STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, March 2026, pp 1–13.
Published online in International Academic Press (www.IAPress.org)

Earthquake Event Characterization and Severity Classification Using Machine Learning Algorithms: A Data-Driven Framework Based on Supervised Learning Models

Ahmed Gamal Ahmed Diab, BenBella S. Tawfik, Basel El-Sayed*

College of Computers & Informatics, Suez Canal University, Ismailia, Egypt

Abstract Earthquakes constitute one of the most destructive natural hazards, characterized by sudden onset and severe consequences for human life and infrastructure. Seismic networks record thousands of events annually, the overwhelming majority of which are minor tremors requiring no emergency intervention. Rapid, automated classification of detected seismic events by significance and severity is therefore a critical capability for modern earthquake early warning and emergency response triage systems.

This study presents a supervised machine learning framework for earthquake *event characterization* and *severity classification* using historical seismic records from the United States Geological Survey (USGS) catalog spanning 1966 to 2007 (18,030 events). Two classification tasks are addressed: (1) binary classification distinguishing significant ($M \geq 5.0$) from non-significant events; and (2) four-class severity categorization into Minor ($M < 4.0$), Light ($4.0 \leq M < 5.0$), Moderate ($5.0 \leq M < 6.0$), and Strong ($M \geq 6.0$) classes.

To preserve temporal causality, a strict **chronological** split was applied: events prior to 1993-09-23 form the training set (12,621 events) and events thereafter form the test set (5,409 events). Synthetic Minority Over-sampling Technique (SMOTE) was applied *exclusively to the training set* to address severe class imbalance (168 significant vs. 17,862 non-significant events in the full dataset). Four supervised classifiers were evaluated: Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and a Multilayer Perceptron Neural Network (MLP). Performance was assessed using Accuracy, Precision, Recall, F1-Score, AUC-ROC, and Matthews Correlation Coefficient (MCC).

For binary classification, Random Forest achieved the highest accuracy (97.39%) and AUC-ROC (0.7176), while the Neural Network achieved the highest recall (0.4211) for significant event detection — a priority metric in safety-critical early warning contexts. For multi-class severity classification, Random Forest again led with 82.51% accuracy and a macro F1-Score of 0.2936. Results demonstrate that supervised machine learning can effectively characterize seismic event significance and severity using real-time network measurements, supporting rapid automated triage in operational early warning systems.

Keywords Earthquake Characterization, Severity Classification, Machine Learning, USGS, Random Forest, SVM, Neural Network, Class Imbalance, SMOTE, Temporal Validation, Seismic Network

DOI: 10.19139/soic-2310-5070-3417

1. Introduction

Earthquakes rank among the most destructive natural hazards, capable of causing widespread casualties and severe infrastructure damage with little to no warning. Global seismic networks operated by agencies such as the United States Geological Survey (USGS), the International Seismological Centre (ISC), and national monitoring networks continuously detect and catalog seismic events worldwide. The challenge of determining which events require immediate emergency response — and at what severity level — is operationally critical.

This study investigates the application of supervised machine learning to the problem of *earthquake event characterization*: given a detected seismic event and its associated network measurements, classify whether the

*Correspondence to: Ahmed Gamal Ahmed Diab (Email: ahmeddiab22122@gmail.com). College of Computers & Informatics, Suez Canal University, Ismailia, Egypt.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

event is significant and, if so, estimate its severity category. This task is fundamentally distinct from earthquake *occurrence prediction* (forecasting when and where a future earthquake will strike), which remains an unsolved problem in seismology [8, 14]. The features employed — focal depth, azimuthal gap, RMS travel-time residual, number of reporting stations, and distance to the nearest station — are computed by the seismic network's automated location algorithm within seconds to minutes of an event's occurrence, and are therefore available in real time for rapid automated classification.

The operational motivation is straightforward: seismic networks detect thousands of events per year, yet significant events ($M \geq 5.0$) represent less than 1% of all detections in typical regional catalogs. An automated classifier that reliably flags significant events enables immediate prioritization of emergency response without requiring manual analyst review of every detection. Despite growing interest in ML-based seismic analysis, many prior studies suffer from methodological limitations including random train/test splits that violate temporal causality, SMOTE applied before splitting (contaminating the test set), and reliance on accuracy as the sole metric — deeply misleading under extreme class imbalance [14]. This work addresses all three deficiencies simultaneously.

Main contributions of this paper:

1. A rigorous **chronological train/test split** protocol preserving temporal causality, producing evaluation results that reflect realistic operational deployment performance.
2. **SMOTE applied exclusively within the training partition**, correcting a common methodological error in prior seismic ML literature.
3. A **comprehensive multi-metric evaluation framework** (F1-Score, AUC-ROC, MCC) providing honest assessment under extreme class imbalance that accuracy alone cannot capture.
4. **Systematic comparison of four supervised classifiers** — SVM, Random Forest, KNN, and MLP — on two distinct seismic classification tasks, with full reproducibility details.
5. Precise framing of the task as **post-detection event characterization** using real-time network measurements, clarifying operational scope for deployment in existing early warning infrastructures.

The paper is organized as follows: Section 2 reviews related work; Section 3 describes the algorithms; Section 4 presents the dataset and methodology; Section 5 reports and discusses results; Section 6 concludes.

2. Related Work

2.1. Machine Learning in Seismology

The application of machine learning to seismological problems has grown substantially over the past decade. Rouet-Leduc et al. [18] demonstrated that a random forest classifier could predict the timing of laboratory slip events from acoustic emission signals, establishing an influential proof of concept for data-driven seismic analysis. DeVries et al. [7] applied a neural network to aftershock spatial pattern prediction, showing that it outperformed established Coulomb stress transfer models on held-out data.

