

Machine learning predicts meter-scale laboratory earthquakes

Received: 26 September 2024

Accepted: 17 September 2025

Published online: 30 October 2025

 Check for updatesReiju Norisugi¹✉, Yoshihiro Kaneko¹ & Bertrand Rouet-Leduc²

In recent years, there has been a growing interest in utilizing machine learning (ML) to investigate the predictability of shear-slip failures, known as laboratory quakes, in centimeter-scale rock-friction experiments. However, the applicability of ML to larger-scale laboratory quakes and natural earthquakes, where important timescales vary by orders of magnitude, remains uncertain. Here, we apply an advanced ML approach to meter-scale laboratory quake data, characterized by accelerating foreshock activity manifesting as increasing numbers of tiny acoustic emission events. We demonstrate that a trained ML model, using a network representation of the event catalog, can accurately predict the time-to-failure of meter-scale mainshocks, from tens of seconds to milliseconds before the upcoming main quakes. These timescales correspond to approximately decades down to weeks in the context of large earthquakes. By comparing our results with a dynamic model of shear failures that replicates the experimental data, we suggest that tracking the evolution of shear stress on creeping fault areas, rather than nominal shear stress, indirectly through the acoustic emission events, enables ML to predict both numerical and laboratory quakes. These findings provide critical insights into fault conditions that may facilitate short-term forecasting of earthquakes in nature.

Forecasting earthquakes with reliable accuracy remains a formidable and unresolved challenge in seismology¹, in part due to the inaccessibility of seismogenic zones within the Earth's crust and the scarcity of in situ data. There was once considerable optimism within earthquake science regarding the potential for short-term forecasting, based on the understanding that dynamic fault failures (earthquakes) cannot occur without a precursory, near-failure phase where the shear stress is just below the fault strength^{2,3}. During this phase, micro-fracturing⁴ or quasi-static preslip⁵ typically accelerates due to the delayed failure characteristics inherent to rock fracture and friction. However, the evolution of physical conditions associated with the precursory phase in natural, heterogeneous fault zones remains inaccessible and poorly understood. Additionally, recent developments have led to growing skepticism regarding the viability of utilizing short-term precursors. Observational studies^{6,7} have shown that small and large earthquakes are indistinguishable until rupture begins to cease, suggesting that the processes occurring during the precursory phase exert minimal

influence on the eventual size of the dynamic rupture, thereby limiting their utility as reliable precursors to large earthquakes. Furthermore, possible precursory signals, such as slow slip events, were identified only retrospectively, after a large earthquake had occurred^{8,9}, and only a small fraction of slow slip events have been linked to the occurrence of large quakes¹⁰.

Nevertheless, recent studies utilizing machine learning (ML) techniques—statistical algorithms designed to uncover hidden patterns in complex, multi-dimensional data—have reignited optimism in the potential for earthquake forecasting. When applied to laboratory experiments that simulate natural faulting through the shearing of rock samples or gel materials, ML techniques have successfully predicted the timing of laboratory quakes by analyzing precursory acoustic emissions (AE)^{11–23}. The evolution of nominal shear stress within laboratory fault systems, as indicated by increasing variance in AE signal amplitude, has been associated with the preparatory phase of quakes^{11,12,24}. Additionally, ML models trained on ultrasonic (high-

¹Department of Geophysics, Kyoto University, Kyoto, Japan. ²Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan.

✉ e-mail: norisugi.reiju.77e@st.kyoto-u.ac.jp

frequency) pulse data have shown promise in predicting both the time to failure and the evolution of shear stress¹⁹. These findings suggest that precursory signals, such as increasing AE variance and the destruction of weak microscopic asperities, may serve as proxies for the onset of foreshock activity^{25–29}.

Despite these advancements, several critical challenges remain unresolved. First, most studies on the predictability of laboratory quakes have predominantly utilized centimeter-scale rock samples or sub-meter-scale gel models^{11–23}. Given the scale dependency observed in frictional behavior during meter-scale rock-friction experiments³⁰, it is essential to investigate predictability in larger-scale rock failures to effectively extrapolate these findings to natural faulting scenarios. Second, the predictive performance immediately preceding laboratory quakes—during the final few percent of recurrence intervals—has been seldom addressed. In the context of large earthquakes, this period corresponds to timescales of several years, making it imperative to develop methodologies that enable reliable predictions on much shorter timescales. Third, the underlying mechanisms behind the predictive capabilities of ML remain elusive. This gap is particularly pronounced when it comes to understanding the physical processes and the evolution of fault conditions, as direct observation of the physical state within fault zones (e.g., local shear stress or slip velocity) is challenging, even in controlled laboratory settings.

In this study, to address these questions, we leverage data from a meter-scale rock-friction laboratory experiment³¹ and apply a machine learning (ML) approach to predict both the timing of laboratory quakes and the evolution of on-fault shear stresses. The experimental fault exhibits a heterogeneous distribution of gouges and damage zones, leading to complex foreshock activity characterized by an inverse Omori's law and a temporal decrease in the Gutenberg-Richter *b*-value preceding laboratory quakes³¹. These phenomena have been reported occasionally in various laboratory experiments^{32–43} and in natural settings^{44–52}. Our focus is on elucidating the evolution of physical variables and associated slip phenomena that may enable short-term forecasting of imminent meter-scale laboratory quakes.

Results

Machine learning predicts time to meter-scale laboratory quakes and nominal shear stress evolution

We apply machine learning (ML), specifically the Random Forest (RF) algorithm, to catalog data derived from a large-scale, rock-friction laboratory experiment³¹ (Fig. 1a). This experiment (LB12-011), conducted using a ground shaking table, featured a fault with a heterogeneous distribution of fault gouge (Fig. 1b) resulting from a prior high-speed sliding test. It also exhibited active acoustic emission (AE) events preceding 1.5-m × 0.1-m laboratory quakes^{30,31}. The rock sample, metagabbro from Tamil Nadu, India, consisted of a lower block fixed to a shaking table and an upper block supported by an isolated reaction force bar (Fig. 1a). Shear loading (0.01 mm/s) was applied parallel to the fault, with a constant normal load of 6.7 MPa from three jacks.

In this study, we refer to the failure events as laboratory quakes and categorize all other AE events as foreshocks. The events are characterized using a network representation⁵³ to generate input variables for the ML model, which we associate with the time remaining before the next laboratory quakes. The experimental data, spanning approximately 900 seconds, were divided into a training set (covering approximately 200 to 600 seconds, during which 19 laboratory quakes occurred, excluding the initial 200 seconds to allow the experiment to reach a steady state), a validation set (from ~600 seconds to 700 seconds, with 4 laboratory quakes), and a test set (from ~700 to 900 seconds, with 11 laboratory quakes). Detailed descriptions of the experimental data and the forecasting technique are provided in the Methods section.

