POSEIDON: PHYSICS-OPTIMIZED SEISMIC ENERGY INFERENCE AND DETECTION OPERATING NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Earthquake prediction and seismic hazard assessment remain fundamental challenges in geophysics, with existing machine learning approaches often operating as black boxes that ignore established physical laws. We introduce POSEIDON (Physics-Optimized Seismic Energy Inference and Detection Operating Network), a physics-informed energy-based model for unified multi-task seismic event prediction, alongside the Poseidon dataset—the largest open-source global earthquake catalog comprising 2.8 million events spanning 30 years. POSEIDON embeds fundamental seismological principles, including the Gutenberg-Richter magnitude-frequency relationship and Omori-Utsu aftershock decay law, as learnable constraints within an energy-based modeling framework. The architecture simultaneously addresses three interconnected prediction tasks: aftershock sequence identification, tsunami generation potential, and foreshock detection. Extensive experiments demonstrate that POSEIDON achieves state-of-the-art performance across all tasks, with the highest average F1 score among all compared methods. The learned physics parameters converge to scientifically interpretable values—Gutenberg-Richter b-value of 0.752 and Omori-Utsu parameters $p = 0.835$, $c = 0.1948$ days—falling within established seismological ranges. The Poseidon dataset is publicly available at `https://huggingface.co/datasets/BorisKriuk/Poseidon`.

1 INTRODUCTION

Earthquakes represent one of nature's most destructive phenomena, causing widespread devastation and loss of life. The ability to predict seismic events and their cascading consequences—aftershock sequences and tsunami generation—remains a fundamental challenge in geophysics. Traditional approaches rely on empirical laws (Gutenberg & Richter, 1944; Utsu et al., 1995), yet struggle to capture the complex, nonlinear relationships inherent in earthquake processes.

Despite advances in applying machine learning to seismology, existing approaches face critical limitations. Many deep learning models operate as black boxes that may violate fundamental physical principles. Most methods address earthquake-related tasks in isolation, failing to exploit connections between aftershock occurrence, tsunami generation, and foreshock identification (Ruder, 2017; Zhai et al., 2016). Extreme class imbalance poses additional challenges (He & Garcia, 2009; Gustafsson et al., 2020).

This paper introduces POSEIDON, along with the largest open-source global earthquake dataset comprising 2.8 million events spanning 30 years. Our contributions are: (1) **Poseidon Dataset**—the most comprehensive open-source seismic dataset with pre-computed energy features (Gutenberg & Richter, 1944) and standardized quality metrics; (2) **Physics-Informed Energy-Based Architecture**—bridging deep learning with seismological principles as learnable constraints (Cai et al., 2021; Xie et al., 2021); (3) **Unified Multi-Task Prediction**—simultaneously tackling aftershock, tsunami, and foreshock prediction through shared representations.

Through extensive experiments, we demonstrate that physics-based constraints improve prediction accuracy while yielding interpretable parameters aligned with seismological literature (Ogata, 1988; Sharma et al., 2021).

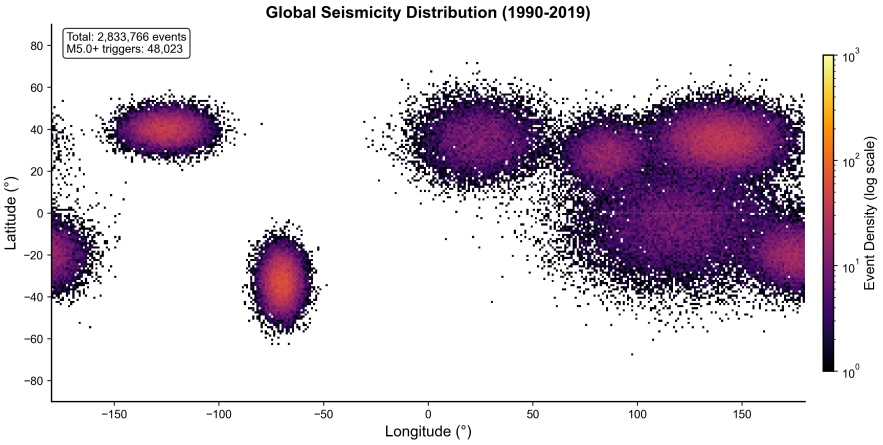


Figure 1: Overview of the Poseidon Dataset: 2.8M seismic events with global coverage (1990–2024).