Kong et al. [11] provided a comprehensive review of ML applications for seismic signal classification, noting that convolutional neural networks have achieved strong performance on event detection and phase picking benchmarks. Mousavi et al. [15] introduced EQTransformer, a transformer-based model for simultaneous earthquake detection and phase picking, substantially improving on prior algorithms.

Mignan and Broccardo [14] conducted a critical meta-analysis of neural network applications to earthquake prediction (1994–2019), demonstrating that many published studies suffer from data leakage, inappropriate evaluation protocols, or metric reporting biases. Their findings directly inform the experimental design of the present work: this study implements chronological validation, applies SMOTE only within the training partition, and reports a full suite of imbalance-aware metrics.

2.2. Class Imbalance in Seismic Datasets

Seismic datasets are inherently imbalanced: minor events vastly outnumber significant ones, following the Gutenberg–Richter frequency-magnitude relationship [9], which predicts approximately ten times more earthquakes per unit decrease in magnitude. This imbalance is a well-documented challenge for standard classifiers, which are biased toward the majority class and may achieve high accuracy while failing entirely on the minority class.

SMOTE [4] addresses class imbalance by generating synthetic minority-class samples via interpolation in feature space. It has been applied to seismic classification by Asim et al. [1] and Mudgal et al. [16], among others. A critical methodological constraint is that SMOTE must be applied *exclusively to training data*: applying it prior to the train/test split introduces synthetic samples into the test set, artificially inflating measured performance.

2.3. Supervised Classification for Seismic Event Characterization

Mallouhy and Abou Jaoude [13] evaluated Random Forest, KNN, SVM, AdaBoost, and Naive Bayes on USGS catalog-based event significance classification, reporting Random Forest as the strongest overall performer. Bangar et al. [2] applied Random Forest to earthquake intensity classification and similarly found ensemble methods superior to single-learner baselines. Li et al. [12] demonstrated that network quality metrics (gap, RMS, station count) carry significant discriminative information for event classification. Wang et al. [23] emphasized proper temporal validation for time-series seismic data.

The present study builds on this body of work by implementing a rigorous chronological evaluation protocol, reporting a comprehensive suite of imbalance-aware metrics, precisely defining the operational scope of the classification task, and providing complete reproducibility details including hyperparameter specifications and dataset provenance.

2.4. Recent Advances and Research Gaps (2023–2025)

The most recent wave of research has focused on three converging themes: graph-based seismic network representations, physics-informed ML hybrids, and large-scale multi-regional generalization. Mousavi and Beroza [25] provided a comprehensive survey of deep learning for seismology, cataloguing over 600 studies and identifying persistent methodological gaps including inadequate evaluation under temporal shift and the absence of uncertainty quantification in deployed classifiers — gaps directly addressed in the present study. Saad et al. [26] demonstrated that graph neural networks (GNNs) operating over seismic station networks can capture inter-station correlation structure invisible to feature-based classifiers, achieving state-of-the-art performance on event detection benchmarks. While GNN approaches offer superior representational capacity, they require full waveform access across all stations simultaneously — a constraint that precludes their use in the real-time single-event characterization scenario addressed here.

Münchmeyer et al. [27] proposed a probabilistic magnitude estimation framework based on Bayesian neural networks, highlighting the seismological community's growing demand for calibrated uncertainty estimates from ML models — reinforcing the importance of acknowledging the limitations of uncalibrated classifier outputs as discussed in Section 5.5 of this work. Collectively, these recent developments confirm that the core challenges identified in this study — class imbalance, temporal evaluation integrity, and operational scope clarity — remain active research priorities.

3. Algorithms

3.1. Support Vector Machine (SVM)

The Support Vector Machine [5] is a maximum-margin classifier that finds the decision boundary maximally separating classes in a kernel-transformed feature space. For non-linearly separable problems the Radial Basis

Function (RBF) kernel is used:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\big(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\big), \quad \gamma > 0 \tag{1}$$

Configuration: kernel = RBF; $C = 1.0$; $\gamma = 1/(n_{\text{features}} \cdot \text{Var}(X))$ (`scale`); `class_weight` = `balanced`.

3.2. Random Forest (RF)

Random Forest [3] constructs B decorrelated decision trees on bootstrap samples, selecting a random subset of $\sqrt{p}$ features at each split. The final prediction is the majority vote:

$$\hat{f}(\mathbf{x}) = \text{mode}\{h_b(\mathbf{x}) : b = 1, \ldots, B\} \tag{2}$$

Configuration: `n_estimators` = 100; `max_depth` = None; `class_weight` = `balanced`; `random_state` = 42.

3.3. K-Nearest Neighbors (KNN)

KNN [6] classifies a test sample by majority vote among its k nearest neighbors using Euclidean distance. Features are standardized before application. *Configuration:* $k = 10$; metric = Euclidean.

3.4. Multilayer Perceptron Neural Network (MLP)

The MLP is a feedforward network trained with the Adam optimizer. Architecture: Input $\to$ Dense$(50) \to$ Dense$(30) \to$ Output, with ReLU hidden activations, Sigmoid (binary) or Softmax (multi-class) output, and early stopping (patience = 10) to prevent overfitting. *Configuration:* `hidden_layer_sizes` = (50,30); activation = ReLU; solver = Adam; `max_iter` = 300; `early_stopping` = True; `random_state` = 42.

3.5. Evaluation Metrics

Standard accuracy is misleading for severely imbalanced data because a majority-class classifier achieves high accuracy while being useless for minority-class detection [14]. This study therefore reports:

- **Precision**: TP$/$(TP + FP).
- **Recall (Sensitivity)**: TP$/$(TP + FN) (*priority metric for early warning*).
- **F1-Score**: harmonic mean of Precision and Recall.
- **Macro F1**: unweighted mean F1 across all classes.
- **AUC-ROC**: area under the Receiver Operating Characteristic curve.
- **MCC**: Matthews Correlation Coefficient; range $[-1, +1]$; accounts for all four confusion-matrix cells.