We find that the trained ML model successfully predicts the timing of laboratory quakes, with predictions ranging from tens of seconds (Fig. 1c) to milliseconds (Fig. 1d) before the quakes. The prediction score is evaluated by the coefficient of determination based on logarithmic values (Eq. (4)), which is $\log R^2_{\text{lab-quake}} = 0.84$ for the test set. Remarkably, the model can accurately forecast the occurrence of the next quake down to the millisecond before it happens (Fig. 1d), based solely on current and past catalog information of tiny foreshocks, despite their sparse distribution in time and space and the potential incompleteness of the event catalog. The model's short-term forecasting accuracy is notably enhanced when the number of foreshocks per quake is relatively large (e.g., the 6th and 7th events in Fig. 1d), allowing predictions down to milliseconds before the quakes, with an accuracy comparable to within 0.1% of the recurrence intervals of the laboratory quakes. However, the ML model encounters challenges in accurately predicting upcoming quakes when the number of foreshocks per quake is relatively small (e.g., the 5th, 8th, 9th, and 10th events in Fig. 1d). This inaccuracy, particularly in the moments leading up to the quakes, is likely due to insufficient event detection and sensor resolution.

The effective features for predicting the time to quakes can also predict the evolution of the nominal shear stress. Fig. 2a, b illustrate the prediction of nominal shear stress by the trained ML model, yielding $\log R^2_{\text{lab-stress}} = 0.81$ for the test set. The nominal shear stress data are sampled during foreshocks and quakes, and it is detrended by using a least squares method on the training set (see details in “Method”). The ML model learns to associate this stress with the same input features used for predicting the time to laboratory quakes.

We also train a simple deep learning (DL) model to predict the time to laboratory quakes and nominal shear stress. The DL models achieve accuracy comparable to ML models, demonstrating the robustness of our results regardless of the predictive model used (see “Methods” and Figs. S5 and S6 for details). Additionally, we construct and apply a simple inter-event time model^{11,19}, which relies solely on recurrence intervals of laboratory quakes (see Methods). With low prediction scores (Fig. S2), this model serves as a baseline, emphasizing the advantage of ML and DL approaches.

These results suggest that a sufficiently trained ML or DL model can infer the current state of the fault using only event catalog data over both long- and short-term scales, and more generally that the foreshock statistics encode the state of the fault.

Examining the evolution of the features used to predict the time to laboratory quakes and the nominal shear stress reveals a systematic pattern (Fig. 2c). Specifically, the average event interval and seismic moment over specific timescales (e.g., Δt_{20f} and $\bar{M}_{20f}$, respectively) either logarithmically decrease or increase as the laboratory quakes approach. The distinct logarithmic differences in these features at various times before a laboratory quake suggest that they play an important role in the accuracy of the predictions.

Evolution of event interval and seismic moment that enables accurate prediction

To understand the relationships between input features of the ML model and the prediction outcomes, we analyze the distribution of input data as a probability density via kernel density estimation. Figure 3a, b show the distribution of event intervals (Δt_{20f}) and seismic moments ($\bar{M}_{20f}$) averaged over specific timescales as functions of time to quake. These features show a clear trend: the event intervals decrease, and seismic moments increase as the quake approaches, as indicated by the green arrows in Fig. 3a, b, starting approximately one second before the quakes occur. The decreasing event interval reflects the characteristics of an inverse Omori's law, while the increasing seismic moment suggests a temporal decrease in the Gutenberg-Richter *b*-value (although the number of events is not sufficient for

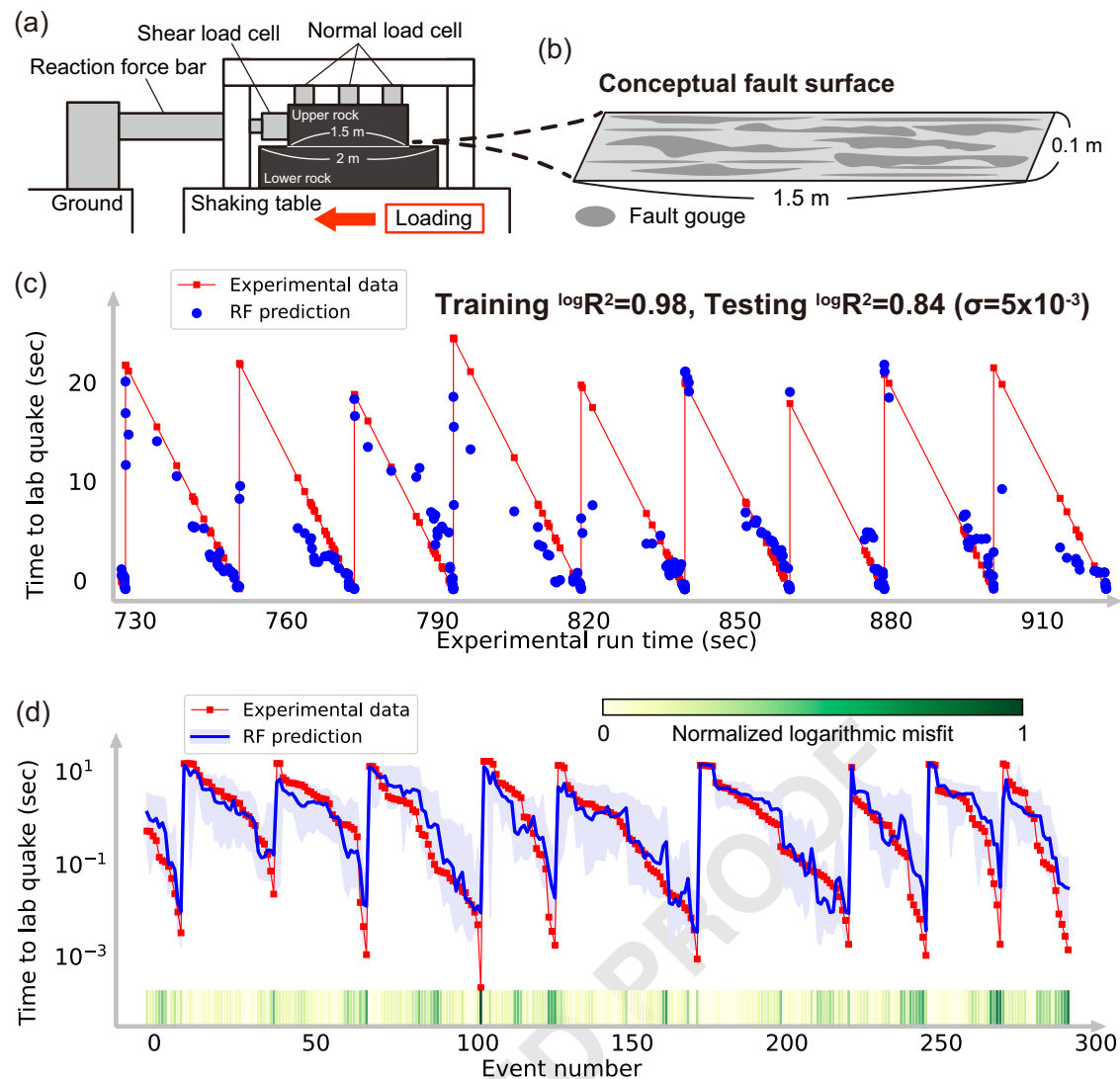


Fig. 1 | Experimental setup and Random Forest (RF) prediction of time to laboratory quakes in the test set. a, b Conceptual illustration of the meter-scale rock-friction laboratory experiment³¹. **a** Schematic of the experimental setup showing two rectangular rock blocks (1.5 m and 2 m in length) being sheared by a large shaking table and a reaction force bar (indicated by the red arrow) under constant loading conditions: normal stress of 6.7 MPa, a loading rate of 0.01 mm/s, and a total slip of approximately 7 mm. **b** Simulated fault surface with a heterogeneous distribution of gouge. **c** Prediction of the time to laboratory quakes by the trained RF model on the scale of tens of seconds **(c)** and on the event number scale **(d)**. The red lines and dots represent the actual target values, while the blue dots

and curves show the RF model predictions. In **d**, the blue-shaded area indicates the 5 to 95 percentile range of predictions of the trees in the RF model, and the green color bar represents the normalized logarithmic misfit. Predictions are irregularly spaced in time, corresponding to the occurrence of events. The event number axis corresponds to the discrete time series, with a focus on the period immediately preceding quakes. The performance is quantified by the coefficient of determination for logarithmic values (as defined in Eq. (4)). The standard deviation σ is obtained by 100 RF models with different random seeds under the optimized hyperparameters to show the stability of R^2 .

stable b -value computation). The event interval also exhibits a distinct decreasing trend during the loading phase, where the nominal shear stress linearly increases with time (Fig. 2a). These features, which serve as inputs to the ML model, allow our approach to effectively capture and quantify these precursory seismicity characteristics, and explain the reason why our ML model can accurately predict the timing of meter-scale laboratory quakes using only catalog information.