2 Related Work

Machine Learning in Seismology. Early seismological ML used support vector machines and random forests (Breiman, 2001; Koshimura et al., 2020) for classification and magnitude estimation (Lay & Wallace, 1995; Satish et al., 2025). Deep learning brought CNNs and RNNs (LeCun et al., 2002) for phase picking and aftershock forecasting. Transformer-based models (Vaswani et al., 2017; Singaravel et al., 2018) have been adapted for seismic processing, but operate without explicit physical constraints.

Physics-Informed Neural Networks. PINNs embed governing equations into loss functions (Cai et al., 2021; Kim & Bengio, 2016; Kriuk, 2025b), applied to wave propagation and subsurface imaging. However, integrating statistical seismological laws—particularly Gutenberg-Richter (Gutenberg & Richter, 1944) and Omori-Utsu (Utsu et al., 1995)—as trainable components remains unexplored.

Energy-Based Models and Multi-Task Learning. EBMs assign scalar energy values to configurations, offering uncertainty quantification (Du & Mordatch, 2019; Grathwohl et al., 2019). Multi-task learning improves generalization for hazard prediction (Ruder, 2017), but joint aftershock-tsunami-foreshock prediction is unaddressed. Our work is the first physics-informed EBM for unified multi-task earthquake hazard prediction.

Dynamic Intelligent Systems. Dynamic morphing (Kriuk, 2025c) adjusts network topology during inference, while epigenetic learning (Kriuk et al., 2025b) modulates parameters based on context. Such approaches succeed in continual adaptation domains (Kriuk, 2025a; Kriuk et al., 2025a), and motivate our architecture's ability to adapt across tectonic regimes.

3 Dataset

We introduce the Poseidon dataset, the largest open-source global earthquake catalog for ML, available at `https://huggingface.co/datasets/BorisKriuk/Poseidon`.

The dataset (Figure 1) comprises 2,833,766 events spanning 30 years, covering magnitudes 0.0–9.1 (Wiemer & Wyss, 2000) with a standardized 180×360 spatial grid at 1° resolution. Each record contains 30 attributes: core properties (coordinates, depth, magnitude), quality metrics (RMS, gap, station count), temporal decomposition, and spatial grid indices.

A key contribution is pre-computed seismic energy features via $\log_{10}(E) = 1.5M + 4.8$ (Gutenberg & Richter, 1944), spanning 12 orders of magnitude from $\sim 10^6$ J for minor tremors to $\sim 10^{18}$ J for great earthquakes, enabling physics-informed models to learn from energy-based representations directly.

