colored by class. The visualization indicates that while some separation exists, the classes are not linearly separable, highlighting the need for non-linear classifiers such as gradient boosting models.

E. Model Selection

Four ensemble-based classifiers were explored: XGBoost, LightGBM, CatBoost, and AdaBoost. Among these, the first three are gradient boosting models, which are particularly well-suited for our dataset due to its non-linear relationships, class imbalance, and partial overlap between mainshock and non-mainshock events. Gradient boosting iteratively builds an ensemble of weak learners to correct previous errors, providing strong predictive performance and robustness to complex feature interactions.

F. Hyperparameter Optimization

To maximize predictive performance, we employed Bayesian Optimization using the Optuna framework. Unlike standard grid search, Bayesian Optimization utilizes a Tree-structured Parzen Estimator (TPE) to intelligently sample the hyperparameter space, tuning critical parameters including learning rate (0.03 - 0.20), tree depth, and regularization terms. The F1-score was employed as the primary optimization metric to balance precision and recall, addressing the inherent class imbalance in the dataset. The dataset was split into training (81%), validation (9%), and test (10%) sets with stratification to preserve class distributions.

G. Training and Evaluation

Optimized models were retrained on the full training set (80%) and evaluated on the hold-out test set (20%) using standard metrics including accuracy, precision, recall, F1-score, and ROC–AUC. Confusion matrices were plotted to visualize prediction distributions for each class. Stratification is used to preserve class distributions.

H. Model Interpretability

To assess feature contributions, SHAP (SHapley Additive exPlanations) values were computed for the best performing model. Both global feature importance, measured by mean absolute SHAP values, and local explanations for individual predictions were generated. These analyses provided insights into the most influential features driving the classification of mainshocks versus non-mainshocks.

IV. Results and Discussion

A. Classification Performance

The performance of the four ensemble classifiers—XGBoost, LightGBM, CatBoost, and AdaBoost—on the test set is summarized in Table I. XGBoost achieved the highest F1-score (0.78 weighted) and accuracy (0.79), followed closely by LightGBM. AdaBoost exhibited the lowest performance among the models.

TABLE I: Test Set Classification Metrics for Mainshock Prediction

Model	Precision	Recall	F1-score (weighted)	Accuracy
XGBoost	0.82 / 0.74	0.85 / 0.71	0.79	0.79
LightGBM	0.81 / 0.73	0.84 / 0.68	0.78	0.78
CatBoost	0.80 / 0.72	0.84 / 0.65	0.77	0.77
AdaBoost	0.78 / 0.65	0.79 / 0.65	0.73	0.73

The confusion matrices shown in Figure 4 indicate that XGBoost and LightGBM better balance the trade-off between detecting mainshocks (class 1) and non-mainshocks (class 0). Notably, all models exhibited higher precision and recall for non-mainshocks, reflecting the imbalance in the dataset, where non-mainshocks were more frequent.

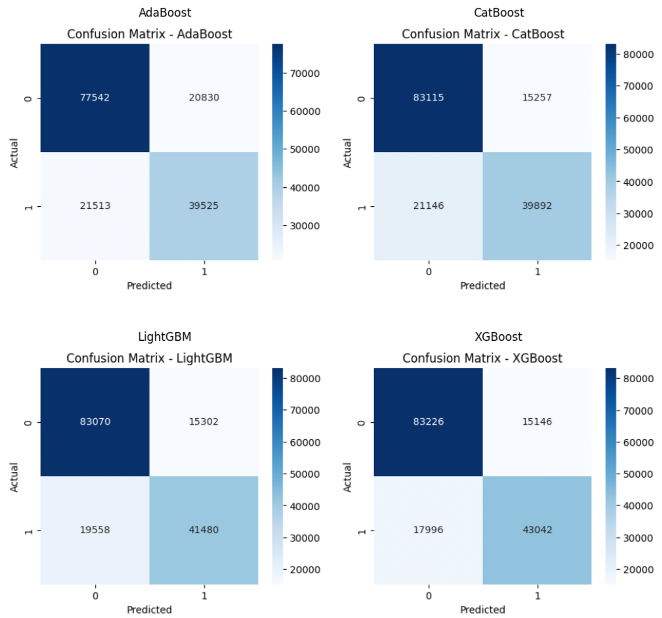

Fig. 4: Confusion Matrices of Gradient Boosting Models for Mainshock Prediction

B. ROC–AUC Performance

The area under the ROC curve (AUC) provides a threshold-independent measure of model discrimination. XGBoost achieved the highest AUC of 0.874, followed by LightGBM (0.863), CatBoost (0.851), and AdaBoost (0.718), as shown in Figure 7. These results suggest that XGBoost offers the most reliable separation between mainshocks and non-mainshocks.