In contrast, the nominal shear stress does not correlate with the input features used in our ML model. The relations between the nominal shear stress and the input features ($\overline{\Delta t_{20f}}$ and $\overline{M_{20f}}$) exhibit a monotonous flat distribution relative to shear stress just before quakes (highlighted by the green circles in Fig. 3c, d). A correlation between the features and the labels is observed only in the event interval during the loading phase (indicated by the gray arrow in Fig. 3c), extending up to approximately one second before the quakes. By comparing Fig. 2b

and 2c, we see that the effective features (such as $\overline{\Delta t_{20f}}$ and $\overline{M_{20f}}$) continue to evolve even during the shear stress stagnation, which contrasts with the trend observed when using the time to quakes as the axis (Fig. 3a, b). When nominal shear stress plateaus roughly one second before quakes (indicated by the nearly horizontal red curve in Fig. 2b), foreshock activity significantly increases, suggesting that the plateau is associated with fault stress unloading via foreshocks. Therefore, although the ML model accurately predicts both the time to quakes and the evolution of nominal shear stress, implying apparent causality (Figs. 1, 2), nominal shear stress itself may not be a direct controlling factor for the predictability of the laboratory quakes during the precursory phase, contrary to inferences from previous studies²⁴. The model's predictive success likely arises from its ability to capture the subtle evolution of other features rather than relying solely on nominal shear stress.

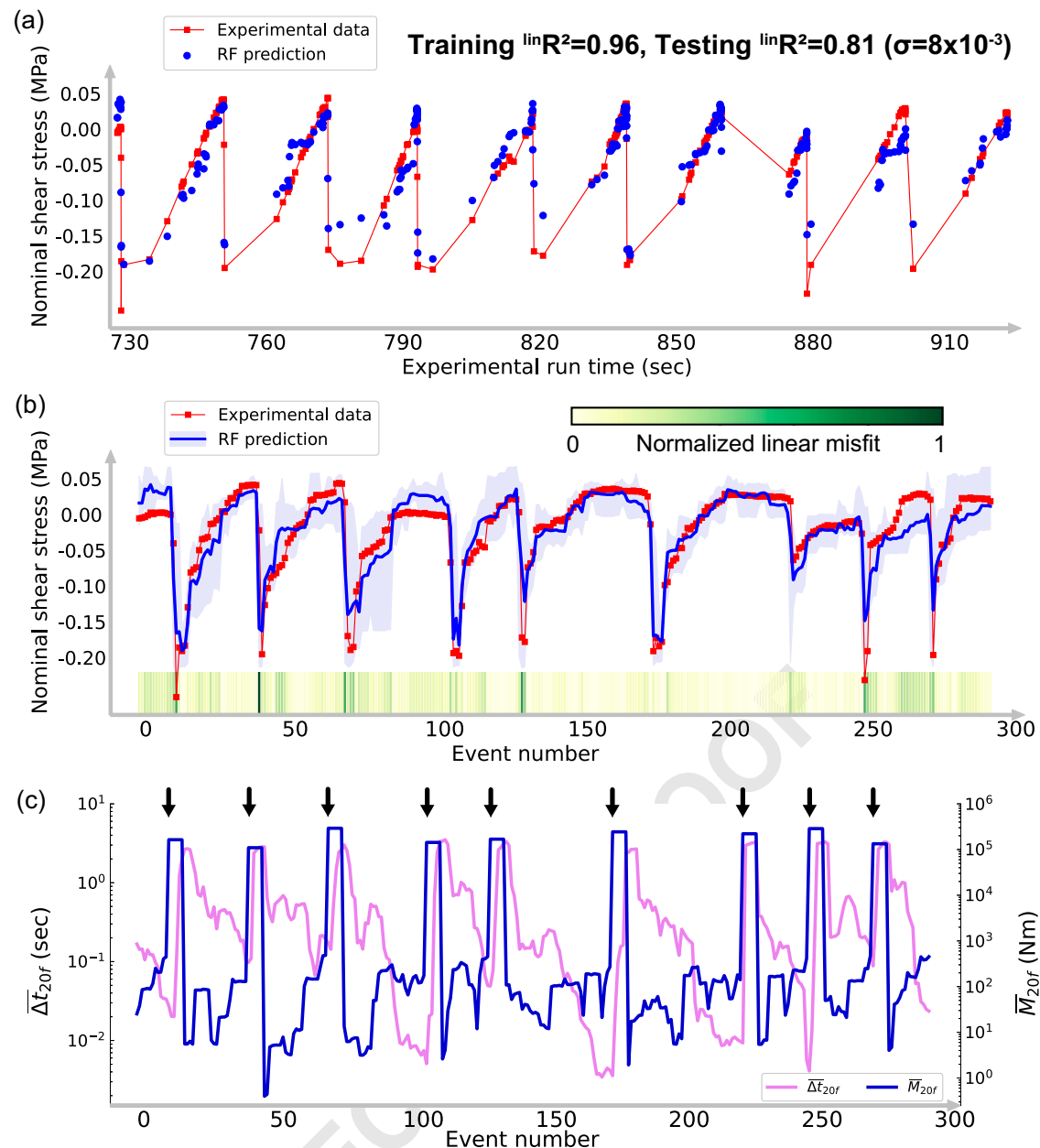


Fig. 2 | Prediction of nominal shear stress in the test set by the trained Random Forest (RF) model. Prediction of the nominal shear stress over a timescale of tens of seconds (a) and on the event number scale (b). The red lines and dots represent the actual target values, while the blue dots and curves indicate the predictions made by the trained RF model. The blue-shaded area represents the 5 to 95% percentile range of the predictions, and the green color bar reflects the normalized linear misfit. The prediction accuracy is quantified using the coefficient of determination linR^2 (as defined in Eq. (5)). The standard deviation of R^2 score σ is

computed from 100 RF models with different random seeds. c Temporal evolution of the event interval (Δt_{20r} , shown in pink) and seismic moment ($\bar{M}_{20r}$, shown in blue) averaged over specific timescales (see Methods). These features are direct inputs for the trained RF model used to generate the predictions shown in Figs. 1c, d and 2a, b. The timings of laboratory quakes are indicated by black arrows. The flat portions of $\bar{M}_{20r}$ following each laboratory quake arise from the significant contribution of individual quake data points at this specific network size.