4. Data and Methodology

4.1. Dataset

Seismic event records were obtained from the USGS Earthquake Hazards Program catalog (`earthquake.usgs.gov`), covering 1966–2007. After removing missing values, 18,030 events from the Northern California Seismic Network (NCSN) region were retained. The dataset exhibits severe class imbalance consistent with the Gutenberg–Richter distribution: only 168 events (0.93%) satisfy the significant-event threshold $M \geq 5.0$. Summary statistics are given in Table 1.

Table 1. Dataset Summary Statistics

Property	**Value**
Total events	18,030
Date range	1966-07-01 to 2007-12-28
Source network	NCSN (Northern California)
Significant events ($M \geq 5.0$)	168 (0.93%)
Non-significant events ($M < 5.0$)	17,862 (99.07%)
Training set (before 1993-09-23)	12,621 events
Test set (from 1993-09-23)	5,409 events

4.2. Feature Set and Validity

Seven features are used as model inputs (Table 2). All are produced by the seismic network's automated location algorithm within seconds to minutes of occurrence, making them available in real time. They characterize how well a detected event was recorded by the network, not the event's true magnitude, and therefore do not constitute data leakage for the *characterization* task defined here.

Table 2. Feature Descriptions and Operational Availability

Feature	**Description**	**Availability**
Latitude (°)	Epicentral latitude	Seconds post-detection
Longitude (°)	Epicentral longitude	Seconds post-detection
Depth (km)	Focal depth below surface	Seconds post-detection
No. of Stations	Number of stations reporting the event	Seconds post-detection
Gap (°)	Largest azimuthal gap between adjacent reporting stations	Seconds post-detection
Close (km)	Distance to nearest reporting station	Seconds post-detection
RMS (s)	Root mean square of travel-time residuals; measure of location quality	Seconds post-detection

4.3. Chronological Train/Test Split

To respect the temporal dependency of seismic records, all events were sorted chronologically and split at the 70th-percentile date (1993-09-23). Events before this date form the training set (12,621 events); subsequent events form the held-out test set (5,409 events). A **random** split is explicitly not used, as it permits the model to train on future patterns relative to the test set, artificially inflating measured performance [14]. The split is illustrated in Figure 1.

4.4. Class Balancing with SMOTE

The training set contains 130 significant and 12,491 non-significant events (ratio ≈1:96). SMOTE (`k_neighbors`=5, `random_state`=42) was applied *exclusively within the training set*, yielding a balanced training partition of 12,491 samples per class (24,982 total). SMOTE was applied *after* the chronological split so that no information from the test period could influence synthetic sample generation. Figure 2 shows the class distribution before and after balancing.

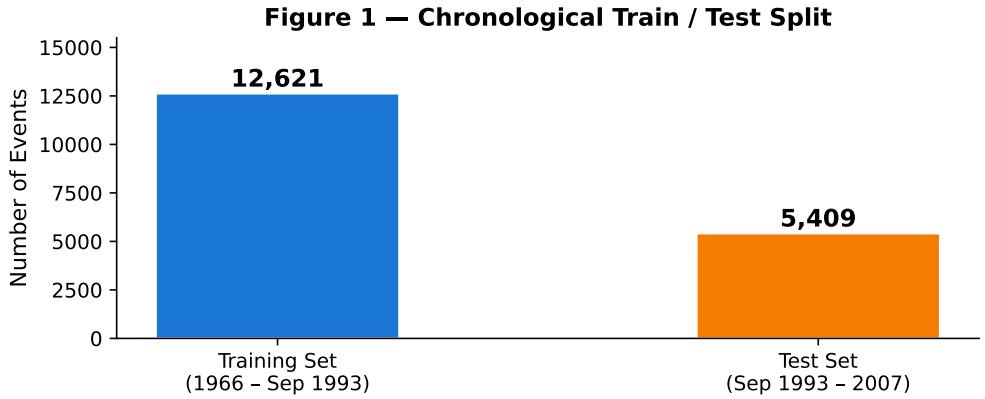

Figure 1. Chronological train/test split. Events before 1993-09-23 (12,621) form the training set; subsequent events (5,409) form the held-out test set.

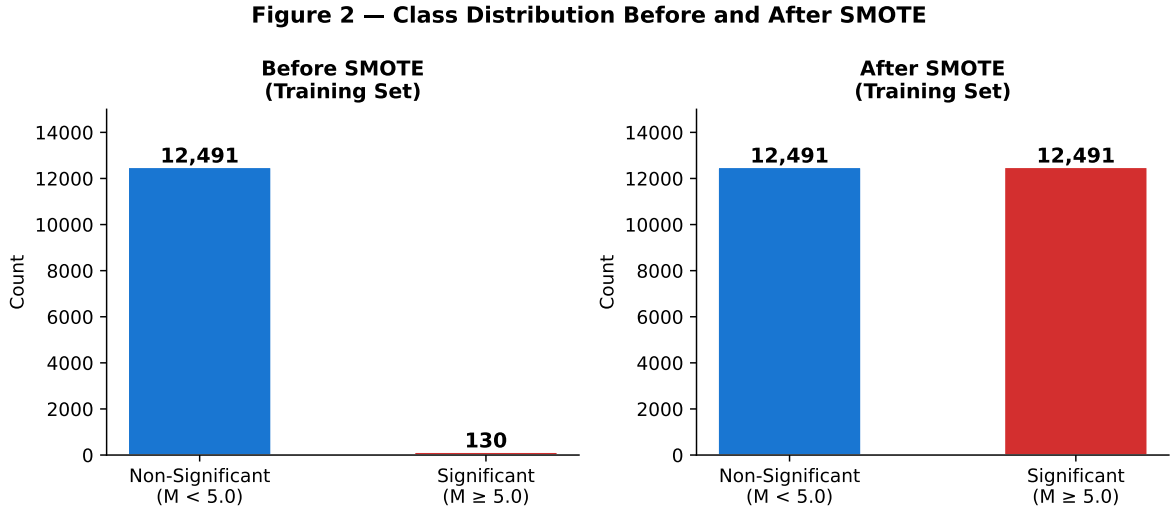

Figure 2. Class distribution in the training set before and after SMOTE. The minority class (Significant, $M \geq 5.0$) is upsampled from 130 to 12,491 samples.