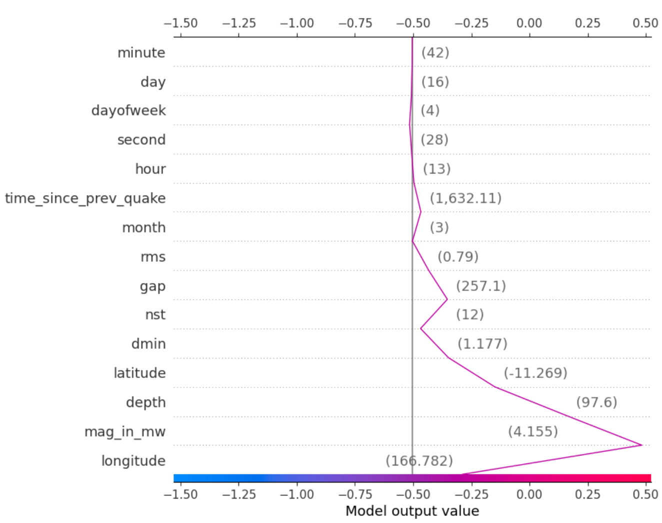

Fig. 5: SHAP Decision Plot for a Single Earthquake Prediction (XGBoost)

C. Feature Importance via SHAP

SHAP analysis for XGBoost highlights the relative importance of input features in predicting mainshocks. Figure 6 visualizes the mean absolute SHAP values. Magnitude (mag_in_mw) emerged as the most influential feature, followed by longitude and latitude. Temporal features such as day, hour, and minute had minimal, but non-negligible impact.

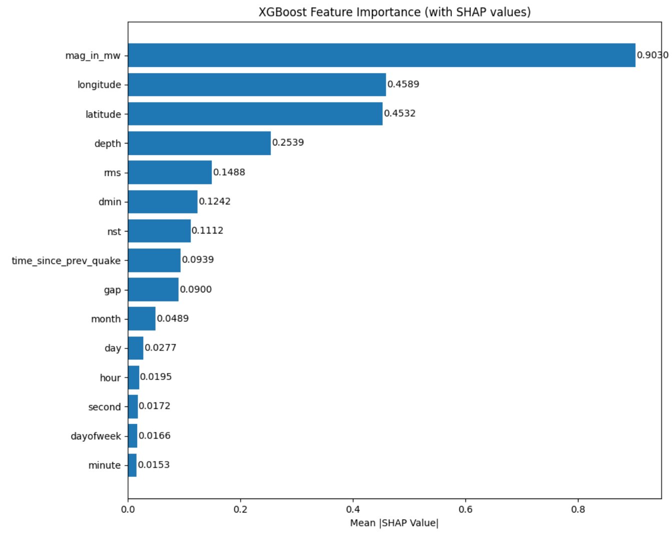

Fig. 6: Top Features by Mean Absolute SHAP Value (XGBoost)

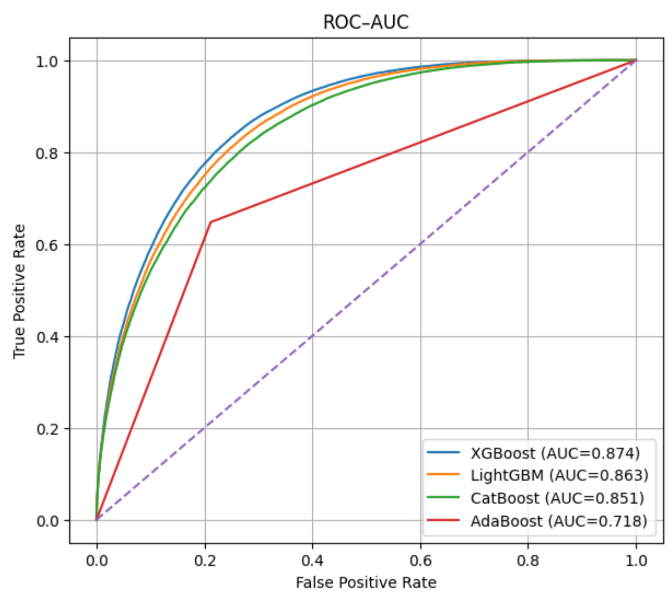

Fig. 7: ROC–AUC Comparison of Gradient Boosting Models for Mainshock Prediction

D. Interpretation of Predictions

SHAP decision plots provide detailed explanations for individual earthquake predictions. For a representative instance as shown in Figure 5, the model predicted a non-mainshock, despite a moderate magnitude (4.155), because other factors such as depth (97.6 km) and azimuthal gap (257.1°) counterbalanced the magnitude effect. Longitude contributed positively, while features like latitude, gap, and depth decreased the likelihood of the event being a mainshock. Features such as time of day or day of the week had minor influence. These observations demonstrate that the model learned complex spatio-temporal interactions that go beyond simple magnitude thresholds. Overall, ensemble models, particularly XGBoost, demonstrated strong capability in distinguishing mainshocks from foreshocks and aftershocks, with magnitude, location, and depth identified as the primary drivers of predictions. The SHAP analysis further enhanced interpretability, offering actionable insights for seismologists in earthquake-prone regions. Despite this, the moderate recall for mainshocks highlights the inherent difficulty of predicting rare, high-magnitude events.