Insights from synthetic seismicity generated by an earthquake-cycle model

To examine why an ML model can accurately predict the timing of the laboratory earthquakes, we set up and simulate seismicity that replicates the laboratory experimental data using a fully dynamic model of earthquake cycles⁵⁴. For simplicity, the model contains a one-dimensional planar strike-slip fault embedded in an elastic continuum (Fig. 4a). The physical parameters are set according to elastic constants and sample size in the laboratory experiment³⁰. Simplified frictional heterogeneity mimicking the response of the fault gouge and bare surface is assumed, with alternating velocity-strengthening (VS)

and velocity-weakening (VW) patches to generate accelerating foreshock activity (Fig. 4b). Details of the model setup are provided in the Methods section.

The model generates synthetic quakes with a recurrence interval of approximately 12 seconds and a moment magnitude of $M_w \sim -2.5$, along with foreshocks in the range of $-6 < M_w < -3$, consistent with the laboratory quakes (see Methods and Fig. S7 for details). To avoid contamination by numerical artifacts, we remove data for earthquakes smaller than the magnitude completeness threshold ($M_w = -4.6$) from the catalog (Fig. S7b). To further replicate the conditions of the laboratory experiments (see Fig. 1d), we exclude foreshock data that

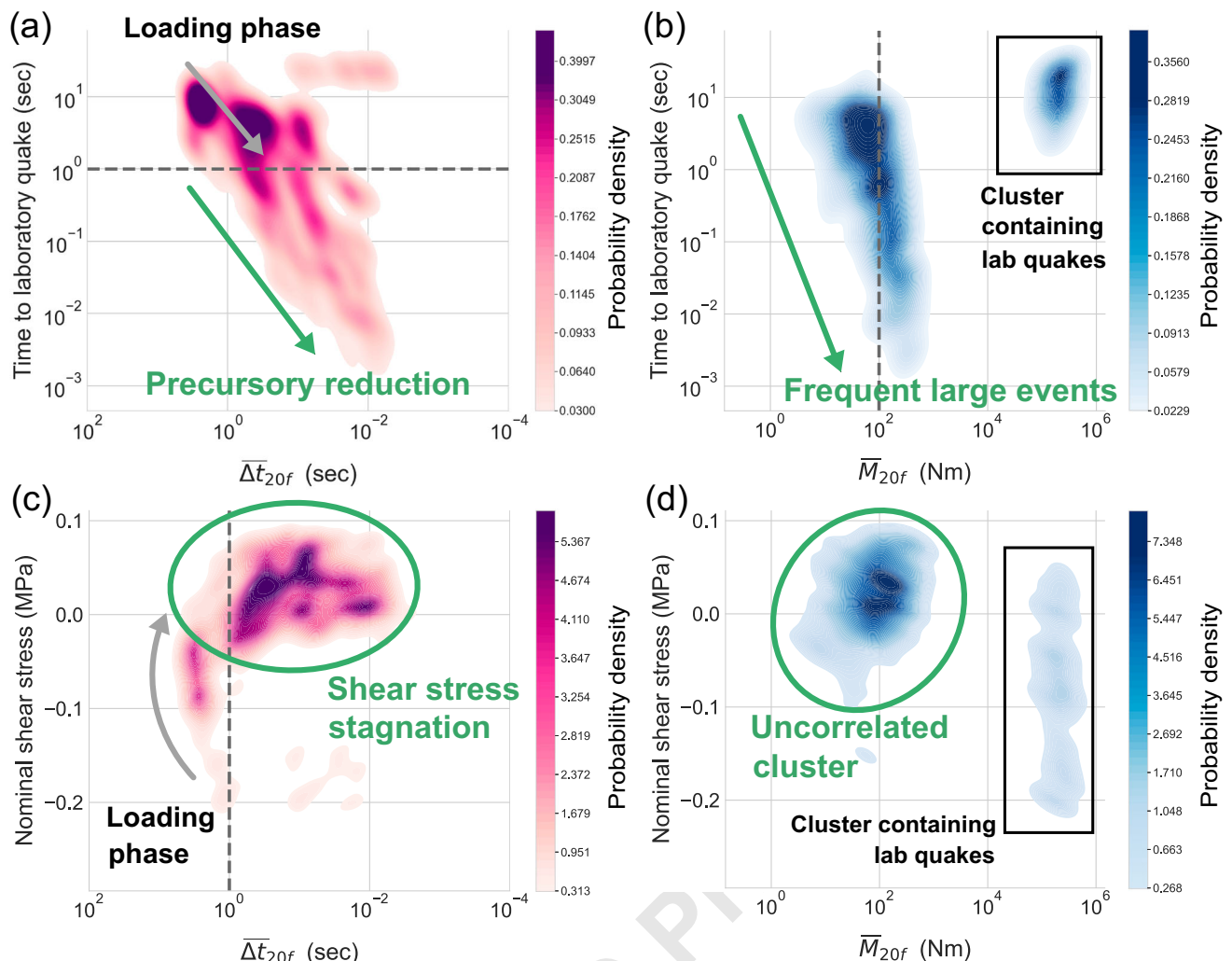


Fig. 3 | Important feature distributions for the accurate prediction of laboratory quakes and nominal on-fault shear stresses. **a, b** Distributions of important features as a function of time to quake: pink and blue indicates the distributions of $\Delta \bar{t}_{20f}$ (event interval averaged over certain timescales), and $\bar{M}_{20f}$ (seismic moment averaged over certain timescales), respectively. **c, d** Distributions of important features as a function of nominal shear stress. The contour lines show the

probability density of the data approximated by kernel density estimation, with dark/light colors indicating high/low data density in the feature-target space. The arrows emphasize the temporal evolution of these features. The right-side separated cluster in **(b)** and **(d)** are data containing laboratory quakes. Other important features (Fig. S15a) depict a similar trend.

occur within 10^{-3} seconds of the main failures, aligning with the laboratory's detectability of tiny events. We calculate foreshock counts and corresponding b -values for the stacked catalog (see Methods) and confirm that the foreshocks exhibit a temporal decrease in b -values⁵⁴ and follow an inverse Omori's law⁵⁵, beginning several milliseconds before synthetic quakes (Fig. S7c). For ML prediction, the dataset is divided as follows: the training set spans 70–770 seconds of simulation time (including 52 quakes, with the first four cycles removed), the validation set covers 770–870 seconds (7 quakes), and the test set spans 870–1070 seconds (15 quakes).

As in the laboratory quake case, the trained ML model can accurately predict the timing of synthetic quakes, achieving $\log R^2_{\text{syn-quake}} = 0.83$ and nominal shear stress with $\ln R^2_{\text{syn-avg}} = 0.76$, over timescales ranging from ten seconds to milliseconds (Fig. 4c, d). Since the numerical model provides direct access to local fault conditions, the prediction of the shear stress averaged over the VS patches (referred to as VS shear stress) is reported, yielding $\ln R^2_{\text{syn-VS}} = 0.70$ (Fig. 4e). Notable mispredictions occur primarily during large peaks in both average shear stress (negative peaks) and VS shear stress (positive peaks) (Fig. 4d, e). These peaks are attributed to the effects of aftershocks, as the model predicts an increase in VS shear stress following the rupture

propagation of synthetic quakes. Due to the rarity of aftershock sequences in this model, the ML model does not sufficiently learn from the limited aftershock data available, leading to wider 5 to 95% percentile ranges during these periods. However, aside from the misfits immediately following synthetic quakes, the predictions are accurate, with the ML model closely tracking the stress evolution, particularly the VS shear stress, which shows a narrow percentile range, indicating that most RF trees generate similar prediction values (Fig. 4d, e).