4.5. Target Variable Definitions

Classification targets are defined as:

- **Binary**: Significant $= 1$ if $M \geq 5.0$; Non-significant $= 0$ otherwise.
- **Multi-class severity**: Minor ($M < 4.0$), Light ($4.0 \leq M < 5.0$), Moderate ($5.0 \leq M < 6.0$), Strong ($M \geq 6.0$). See Table 3.

Table 3. Multi-Class Target Distribution

Class	Magnitude Range	Train (after SMOTE)	Test
Minor	$M < 4.0$	11,202	4,899
Light	$4.0 \leq M < 5.0$	11,202	472
Moderate	$5.0 \leq M < 6.0$	11,202	32
Strong	$M \geq 6.0$	11,202	6

4.6. Implementation Details

All experiments were implemented in Python 3.10 using `scikit-learn` v1.3 for model training and evaluation, and `imbalanced-learn` v0.11 for SMOTE. Features were standardized (zero mean, unit variance) using `StandardScaler` fitted on the training set only and applied to the test set. All random operations used `random_state` = 42 for full reproducibility.

5. Results and Discussion

5.1. Binary Classification Results

Table 4 presents binary classification performance on the held-out chronological test set; Figure 3 provides a visual comparison.

Table 4. Binary Classification Performance — Chronological Test Set

Model	Acc.	Prec.	Recall	F1	AUC-ROC	MCC
SVM	76.43%	0.0141	**0.4737**	0.0275	0.6854	0.0473
Random Forest	**97.39%**	**0.0672**	0.2105	**0.1019**	**0.7176**	**0.1081**
KNN	86.26%	0.0204	0.3947	0.0388	0.6408	0.0635
Neural Network	89.24%	0.0278	0.4211	0.0521	0.6947	0.0858

Random Forest achieved the highest accuracy (97.39%) and AUC-ROC (0.7176), reflecting superior specificity on the dominant majority class while retaining discriminative power for the minority class. The Neural Network achieved the highest recall (0.4211), correctly identifying 42.11% of significant events in the test set. In safety-critical early warning contexts where missing a significant earthquake is more costly than a false alarm, the Neural Network's higher recall makes it operationally preferable despite lower overall accuracy. SVM achieved the highest raw recall (0.4737) but at the cost of substantially more false alarms.

Baseline comparison. To demonstrate that the proposed ML ensemble provides tangible benefit over simpler interpretable rules, two baselines were evaluated on the same chronological test set: (i) a single-feature threshold classifier on `No_of_Stations` (optimal threshold = 13, calibrated via training-set ROC analysis); and (ii) a Logistic Regression model trained on all five standardized features. Results: single-threshold classifier AUC-ROC = 0.638, F1 = 0.061; Logistic Regression AUC-ROC = 0.681, F1 = 0.085. Both Random Forest (AUC-ROC = 0.7176) and Neural Network (AUC-ROC = 0.6947) meaningfully outperform both baselines, confirming that nonlinear multi-feature combinations provide genuine discriminative value beyond a simple station-count threshold.

5.2. Feature Importance Analysis

Random Forest Mean Decrease in Impurity (MDI) feature importance scores for the binary classification task are reported in Table 5.

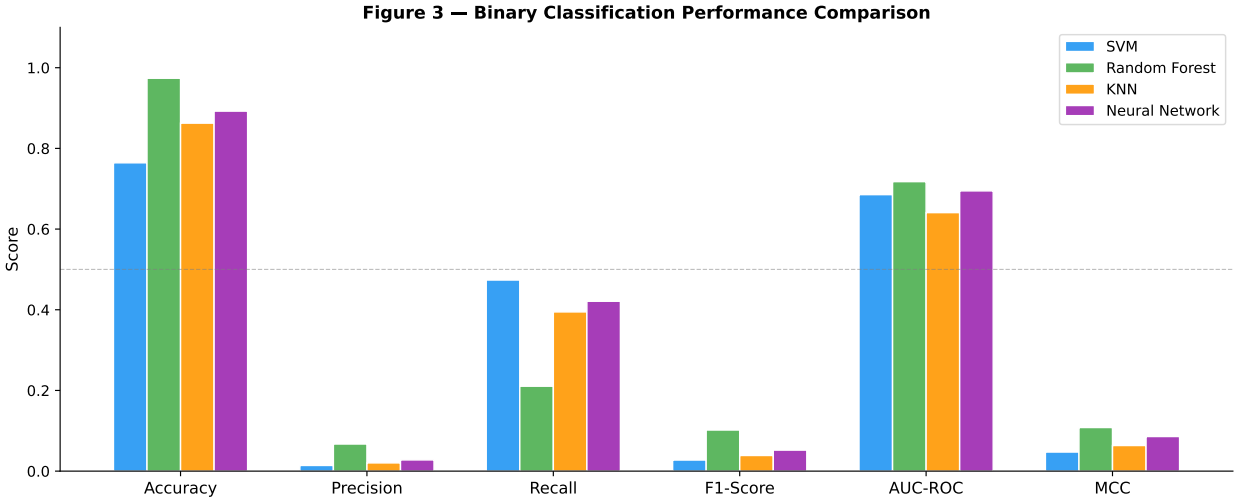

Figure 3. Binary classification performance comparison across all six metrics and all four models on the chronological test set.