V. Conclusion

This study presented a data-driven framework to address the critical seismological challenge of distinguishing mainshocks from foreshocks in near real-time. By leveraging a historical catalog spanning 88 years and introducing a novel Bi-directional Spatiotemporal Windowing algorithm, we successfully generated ground-truth labels to train supervised machine learning models. Our comprehensive evaluation of gradient boosting algorithms revealed that XGBoost, optimized via Bayesian hyperparameter tuning, offers the most robust performance for this classification task, achieving a superior F1-score of 0.79 and an AUC of 0.87. The model effectively handled the inherent class imbalance and missing data through iterative imputation strategies. Furthermore, the integration of SHAP analysis provided vital geophysical interpretability, identifying that while magnitude is the primary indicator, secondary factors such as focal depth and station gap play decisive roles in differentiating isolated mainshocks from complex foreshock sequences. The implications of this research are significant for disaster risk reduction in developing nations. A reliable "Mainshock vs. Non-Mainshock" classifier can drastically reduce panic and optimize emergency response strategies immediately following a seismic event. Future work could incorporate sequence-aware architectures like LSTM, additional geophysical features or real-time seismic sensor data to improve early-warning capabilities. Moreover, careful tuning of the time window for considering earthquake sequences and the spatial distance threshold for grouping related events may further enhance prediction accuracy.

References

[1] Kenta Takeda and Kazuo Inaba. The damage and reconstruction of the kumamoto earthquake: an analysis on the impact of changes in expenditures with multi-regional input–output table for kumamoto prefecture. *Economic Structures*, 11(1):20, 2022.

[2] M. G. Bonilla, R. K. Mark, and J. J. Lienkaemper. Statistical relations among earthquake magnitude, surface rupture length, and surface fault displacement. *Bulletin of the Seismological Society of America*, 74(6):2379–2411, 1984.

[3] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(40):9–27, 2012.

[4] Zhou Zhou, Meng Wang, Miao Han, Xiaohui Yu, and Dagang Lu. Prediction of damage potential in mainshock–aftershock sequences using machine learning algorithms. *Earthquake Engineering and Engineering Vibration*, 23:919–938, 2024.

[5] Maosheng Gong, Bo Liu, Xiaomin Wang, Baofeng Zhou, Yinan Zhao, and Jia Jia. Damage assessment of reinforced concrete frame under mainshock-aftershock based on deep learning considering pre-earthquake damage. *Journal of Building Engineering*, 100:111729, 2025.

[6] Mojtaba Harati and John W. van de Lindt. Mainshock-aftershock fragility surfaces: A machine learning approach using direct successive simulation. *International Journal of Disaster Risk Reduction*, 131:105886, 2025.

[7] Neda Asgarkhani, Farzin Kazemi, and Robert Jankowski. Active learning on ensemble machine-learning model to retrofit buildings under seismic mainshock-aftershock sequence. In *Computational Science – ICCS 2024*, pages 470–478. Springer, 2024.

[8] Xiaohui Yu, Meng Wang, Chaolie Ning, and Kun Ji. Predicting largest expected aftershock ground motions using automated machine learning (automl)-based scheme. *Scientific Reports*, 15:942, 2025.

[9] Yinjun Ding, Jun Chen, and Jiaxu Shen. Prediction of spectral accelerations of aftershock ground motion with deep learning method. *Soil Dynamics and Earthquake Engineering*, 150:106951, 2021.

[10] S. Gentili and R. Di Giovambattista. Forecasting strong subsequent earthquakes in california clusters by machine learning. *Physics of the Earth and Planetary Interiors*, 327:106879, 2022.

[11] Eleni-Apostolia Anyfadi, Stefania Gentili, Piero Brondi, and Filippos Vallianatos. Forecasting strong subsequent earthquakes in greece with the machine learning algorithm nestore. *Entropy*, 25(5):797, 2023.

[12] S. Gentili, G.D. Chiappetta, G. Petrillo, P. Brondi, and J. Zhuang. Forecasting strong subsequent earthquakes in japan using an improved version of nestore machine learning algorithm. *Geoscience Frontiers*, 16(3):102016, 2025.

[13] Mustafa Abdul Salam, Lobna Ibrahim, and Diaa Salama Abdelminaam. Earthquake prediction using hybrid machine learning techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(5), 2021.

[14] Yuxuan Tao, Zhao-Dong Xu, Yaxin Wei, Xin-Yu Liu, Yao-Rong Dong, and Jun Dai. Integrating deep learning into an energy framework for rapid regional damage assessment and fragility analysis under mainshock-aftershock sequences. *Earthquake Engineering Structural Dynamics*, 54(6):1678–1697, 2025.

[15] Shan He, Yuchen Liao, Peng Patrick Sun, and Ruiyang Zhang. Deep learning enabled seismic fragility evaluation of structures subjected to mainshock-aftershock earthquakes. *Urban Lifeline*, 2(2), 2024.

[16] Anbazhagan Panjamani and A. Balakumar. Seismic magnitude conversion and its effect on seismic hazard analysis. *Journal of Seismology*, 23(2):435–454, 2019.