The machine learning model tracks the evolution of shear stresses on creeping fault patches, enabling accurate prediction

Upon analyzing the evolution of key input features critical for accurate predictions, we observe similarities between the synthetic and laboratory data, despite the simplification of the simulation. In the synthetic catalog, the average event interval ($\Delta \bar{t}_{20f}$) decreases logarithmically as a quake approaches (indicated by the green arrow in Fig. 5a), while the seismic moment ($\bar{M}_{20f}$) concurrently increases (green arrow in Fig. 5b). These trends, which reflect an inverse Omori's law and a decrease in the Gutenberg-Richter b -value, allow the ML model to predict the timing of synthetic quakes across a wide range of timescales, from ten seconds to milliseconds before the events.

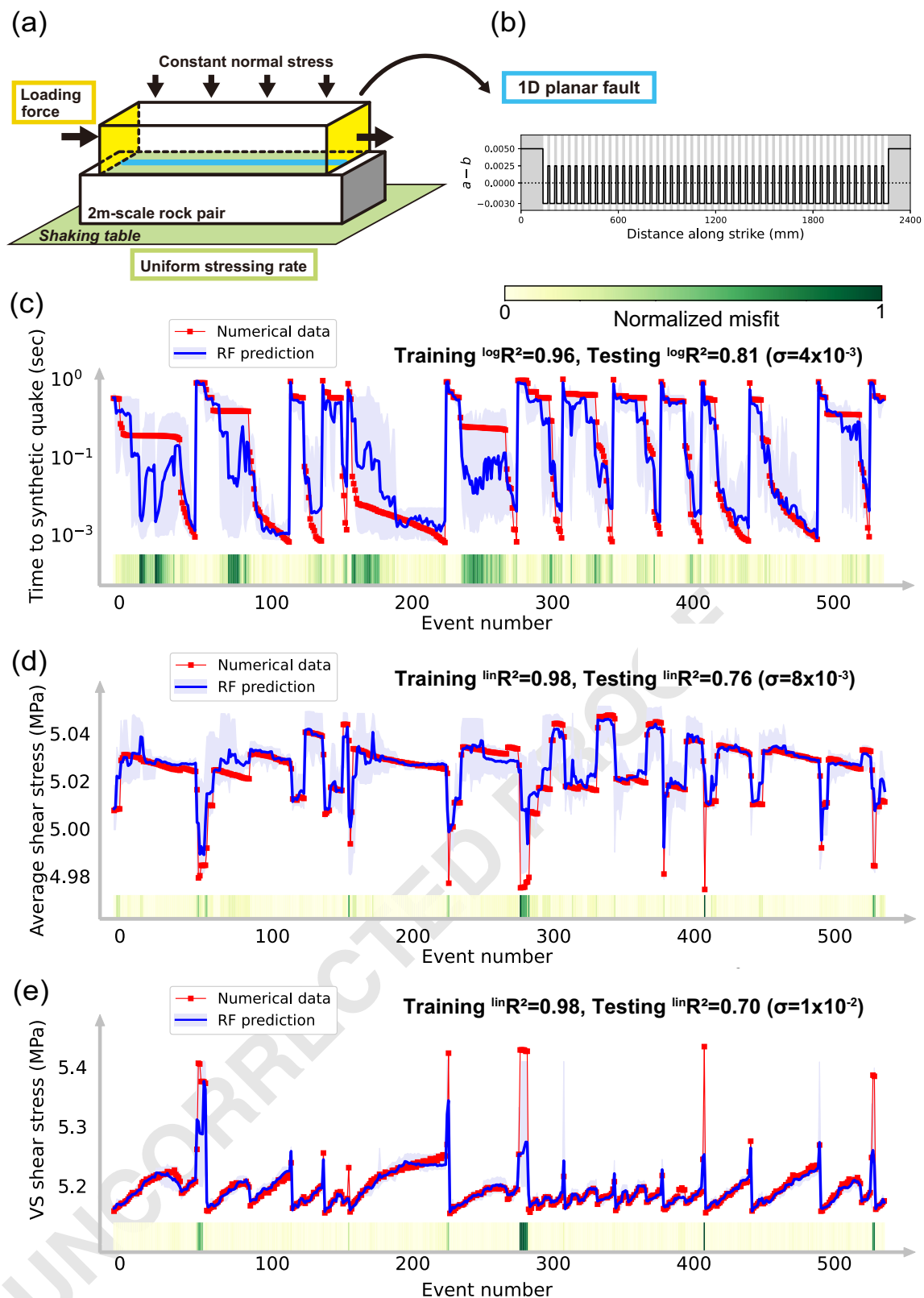


Fig. 4 | Setup of numerical model and the Random Forest (RF) prediction of synthetic quakes and shear stresses in the test set. a Conceptual illustration of the fully-dynamic earthquake-cycle model. **b** The fault surface parameterized by a heterogeneous distribution of velocity-strengthening (VS) and velocity-weakening (VW) friction. Predictions of time to synthetic quakes (c), average shear stress (analogous to nominal shear stress in the laboratory experiment) (d), and shear stress averaged over the VS patches (e). The red curves represent true values, while

the blue curves show predictions by the trained RF model. The blue-shaded areas indicate the 5 to 95% percentile range of the predictions. The green color bars indicate the normalized logarithmic misfit in (c) and the normalized linear misfit in (d, e). The definitions of R^2 follow $\log R^2$ (Eq. (4)) in (c) and $\text{lin} R^2$ (Eq. (5)) in (d, e). The standard deviation of prediction score σ is computed from 100 RF models randomly trained with different seeds.