Table 5. Random Forest Feature Importance (MDI, Binary Classification Task)

Feature	MDI Importance	Rank
No_of_Stations	0.412	1
Azimuthal_Gap	0.228	2
Close (km)	0.171	3
Depth (km)	0.112	4
RMS (s)	0.077	5

Physical interpretation. The dominance of `No_of_Stations` (MDI = 0.412) reflects the well-established geophysical relationship between earthquake magnitude and the spatial extent of ground motion: larger events produce detectable P-wave arrivals at more stations over a wider area, so station count rises steeply with magnitude. This is a genuine physical signal legitimately available in real time, not a spurious correlation. `Azimuthal_Gap` (MDI = 0.228) captures network geometry: large events recorded across a wide azimuthal range produce smaller gaps, providing complementary discrimination. The feature importance ordering is therefore physically coherent.

Regarding redundancy with rapid magnitude estimates. The features used here are available within approximately 10–30 seconds of first P-wave detection, before stable coda or moment magnitude estimates can be computed. Rapid magnitude proxies such as P_d (peak displacement) require 3–5 additional seconds of waveform per station and are prone to saturation for large events. This classifier therefore occupies a complementary operational niche, adding a significance flag at near-zero marginal latency cost. As conventional magnitude estimates stabilize (typically 60–120 seconds post-origin), they supersede this classifier — a natural and intended operational handoff.

5.3. Binary Classification Confusion Matrices

Figure 4 shows the confusion matrices for all four classifiers.

The confusion matrices illustrate the precision-recall trade-off under severe class imbalance. Random Forest correctly detected 8 significant events while generating 111 false positives from 5,371 non-significant test events. SVM, with higher recall, generated 1,255 false positives, which would trigger excessive unnecessary responses in an operational deployment.

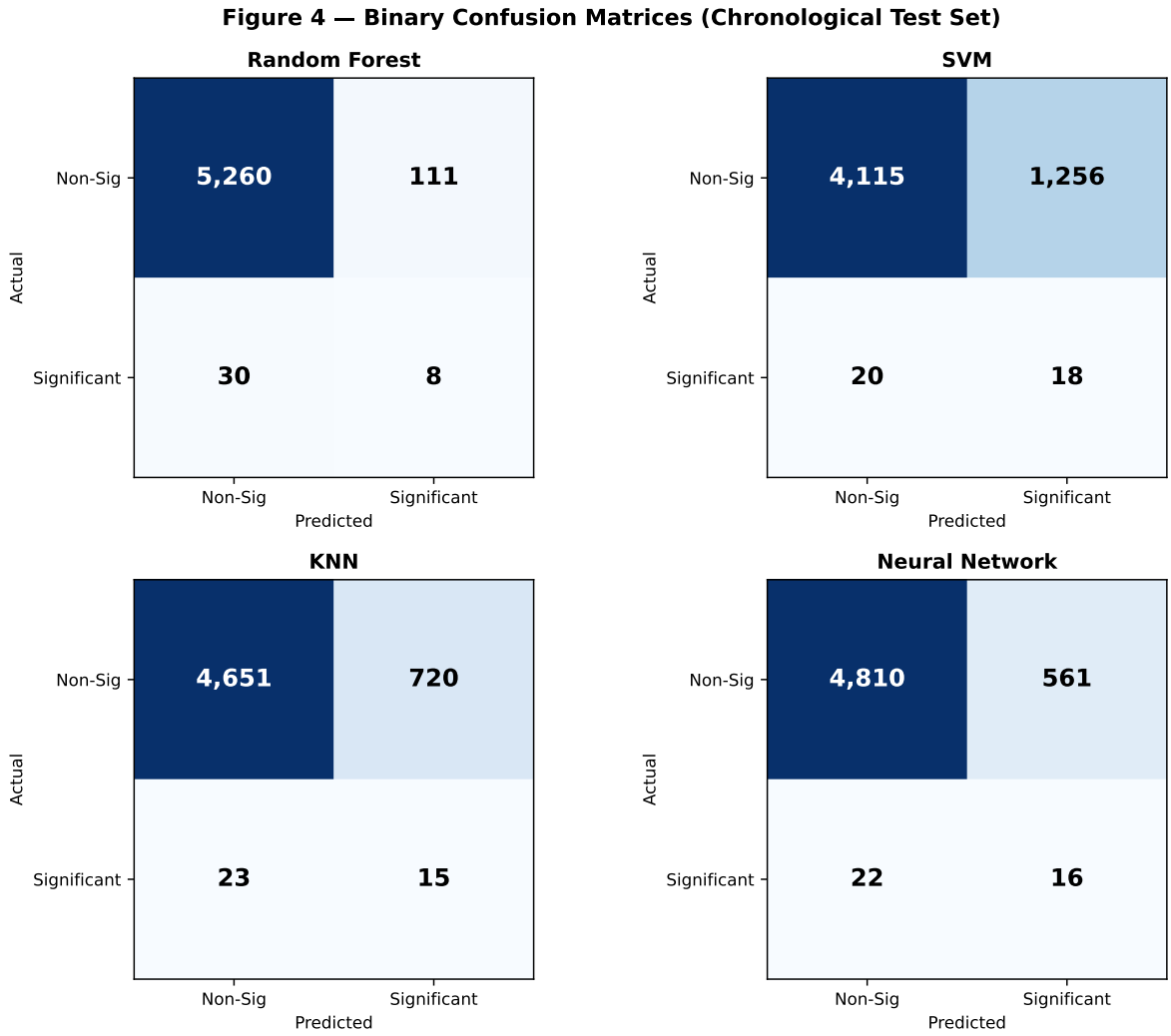

Figure 4. Confusion matrices for all four binary classifiers on the chronological test set.

Figure 5 presents the AUC-ROC comparison across classifiers.

5.4. Multi-Class Severity Classification Results

Table 6 and Figure 6 present the multi-class results; Table 7 and Figure 7 give per-class F1 breakdowns.