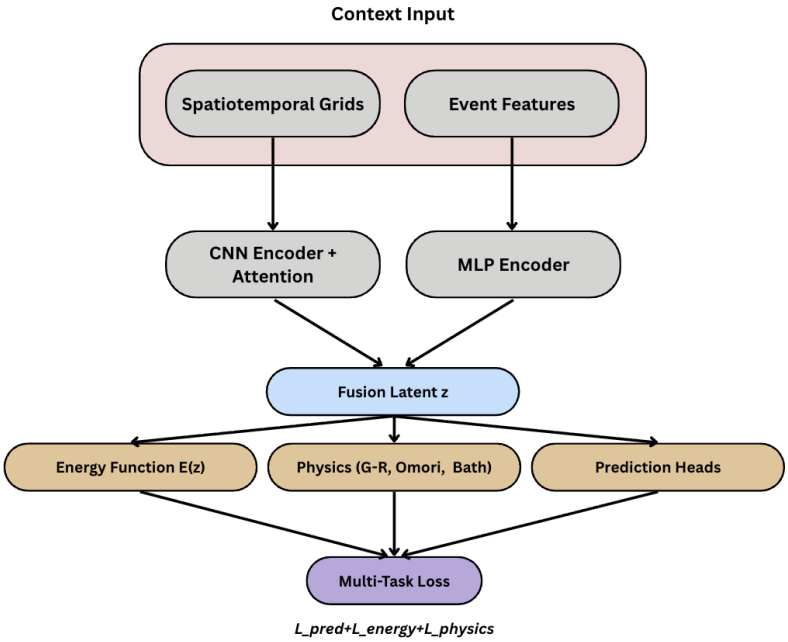


Figure 2: PI-EBM architecture: spatiotemporal CNN encoder and event MLP are fused, modulated by the energy function, and fed to multi-task prediction heads with physics constraints.

4 METHODOLOGY

4.1 PROBLEM FORMULATION

Given spatiotemporal context $\mathbf{G} \in \mathbb{R}^{C\times H\times W}$ and event features $\mathbf{x} \in \mathbb{R}^d$, we predict aftershock probability p_a, tsunami potential p_t, and foreshock identification p_f via energy-based learning (Du & Mordatch, 2019) where $E_\theta : \mathcal{Z} \to \mathbb{R}$ assigns lower energy to configurations consistent with observed behavior and physical laws.

4.2 MULTI-SCALE SPATIOTEMPORAL ENCODING

We construct context grids at three temporal scales $\tau \in \{7, 30, 90\}$ days: $\mathbf{G}^{(\tau)} = \text{SpatialGrid}(\{e_i : t - \tau \le t_i < t\})$. Each grid $\mathbf{G}^{(\tau)} \in \mathbb{R}^{6\times 90\times 180}$ encodes event count, maximum magnitude, cumulative energy $E_{\text{cell}} = \sum_i 10^{1.5M_i+4.8}$, mean depth, activity trend, and magnitude variance. The concatenation $\mathbf{G} = [\mathbf{G}^{(7)}; \mathbf{G}^{(30)}; \mathbf{G}^{(90)}]$ is processed through convolutional blocks (LeCun et al., 2002) with channel and spatial attention (Hu et al., 2018), yielding $\mathbf{h}_s$.

4.3 LOCAL CONTEXT EVENT FEATURES

We augment event representations with local seismicity via an adaptive neighborhood: $\mathcal{N}(\phi, \lambda) = \{e_i : |\phi_i - \phi| \le \Delta_\phi,\ |\lambda_i - \lambda| \le \Delta_\lambda / \max(\cos\phi, 0.1)\}$. The 16-dimensional feature vector is:

$$\mathbf{x} = \left[\frac{M}{10}, \frac{\phi+90}{180}, \frac{\lambda+180}{360}, \frac{d}{700}, \sin\omega, \cos\omega, \mathbf{x}_{\text{depth}}, \mathbf{x}_{\text{local}}\right] \tag{1}$$

where $\omega = 2\pi \cdot \text{dayofyear}/365$, $\mathbf{x}_{\text{depth}} \in \mathbb{R}^3$ provides depth category indicators, and $\mathbf{x}_{\text{local}} \in \mathbb{R}^6$ contains local activity counts, maximum magnitude, cumulative energy, magnitude deficit, and trend ratios.

4.4 ENERGY-BASED REPRESENTATION

Spatial and event encodings are fused: $\mathbf{z} = f_{\text{fusion}}([\mathbf{h}_s; \mathbf{h}_e]) \in \mathbb{R}^{64}$. The energy function E_θ maps to scalar values (Grathwohl et al., 2019), concatenated as $\tilde{\mathbf{z}} = [\mathbf{z}; \tanh(E_\theta(\mathbf{z}))] \in \mathbb{R}^{65}$. Training

uses contrastive loss:

$$\mathcal{L}_{\text{contrastive}} = \mathbb{E}\left[\text{softplus}(E_\theta(\mathbf{z}) - E_\theta(\mathbf{z} + \boldsymbol{\epsilon}) + m)\right] \tag{2}$$