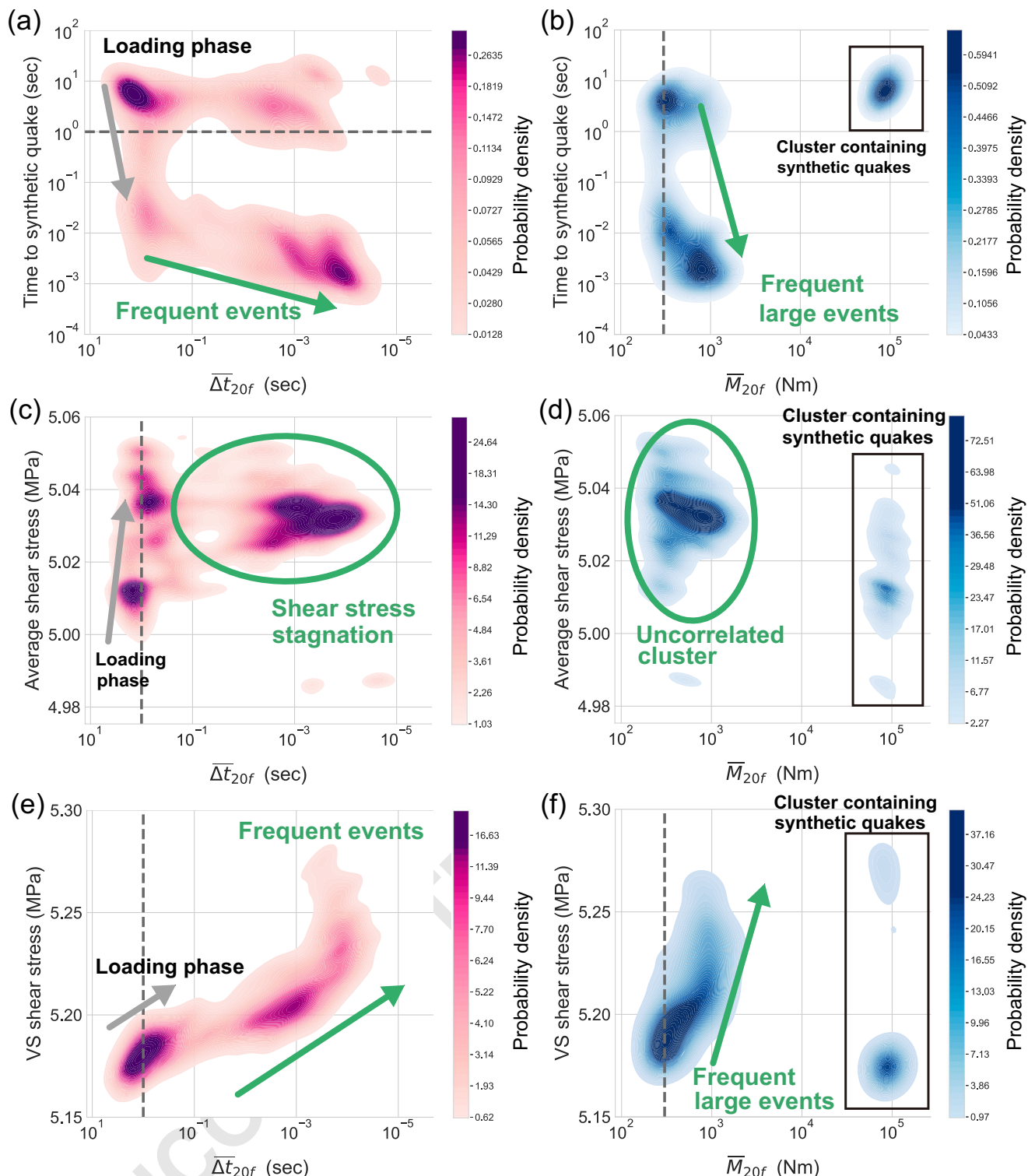


Fig. 5 | Important feature distributions for the accurate prediction of synthetic quakes, nominal on-fault shear stress, and velocity-strengthening (VS) shear stress. Pink and blue colors represent the event interval ($\overline{\Delta t}_{20f}$) and seismic moment ($\overline{M}_{20f}$), respectively, averaged over certain timescales as functions of the time to synthetic quakes (a, b), average shear stress (c, d), and VS shear stress (e, f). The contour lines depict the probability density of the data, estimated via kernel

density estimation, with dark colors indicating high data density and light colors indicating low data density within the feature-target space. Arrows emphasize the temporal evolution of these features. The right-side separated clusters in (b), (d), and (f) represent data involving synthetic quakes. Other important features (Fig. S15b) depict a similar trend.

Furthermore, the average shear stress correlates with $\overline{\Delta t}_{20f}$ only during the loading phase where the shear stress continues to rise, up to one second before the synthetic quakes (as depicted by the gray arrow in Fig. 5c). This leads to a flattened, uncorrelated pattern just before the

synthetic quakes, as the average shear stress nearly stagnates (highlighted by the green circles in Fig. 5c). Similarly, $\overline{M}_{20f}$ forms an uncorrelated cluster when plotted against average shear stress (green circle in Fig. 5d).

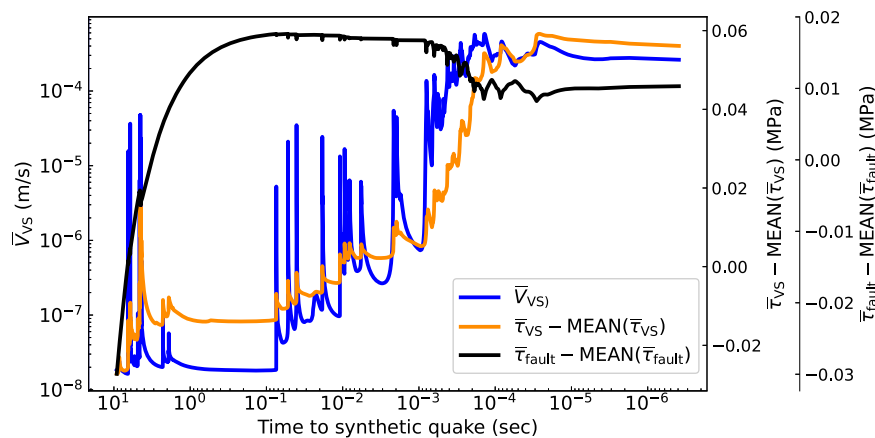


Fig. 6 | The evolution of fault slip velocity, average shear stress, and velocity-strengthening (VS) shear stress over a representative synthetic-quake cycle.

The time interval corresponds to the period from 5 seconds after the 39th up to the 40th synthetic quake within the training set. The blue curve shows the fault slip

velocity averaged over the VS patches $\bar{V}_{VS}$. The orange and black curves are shear stress averaged over the VS patches ($\bar{\tau}_{VS}$) and the entire fault ($\bar{\tau}_{fault}$), respectively. The mean values ($MEAN(\bar{\tau}_{VS})$ and $MEAN(\bar{\tau}_{fault})$) are removed from each stress curve.

Surprisingly, the relationships between input features and VS shear stress exhibit clear correlations both during the loading phase (gray arrow in Fig. 5e) and immediately before synthetic quakes (green arrow in Fig. 5e), with a logarithmic decrease as the VS shear stress increases linearly. Likewise, $\bar{M}_{20f}$ shows a logarithmic increase (green line in Fig. 5f). These results suggest that precursory seismicity is primarily driven by an increase in VS shear stress, rather than the nominal shear stress on the fault, thereby enabling the ML model to accurately predict synthetic quakes (Fig. 4c) –an insight that is difficult to obtain from laboratory data alone.

To directly investigate how VS shear stress increases leading up to a synthetic quake, we analyze the relationship between slip velocity and VS shear stress throughout a single quake cycle in the numerical model (Fig. 6). During the loading phase, the VS shear stress gradually rises over time, accompanied by sporadic foreshocks. In contrast, during the precursory phase, slip velocity on the VS patches escalates rapidly due to afterslip triggered by numerous foreshocks (orange and blue jagged curves in Figure 6). When a foreshock rupture arrests on VS patches, both local slip velocity and shear stress increase, enhancing the likelihood of rupture nucleation in neighboring VW patches and increasing the probability that ruptures will propagate through the VS patches (Fig. S8). This process leads to more frequent larger foreshocks and a reduction in b -values⁵⁴. This positive feedback mechanism occurs when the nominal shear stress is elevated and begins to plateau, resulting in the distinct evolution of average shear stress versus VS shear stress (Fig. 6). Ultimately, one of the ruptures originating from a VW patch, similar to the foreshocks, propagates through multiple VS barriers, culminating in a main quake. As a result, precursory seismicity is systematically driven by local stress increases on creeping barriers, which are captured indirectly through the seismicity catalog via the network representation.