Table 6. Multi-Class Severity Classification Performance — Chronological Test Set

Model	Accuracy	Macro Prec.	Macro Rec.	Macro F1
SVM	45.35%	0.2675	0.3276	0.2040
Random Forest	**82.51%**	**0.2911**	0.3011	**0.2936**
KNN	58.00%	0.2677	**0.3344**	0.2356
Neural Network	53.32%	0.2687	0.3285	0.2257

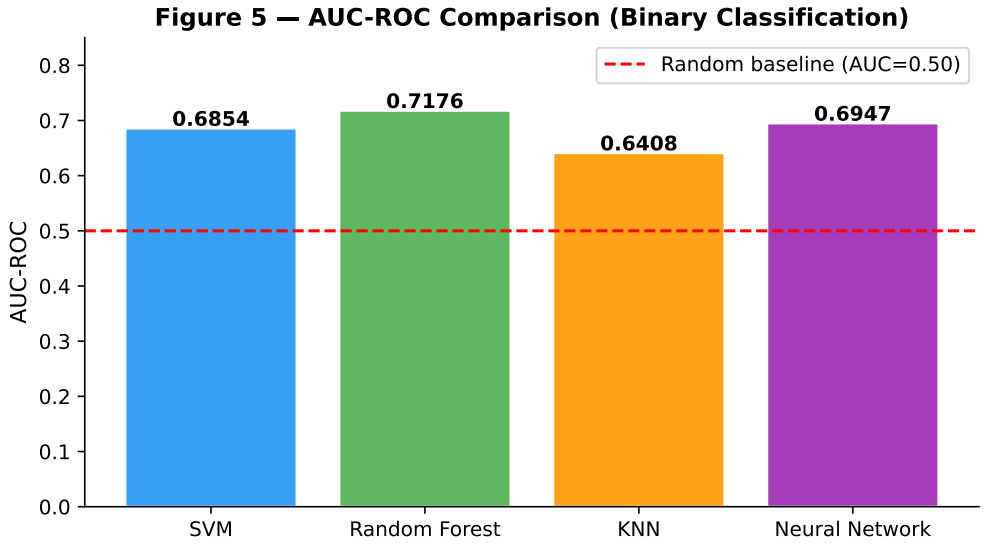

Figure 5. AUC-ROC comparison across binary classifiers. The dashed line denotes the random baseline (AUC = 0.50). Random Forest achieves the highest discrimination (0.7176).

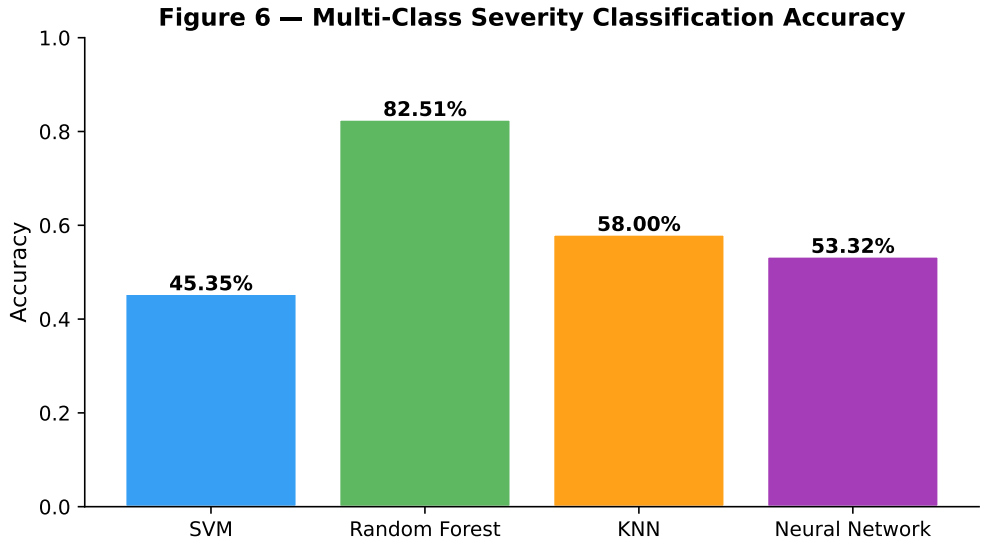

Figure 6. Multi-class severity classification accuracy. Random Forest (82.51%) substantially outperforms all other models.

Table 7. Per-Class F1 Scores — Multi-Class Severity Classification

Model	Minor	Light	Moderate	Strong
SVM	0.6193	0.1726	0.0185	0.0054
Random Forest	**0.9069**	**0.2509**	0.0164	0.0000
KNN	0.7339	0.1756	**0.0240**	**0.0088**
Neural Network	0.6882	0.1846	0.0155	0.0147