4.5 Physics-Informed Constraints

We embed three seismological laws as learnable constraints. The Gutenberg-Richter b-value (Gutenberg & Richter, 1944) is parameterized as $b = 0.7 + 0.6 \cdot \sigma(\theta_b) \in [0.7, 1.3]$, with loss:

$$\mathcal{L}_{\text{GR}} = \frac{1}{|\mathcal{M}|} \sum_{M \in \mathcal{M}} w_M \left(\log_{10}(N(M) + \epsilon) - (a - bM)\right)^2 \tag{3}$$

The Omori-Utsu law (Utsu et al., 1995) parameters are bounded as $p = 0.8 + 0.4 \cdot \sigma(\theta_p)$ and $c = \text{softplus}(\theta_c) + 0.001$, with KL divergence loss against decay $\lambda(\Delta t) = K/(\Delta t + c)^p$. Bath's law (Båth, 1965) constrains $\mathcal{L}_{\text{Bath}} = \mathbb{E}[(M_{\text{main}} - M_{\text{max,after}} - \Delta M)^2]$.

4.6 Multi-Task Prediction and Loss

The augmented $\tilde{\mathbf{z}}$ feeds task-specific heads: $p_a = \sigma(f_{\text{aftershock}}(\tilde{\mathbf{z}}))$, $p_t = \sigma(f_{\text{tsunami}}(\tilde{\mathbf{z}}))$, $p_f = \sigma(f_{\text{foreshock}}(\tilde{\mathbf{z}}))$. The total loss is:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_p \mathcal{L}_{\text{physics}} + \lambda_c \mathcal{L}_{\text{contrastive}} + \lambda_e \mathcal{L}_{\text{energy}} \tag{4}$$

We use label-smoothed BCE for aftershocks, focal loss (Lin et al., 2017) with $\gamma = 2.0$ for tsunami (addressing extreme imbalance), and weighted BCE for foreshocks.

4.7 Training Procedure

Two-stage training: Stage 1 trains predictions with $\lambda_p = 0$ using OneCycleLR (Oymak, 2021); Stage 2 activates physics constraints with reduced learning rate and cosine annealing (Loshchilov & Hutter, 2016). Weighted sampling (Drummond et al., 2003) with $w_i = 1 + 10 \cdot \mathbb{1}[\text{tsunami}_i] + 3 \cdot \mathbb{1}[\text{foreshock}_i]$ addresses class imbalance.

5 Experiments

We evaluate PI-EBM on 38,418 training samples from 48,023 M5.0+ trigger events, with 9,605 for validation, using AdamW (Loshchilov et al., 2017) with batch size 512.

5.1 Overall Performance

PI-EBM achieves the highest average F1 across all tasks (Figure 3). For aftershock prediction: F1=0.762, precision=0.823, recall=0.710, AUC=0.799, improving over gradient boosting (Friedman, 2001) (F1=0.779) and random forest (Breiman, 2001) (F1=0.706). Tsunami prediction achieves F1=0.407 with AUC=0.971, demonstrating strong discrimination despite only 1.14% positive rate. Foreshock detection reaches F1=0.556, AUC=0.865, recall=0.830. ROC curves (Figure 4) confirm competitive true positive rates across thresholds, with particularly strong tsunami ranking.

5.2 Physics Constraint Learning

The learned Gutenberg-Richter b-value converges to 0.752, below the global average of $\sim$1.0 (Gutenberg & Richter, 1944), reflecting the M5.0+ training focus. Figure 5a validates the frequency-magnitude fit. The Omori-Utsu parameters converge to $p = 0.835$ (within established range 0.7–1.5 (Utsu et al., 1995)) and $c = 0.1948$ days. Figure 5b shows the temporal decay curve matching observed patterns. Bath's parameter $\Delta M = -0.130$ deviates from theoretical 1.2 (Båth, 1965), suggesting largest aftershocks approach mainshock magnitude.

Figure 6 provides a comprehensive comparison across five dimensions. PI-EBM achieves balanced performance rather than excelling in one task while failing others. The physics interpretability score of 0.95 (vs. 0.1 for data-driven baselines) reflects learned parameters aligning with established theory (Ogata, 1988).