Discussion

Based on the consistency between our laboratory and numerical results, we propose a plausible explanation as to why machine learning (ML) can predict laboratory earthquakes. At the onset of quake cycles, foreshocks occur sporadically (white circles in Fig. 7a) at points of stress concentration, most commonly at the boundaries between gouge layers and bare fault surfaces. The ruptures of these foreshocks are likely arrested by creeping barriers, such as gouge layers (depicted as gray areas in Fig. 7a). As the fault becomes critically stressed, meaning the nominal shear stress on the fault reaches a critical level (approximately one second before a quake in this experiment), a

positive feedback between the nucleation of foreshocks and subsequent prolonged creep on VS patches accelerates toward a laboratory quake (Figs 3 and 5). This precursory phase is marked by a stagnation of the nominal shear stress, an expanding zone of foreshocks, an inverse Omori's law and a b -value reduction³¹, while the shear stress on the creeping patches increases rapidly (Figs. 6 and 7d, e). The likelihood of one of the nucleating ruptures growing into a quake larger than the foreshock zone increases, eventually leading to a main quake (green star in Fig. 7b). After the main quake, a qualitatively similar process repeats, although the characteristics of seismicity—such as the number of foreshocks and their nucleation sites—vary depending on the heterogeneous distributions of residual shear stress and strength (Fig. 7c), even after a complete failure in the earthquake cycle.

Our results demonstrate that a trained ML model, utilizing a network representation of the event catalog, can accurately predict the time-to-failure of meter-scale mainshocks, ranging from tens of seconds to milliseconds before the upcoming main quakes. These timescales, corresponding to approximately decades down to weeks transposed in the context of large earthquakes, could potentially offer short-term warnings if applied to natural large quakes. However, both the laboratory and numerical faults in our study were highly idealized compared to natural faults.

To assess the impact of different degrees of frictional heterogeneity on ML prediction accuracy, we analyze numerical models with varying heterogeneity levels (i.e., different $a - b$ of velocity-weakening patches; Figs. S9–S11). While prediction accuracy varies with the degree of heterogeneity (Figs S9 and S10), as long as foreshocks arise from the same underlying physical mechanism, the reason for accurate mainshock prediction remains consistent with the default case shown in Fig. 4. However, when barriers to rupture become more effective, as parameterized by weaker velocity-weakening patches, earthquake swarm-like behavior emerges without distinct mainshocks (Fig. S11). In this scenario, prediction scores drop significantly because well-defined mainshocks are absent (Fig. S11).

An additional complexity expected for natural faults is the occurrence of confined ruptures that do not reach the fault edges. In the laboratory experiment with the elongated rock sample we considered, all mainshocks ruptured the entire fault surface³¹. To evaluate the applicability of ML prediction to confined ruptures, we develop a numerical model that simulates both partial and full ruptures (see “Methods” and Fig. S12). Despite the presence of partial ruptures, the ML model accurately predicts the timing of both rupture types (Fig. S12), suggesting that ML predictions remain effective for confined

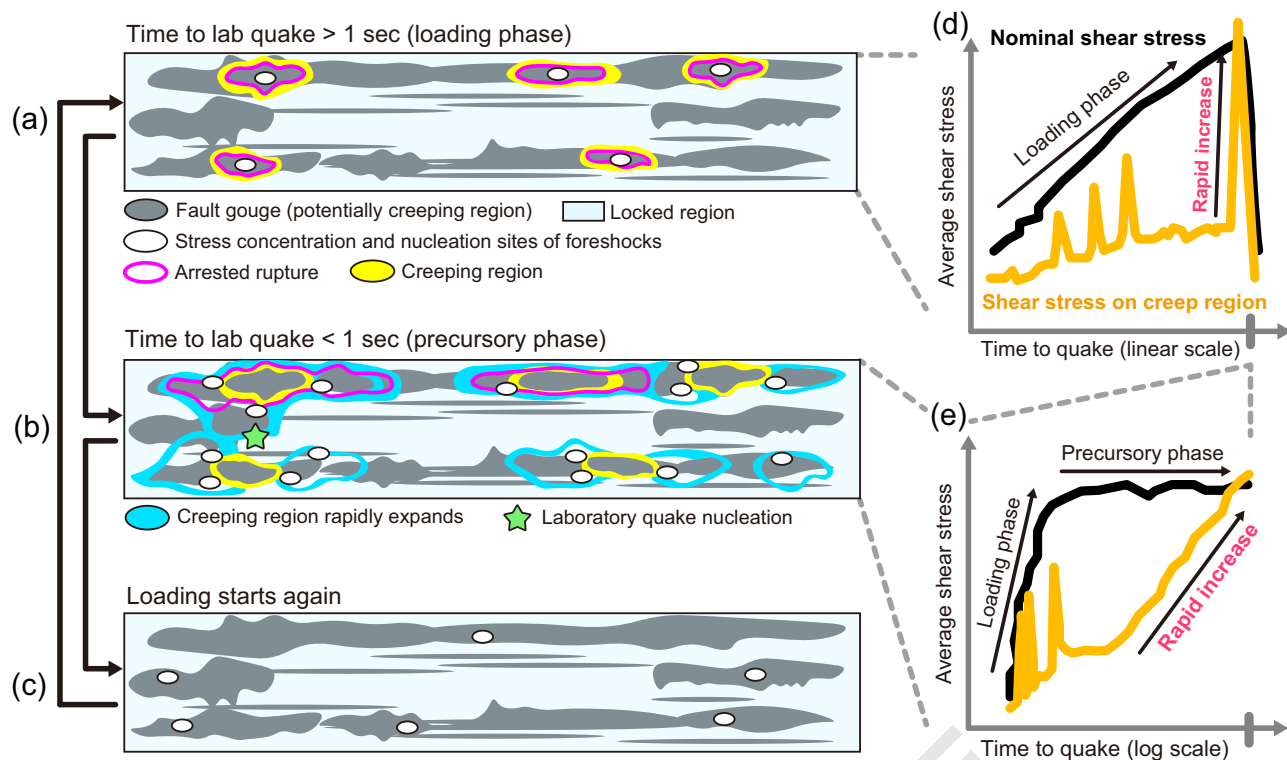


Fig. 7 | Conceptual model illustrating how machine learning predicts a mainshock on a fault. a–c A phenomenological depiction of plausible scenarios occurring on the fault surface during a single laboratory quake cycle. **d, e** The evolution of key physical parameters relevant to the emergence of precursory

seismicity. The event cycle is divided into two phases: the loading phase, where the nominal shear stress progressively increases, and the precursory phase, where the nominal shear stress plateaus.

ruptures as long as they exhibit distinct magnitude-frequency distributions from the rest of the foreshocks.

Neither the laboratory experiment nor the numerical model in this study accounts for off-fault seismicity, such as that occurring on subsidiary faults and within fault damage zones, which would introduce additional complexity in applying ML to predict mainshocks. In addition, fault geometrical complexity and pore fluid pressure evolution, which are not considered in our current laboratory and numerical setting, could influence the occurrence of foreshocks⁴³. Furthermore, unlike in laboratory experiments, foreshocks are relatively rare in natural settings, possibly due to detection limitations, posing a challenge for applying catalog-based ML or DL approaches. Indeed, a less heterogeneous case in a meter-scale laboratory experiment³¹ produced too few foreshocks for our catalog-based ML method to be applicable. However, ML or DL models that incorporate a network representation of continuous waveform data may help overcome this limitation.