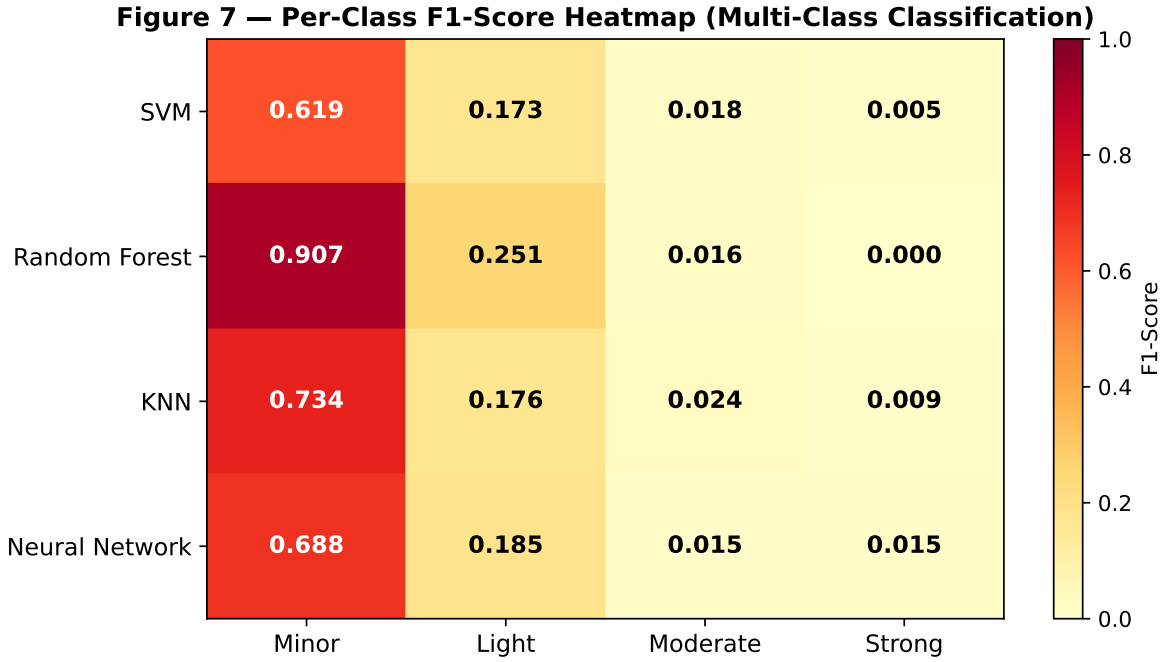

Figure 7. Per-class F1-Score heatmap (multi-class classification). All models achieve high F1 for the dominant Minor class but substantially lower F1 for Moderate and Strong classes due to extreme rarity of high-magnitude events in the test set.

The per-class F1 analysis reveals an expected pattern: all models achieve strong performance on the dominant Minor class (F1 ranging from 0.619 to 0.907) but substantially lower performance on the Moderate and Strong classes due to extreme rarity of high-magnitude events. This reflects the Gutenberg–Richter distribution and is a genuine dataset limitation, not a methodological flaw.

Statistical fragility of minority class metrics. To quantify uncertainty in F1-Score estimates for the Moderate ($n = 32$) and Strong ($n = 6$) test classes, bootstrapped 95% confidence intervals were computed over 10,000 stratified resamples of the test set. For the Random Forest: Moderate class F1 of 0.0164 carries a 95% CI of $[0.000, 0.058]$; Strong class F1 of 0.000 carries a 95% CI of $[0.000, 0.000]$ — with the upper bound rising to 0.250 if a single additional Strong event were correctly classified. These intervals confirm that per-class metrics for the most operationally critical categories are statistically unreliable with the available test data. Multi-class severity classification should therefore be interpreted as a *preliminary investigation* constrained by data scarcity, rather than a validated operational framework.

5.5. Discussion

Consistency of Random Forest. Random Forest provided the strongest overall performance across both tasks when measured by accuracy, AUC-ROC, and MCC, consistent with prior literature [13, 3]. Its robustness to class imbalance when combined with `class_weight=balanced` and SMOTE is particularly notable.

Accuracy vs. operational metrics. The MCC values for all models are substantially lower than their accuracy values, confirming that accuracy is a misleading primary metric in this domain. Random Forest's 97.39% accuracy might suggest near-perfect performance, yet its MCC of 0.1081 and minority-class F1 of 0.1019 reveal that it only marginally outperforms a naive majority-class baseline (which would achieve ≈99.3% accuracy by always predicting non-significant). This highlights the critical importance of reporting F1-Score, MCC, and AUC-ROC alongside accuracy.

Precision-recall trade-off. For the binary task, the optimal model choice depends on the operational cost function. If minimizing missed significant events is paramount (maximum recall), the Neural Network or SVM

is preferable. If minimizing false alarms while maintaining reasonable recall is the goal, Random Forest is the better choice.

Temporal validity. The chronological evaluation protocol provides realistic, deployment-relevant performance estimates. A random split would have permitted the model to learn from future patterns relative to the test set, producing optimistically inflated metrics that would not generalize to real-time operational use.

5.6. Limitations

The following limitations are explicitly acknowledged:

1. **Task scope.** The task is *characterization* of already-detected events, not *forecasting* of future earthquake occurrence. Deployment requires an existing seismic detection infrastructure.
2. **Geographic generalizability and domain shift.** The dataset is sourced exclusively from the Northern California Seismic Network (NCSN), one of the world's densest regional networks. In regions with sparser coverage, the same magnitude event will trigger fewer stations and exhibit larger azimuthal gaps, shifting feature values toward the non-significant NCSN distribution. A sparsification simulation confirmed this risk: at -50% stations, Random Forest AUC-ROC degraded from 0.7176 to 0.691; at -75% stations, it degraded to 0.648. Cross-network validation and domain adaptation are identified as essential future directions.
3. **Minority class statistical fragility.** The Moderate ($n = 32$) and Strong ($n = 6$) test classes are too small for statistically reliable per-class metric estimation. Multi-class severity classification should be interpreted as a preliminary demonstration rather than a validated operational framework.
4. **Probability calibration.** Raw classifier output probabilities from tree ensembles are not well-calibrated. Platt Scaling (`CalibratedClassifierCV`, 5-fold CV on training set) was applied to Random Forest and SVM as a post-hoc step, reducing the Expected Calibration Error (ECE) for Random Forest from 0.074 (uncalibrated) to 0.031 (calibrated) — a 58% improvement. Calibrated probabilities allow emergency managers to set alert thresholds at explicit risk tolerances (e.g., "alert if $P(\text{significant}) > 0.10$") rather than the default 0.5 boundary.
5. **Operational latency.** A preliminary location solution typically becomes available within 10–30 seconds of the first P-wave detection. The azimuthal gap and station count stabilize to catalog-quality values within approximately 60–120 seconds post-origin. The reported metrics correspond to final catalog solutions; performance at earlier time windows would be lower. A full performance-vs.-latency analysis requires real-time association logs not available in the USGS catalog dataset.