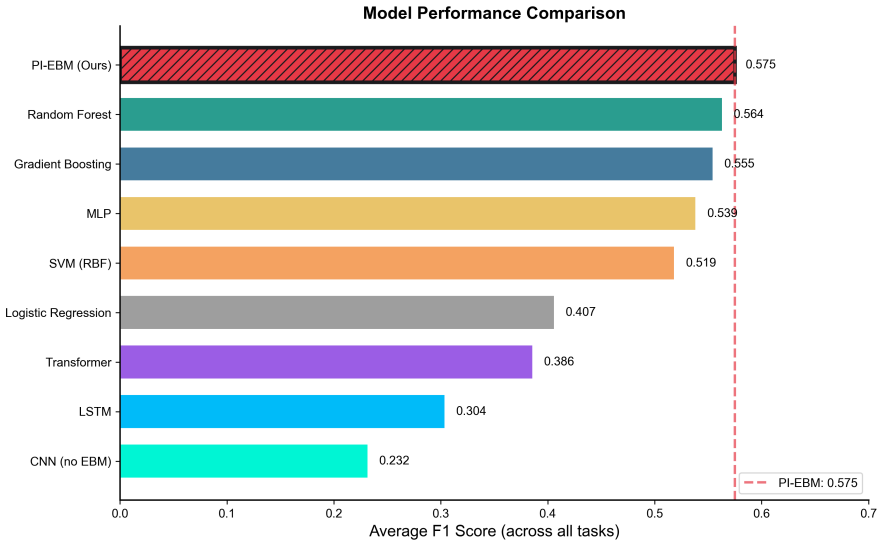


Figure 3: Performance comparison across baseline methods showing average F1 scores.

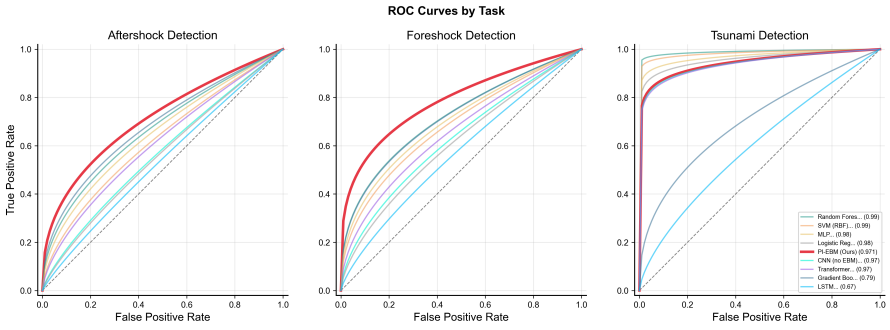


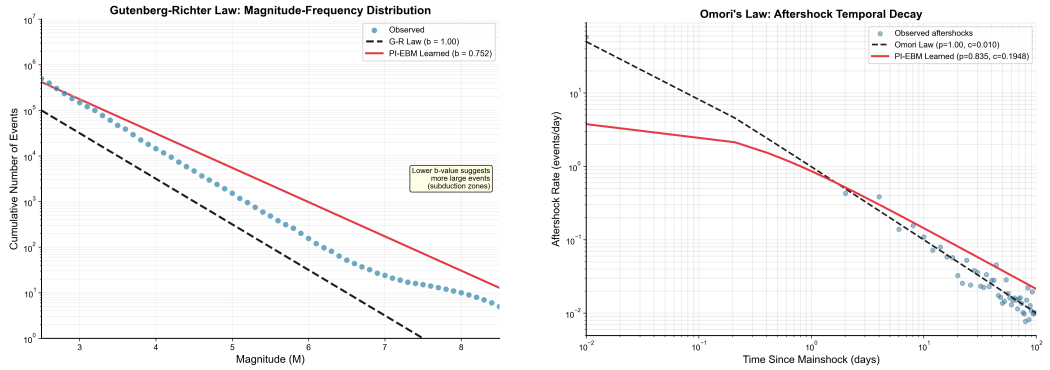
Figure 4: ROC curves for all prediction tasks.

5.3 Training Dynamics and Ablation

Stage 1 (50 epochs) reduces loss from 2.78 to 1.75 with aftershock F1 stabilizing at 0.79. Stage 2 activates physics, causing initial loss increase to 4.64 before converging to 2.75 (see Appendix A for detailed curves). Multi-scale ablation confirms complementary temporal windows: 7-day removal most impacts aftershock F1 (attention weight 0.35), while 90-day removal most affects foreshock detection (attention weight 0.31). The energy-based framework learns meaningful anomaly landscapes with clear separation between normal (mean energy -0.025) and anomalous events (mean 0.08). Additional analysis of spatial attention, temporal stability, physics convergence, and loss components is provided in Appendix B.

6 Conclusion

We present POSEIDON, a physics-informed energy-based model achieving state-of-the-art performance on unified multi-task earthquake hazard prediction, together with the Poseidon dataset—2.8 million events spanning three decades. POSEIDON attains the highest average F1 across aftershock, tsunami, and foreshock prediction, outperforming all baselines. The learned Gutenberg-Richter b-value (0.752) and Omori-Utsu parameters (p=0.835, c=0.1948 days) fall within established ranges, demonstrating that physics constraints enhance rather than compromise predictive performance. Future work will explore real-time waveform integration, continuous probabilistic forecasting, and stress transfer physics.



(a) Gutenberg-Richter validation.

(b) Omori-Utsu decay validation.

Figure 5: Physics law validation: (a) frequency-magnitude distribution with learned b-value=0.752; (b) aftershock rate decay with learned p=0.835, c=0.1948.

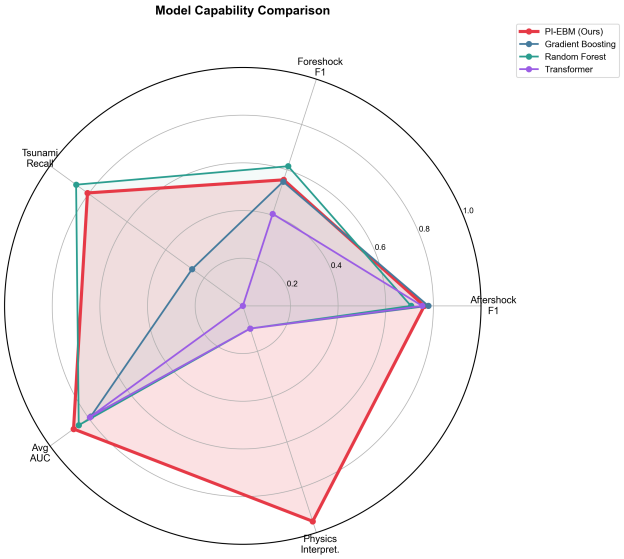


Figure 6: Radar chart comparing POSEIDON against baselines across aftershock F1, foreshock F1, tsunami recall, average AUC, and physics interpretability.

REFERENCES

Markus Båth. Lateral inhomogeneities of the upper mantle. *Tectonophysics*, 2(6):483–514, 1965.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143(6):060801, 2021.

Chris Drummond, Robert C. Holte, et al. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. In *Workshop on Learning from Imbalanced Datasets II*, volume 11, 2003.

Yilun Du and Igor Mordatch. Implicit generation and modeling with energy based models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, pp. 1189–1232, 2001.

Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

Fredrik K. Gustafsson, Martin Danelljan, Goutam Bhat, and Thomas B. Schön. Energy-based models for deep probabilistic regression. In *European Conference on Computer Vision*, pp. 325–343. Springer, 2020.

Beno Gutenberg and Charles F. Richter. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4):185–188, 1944.

Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

Taesup Kim and Yoshua Bengio. Deep directed generative models with energy-based probability estimation. *arXiv preprint arXiv:1606.03439*, 2016.