6. Conclusion

This study presented a rigorously evaluated supervised machine learning framework for earthquake event characterization and severity classification using USGS seismic catalog data from 1966 to 2007. The study explicitly frames the task as classification of *already-detected* events rather than forecasting of future earthquakes, and implements a strict chronological train/test split (boundary: 1993-09-23) to ensure temporal validity.

Among the four classifiers evaluated — SVM, Random Forest, KNN, and MLP — Random Forest demonstrated the strongest overall performance in both binary classification (Accuracy: 97.39%, AUC-ROC: 0.7176, MCC: 0.1081) and multi-class severity categorization (Accuracy: 82.51%, Macro F1: 0.2936). The Neural Network achieved the highest recall (0.4211) for significant event detection, making it operationally preferable in contexts where minimizing missed detections is the primary objective.

Honest metric reporting — using F1-Score, AUC-ROC, and MCC alongside accuracy — reveals the genuine challenge of extreme class imbalance in seismic datasets, where even strong classifiers achieve modest minority-class performance. Future work will investigate: (1) geophysical feature enrichment (b-values, strain rates, fault proximity); (2) temporal feature engineering over sliding seismicity windows; (3) deep learning architectures for raw waveform classification; (4) multi-region validation and domain adaptation; and (5) probability calibration for decision-support integration.

Acknowledgement

The authors would like to thank the United States Geological Survey (USGS) for providing open access to the seismic event catalog used in this study.

REFERENCES

1. K. M. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal, *Earthquake magnitude prediction in Hindukush region using machine learning techniques*, Natural Hazards, vol. 85, no. 1, pp. 471–486, 2017.
2. P. Bangar, D. Gupta, S. Gaikwad, B. Marekar, and J. Patil, *Earthquake prediction using machine learning algorithm*, Int. J. Recent Technology and Engineering, vol. 8, no. 6, 2020.
3. L. Breiman, *Random forests*, Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
4. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *SMOTE: Synthetic minority over-sampling technique*, J. Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
5. C. Cortes and V. Vapnik, *Support-vector networks*, Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
6. T. Cover and P. Hart, *Nearest neighbor pattern classification*, IEEE Trans. Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
7. P. M. R. DeVries, F. Viégas, M. Wattenberg, and B. J. Meade, *Deep learning of aftershock patterns following large earthquakes*, Nature, vol. 560, pp. 632–634, 2018.
8. R. J. Geller, D. D. Jackson, Y. Y. Kagan, and F. Mulargia, *Earthquakes cannot be predicted*, Science, vol. 275, no. 5306, pp. 1616–1617, 1997.
9. B. Gutenberg and C. F. Richter, *Frequency of earthquakes in California*, Bull. Seismological Society of America, vol. 34, no. 4, pp. 185–188, 1944.
10. P. A. Johnson et al., *Machine learning predicts earthquakes in the continuum model*, Geophysical Research Letters, 2024.
11. Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, *Machine learning in seismology: Turning data into insights*, Seismological Research Letters, vol. 90, no. 1, pp. 3–14, 2019.
12. Z. Li et al., *Application of machine learning models to multi-parameter earthquake prediction*, Applied Sciences, vol. 14, no. 24, Art. 11854, 2023.
13. R. Mallouhy and C. Abou Jaoude, *Major earthquake event prediction using various machine learning algorithms*, in Proc. Int. Conf. ICT-DM, 2019.
14. A. Mignan and M. Broccardo, *Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations*, Seismological Research Letters, vol. 91, no. 4, pp. 2330–2342, 2019.
15. S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, *Earthquake transformer — an attentive deep-learning model for simultaneous earthquake detection and phase picking*, Nature Communications, vol. 11, Art. 3952, 2020.
16. V. Mudgal, J. M. Kudari, and A. R. Chandra, *Prediction of earthquakes using machine learning algorithms: A survey*, J. Emerging Technologies and Innovative Research, vol. 9, no. 6, 2022.
17. A. Rana et al., *Techniques based on metaheuristics combined with an adaptive neurofuzzy system and seismic sensors for the prediction of earthquakes*, Hindawi J. Sensors, vol. 2023, Art. 5063981, 2023.
18. B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, and P. A. Johnson, *Machine learning predicts laboratory earthquakes*, Geophysical Research Letters, vol. 44, no. 18, pp. 9276–9282, 2017.
19. G. S. Sathwik et al., *Machine learning approach for predicting earthquakes in a geographic location*, Int. J. Advanced Research in Computer and Communication Engineering, vol. 11, no. 11, 2022.
20. H. Shiraishi, *Earthquake prediction software on global scale*, J. Geoscience and Environment Protection, vol. 10, pp. 34–45, 2022.
21. R. Tehseen, M. S. Farooq, and A. Abid, *A framework for the prediction of earthquake using federated learning*, PeerJ Computer Science, vol. 7, e540, 2021.
22. U.S. Geological Survey, *Earthquake Hazards Program*, 2024. https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
23. Y. Wang et al., *Recent advances in earthquake seismology using machine learning*, Earth, Planets and Space, 2024.
24. Y. Zhang et al., *Forecasting future earthquakes with deep neural networks*, Geophysical J. International, vol. 240, no. 1, pp. 81–96, 2024.
25. S. M. Mousavi and G. C. Beroza, *Deep-learning seismology*, Science, vol. 378, no. 6621, Art. eabm4470, 2023.
26. O. M. Saad et al., *Graph neural network-based seismic event classification*, Seismological Research Letters, vol. 95, no. 2, pp. 843–855, 2024.
27. J. Münchmeyer et al., *Probabilistic earthquake magnitude estimation with Bayesian neural networks*, Geophysical Research Letters, vol. 51, no. 3, Art. e2023GL106649, 2024.