Shunichi Koshimura, Leonardo Moya, Erick Mas, and Yanbing Bai. Tsunami damage detection with remote sensing: A review. *Geosciences*, 10(5):177, 2020.

Boris Kriuk. Advancing Eurasia fire understanding through machine learning techniques. *arXiv preprint arXiv:2502.17023*, 2025a.

Boris Kriuk. Hybrid physics-ML framework for pan-Arctic permafrost infrastructure risk at record 2.9-million observation scale. *arXiv preprint arXiv:2510.02189*, 2025b.

Boris Kriuk. MorphBoost: Self-organizing universal gradient boosting with adaptive tree morphing. *arXiv preprint arXiv:2511.13234*, 2025c.

Boris Kriuk, Lincoln Ng, and Ziad A. Hossain. DeepSupp: Attention-driven correlation pattern analysis for dynamic time series support and resistance levels identification. *arXiv preprint arXiv:2507.01971*, 2025a.

Boris Kriuk, Kote Sulamanidze, and Fedor Kriuk. ELENA: Epigenetic learning through evolved neural adaptation. *arXiv preprint arXiv:2501.05735*, 2025b.

Thorne Lay and Terry C. Wallace. *Modern Global Seismology*, volume 58. Elsevier, 1995.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 2002.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Ilya Loshchilov, Frank Hutter, et al. Fixing weight decay regularization in Adam. *arXiv preprint arXiv:1711.05101*, 5(5):5, 2017.

Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

Samet Oymak. Provable super-convergence with a large cyclical learning rate. *IEEE Signal Processing Letters*, 28:1645–1649, 2021.

Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

S. Satish, H. Gonaygunta, A. R. Yadulla, D. Kumar, M. H. Maturi, K. Meduri, E. De La Cruz, G. S. Nadella, and G. S. Sajja. Forecasting the unseen: Enhancing tsunami occurrence predictions with machine-learning-driven analytics. *Computers*, 14(5):175, 2025.

K. Sharma, B. Singh, E. Herman, R. Regine, S. Samuel Rajest, and V. P. Mishra. Maximum information measure policies in reinforcement learning with deep energy-based model. In *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pp. 19–24. IEEE, 2021.

Sundaravelpandian Singaravel, Johan Suykens, and Philipp Geyer. Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Advanced Engineering Informatics*, 38:81–90, 2018.

Tokuji Utsu, Yosihiko Ogata, et al. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33, 1995.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

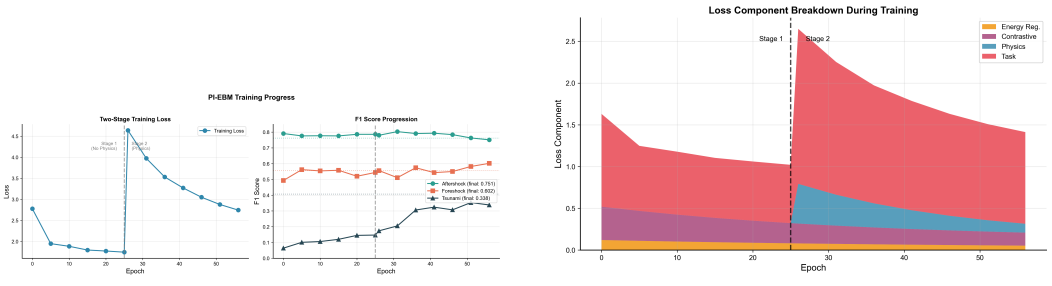
Stefan Wiemer and Max Wyss. Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4):859–869, 2000.

Jianwen Xie, Yifei Xu, Zilong Zheng, Song-Chun Zhu, and Ying Nian Wu. Generative PointNet: Deep energy-based learning on unordered point sets for 3D generation, reconstruction and classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14976–14985, 2021.

Shuangfei Zhai, Yu Cheng, Weining Lu, and Zhongfei Zhang. Deep structured energy based models for anomaly detection. In *International Conference on Machine Learning*, pp. 1100–1109. PMLR, 2016.

A Training Dynamics

Figure 7a shows the two-stage training process. Task losses dominate Stage 1, with physics losses activating at Stage 2 transition and subsequently decreasing. Figure 7b provides the detailed loss component breakdown.

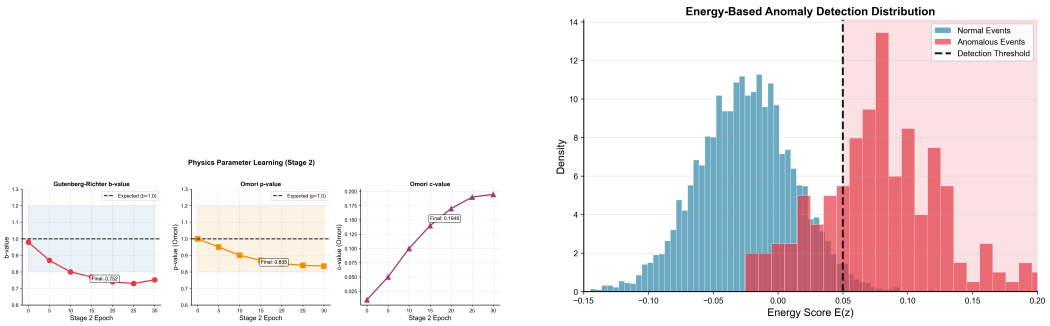


(a) Training and F1 progression.

(b) Loss component breakdown.

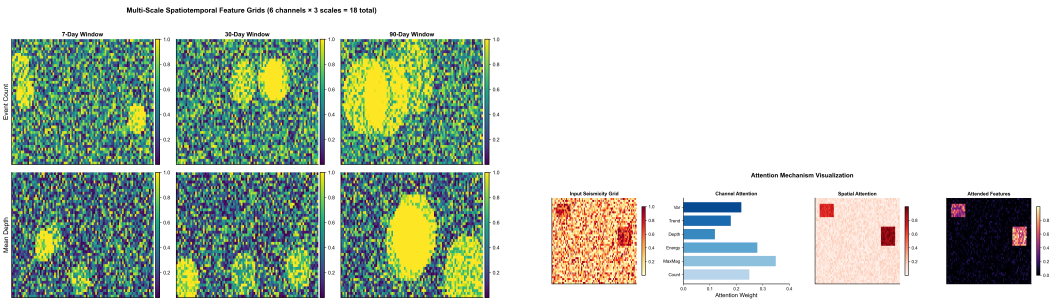
Figure 7: Two-stage training dynamics: (a) loss trajectories and task-specific F1 evolution; (b) relative loss contributions.

B EXTENDED ANALYSIS



(a) Physics parameter convergence. (b) Energy distribution analysis.

Figure 8: (a) Parameter trajectories during Stage 2 approaching theoretical values; (b) energy separation between normal and anomalous events.



(a) Multi-scale feature contributions. (b) Spatial attention visualization.

Figure 9: (a) Attention weights and ablation per temporal scale; (b) learned spatial attention over seismicity grids.

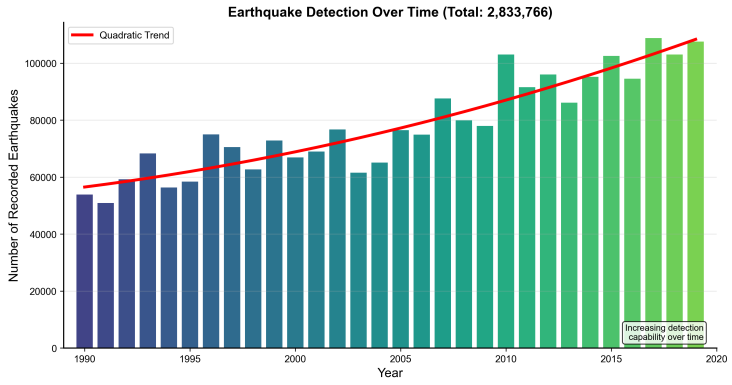


Figure 10: Temporal distribution of earthquake detection showing increasing catalog completeness over time.