Another key challenge lies in the limited availability of seismicity catalogs that span multiple mainshock cycles, which are essential for effectively training ML models. As demonstrated with laboratory data, mainshock prediction accuracy strongly depends on training data size (Fig. S13). A promising approach to addressing this issue is the incorporation of transfer learning techniques²⁰, where synthetic data (e.g., from numerical or laboratory seismicity) is used to train an ML model, which can then be adapted to real-world catalogs with sparse data. This approach, when combined with the network representation method, may help overcome the current limitations in predicting natural seismic events.

The underlying mechanisms behind the predictive capabilities of ML allow us to extrapolate these findings to natural faulting scenarios. In natural, heterogeneous fault zones, the evolution of physical

conditions within a foreshock zone during the precursory phase may be characterized by increasing shear stresses on localized, slowly slipping parts of faults. Although it remains difficult to predict the eventual size of an earthquake before rupture begins, our results suggest that the likelihood of a rupture expanding within the foreshock zone increases during this precursory phase. Furthermore, the timing of mainshocks appears to be closely linked to these slowly slipping regions, which could contribute to the emergence of precursory seismicity and hence more predictable mainshock patterns. In natural fault zones, such slowly slipping regions are often associated with various types of slow earthquake activities⁵⁶. Therefore, more detailed monitoring of slow earthquakes and local fault slip velocities could enhance our understanding of the predictability of large earthquakes.

Methods

Experimental settings

We briefly summarize the laboratory experiment settings used to measure acoustic emissions and mechanical data³¹. Acoustic emissions were recorded using 64 shear-mode piezoelectric transducers with a response frequency of 500 kHz. These sensors were attached to both long sides of the lower sample, with 32 sensors on each side. The signals were amplified 20 times and continuously sampled at 10 MHz. The shear load was measured using a load cell. For further details, refer to the original experimental studies^{30,31}.

Event catalog in the meter-scale laboratory experiment

We utilize a catalog of laboratory quake cycles containing foreshock activity³¹. This catalog is generated using the STA/LTA detection technique combined with a grid search for spatial (x) and temporal (t) information, and the ball drop calibration technique⁵⁷ to determine

moment magnitude (M_w). The seismicity shows a non-patterned spatial distribution (Fig. S1a, 1b) and follows a tapered Gutenberg-Richter magnitude-frequency relation^{31,58}.

The published catalog does not include detailed laboratory quake information beyond their origin times, due to challenges in precisely estimating magnitudes and hypocenters. However, our forecasting method requires both quake and foreshock data. To address this, we roughly estimate the quake information (Figure S1c, blue histogram) and validate these extrapolations post-prediction (see Methods for details).

Seismic moment is estimated according to $M = \mu DA$, where μ is the shear modulus, D is the co-seismic slip amount, and A is the co-seismic slip area. The rigidity of the host rock in the laboratory experiment is $\mu = 41.2 \text{ GPa}$ ³⁰. Assuming the entire fault co-seismically slips during laboratory quakes, we estimate A as $A = 0.1 \times 1.5 = 0.15 \text{ m}^2$. We further assume D is uniform across the fault, estimating it at $0.1\text{--}0.2 \text{ mm}$ based on displacement data. The estimated seismic moment is thus $M = 0.62\text{--}1.24 \times 10^6 \text{ N m}$, corresponding to moment magnitudes of $M_w = -2.2\text{--}2.0$. To account for potential non-uniform slip, we introduce a random error of 10% to M .

The effective rigidity during co-seismic slip might be lower than that of the host rock, as the simulated fault surface does not fully arrest the entire rupture. As a result, the estimated seismic moment could carry an error of roughly an order of magnitude. Therefore, we use $M_w = -2.0 + \text{error}$ as the moment magnitude of the quakes and validate the prediction results within the range $M_w = -3.0$ to -2.0 for quake moment magnitude. Ultimately, the extrapolation does not significantly impact the prediction score, as shown in Fig. S3; the low moment magnitude of the laboratory quakes does not meaningfully reduce the prediction accuracy.

Preparation of nominal shear stress data

We use nominal shear force data measured on the reaction force bar (Fig. 1a), provided by F. Yamashita. The shear force is converted to shear stress and subsequently detrended (only the training set is used to determine the trend). To account for the contribution of shear force to fault shear stress, we divide the shear force by the fault area of 0.15 m^2 to obtain the nominal shear stress. After the initial four quake-cycles, the nominal shear stress stabilizes at a consistent level throughout the experiment. However, a slight increasing trend, possibly due to hardening and compaction, is observed, which could potentially impact the performance of the ML and DL models. To mitigate this, we assume that the increasing trend is linear and remove the trend using a linear regression based on the least squares method (Fig. S4). Note that the regression coefficient was calculated using only the training set, to prevent any information leakage into the validation and test sets. The resulting coefficient is: $\tau_{\text{lin}}(t) = 40.3 \times 10^{-4}t + 66.9 \text{ MPa}$, where τ_{lin} is the linear regression line and t is the time in seconds. Our training, validation, and test sets are continuous in time, and we know the time stamp of each data point in the validation and test sets. Thus, the detrending of validation and test sets can be done by subtracting $\tau_{\text{lin}}(t_{\text{valid}}$ or $t_{\text{test}})$ from the raw shear stress data. This process is even valid in real-time data processing if linear regression is a reasonable assumption. We refer to this detrended nominal shear stress as nominal shear stress in the main text. To effectively train the DL model and ensure consistency with the ML model training conditions, the nominal shear stress is normalized using the min-max values from the training set.

Description of random forest algorithm

The Random Forest (RF) algorithm, developed by Breiman (2001)⁵⁹ and implemented in the scikit-learn Python package by Louppe (2013)⁶⁰ and Pedregosa (2011)⁶¹, is a classification algorithm that aggregates the outputs of multiple decision trees. Each decision tree classifies input data based on statistical classification using input features and target labels. At each node, the tree evaluates whether

feature k is above or below the threshold t_k . If the feature value is above/below the threshold, the data proceeds to the next left/right node, and each subsequent node repeats this process. The threshold at each split is determined by the Classification and Regression Tree (CART) algorithm, which aims to divide the current node into the two purest possible next nodes by minimizing the cost function J ,

$$J(k, t_k)_{s,j} = \frac{m_{j,\text{left}}}{m_j} \text{MSE}_{j,\text{left}} + \frac{m_{j,\text{right}}}{m_j} \text{MSE}_{j,\text{right}} \quad (1)$$

$$\text{MSE}_{j,\text{node}} = \sum_{i \in \text{node}} (\hat{y}_{j,\text{node}} - y_{j,i})^2, \quad \hat{y}_{j,\text{node}} = \frac{1}{m_{j,\text{node}}} \sum_{i \in \text{node}} y_{j,i} \quad (2)$$

where $J(k, t_k)_{s,j}$ is the cost function when the current node j is separated by split s using feature k at the threshold t_k , $m_{j,\text{node}}$ is the number of data points in the next node, m_j is the number of data points before the split, $\hat{y}_{j,\text{node}}$ is the averaged value in the next node, and $y_{j,i}$ is the label values in the next node. Minimizing the cost function $J(k, t_k)_{s,j}$ corresponds to finding the feature k and threshold t_k such that the data within each split node is as homogeneous as possible, while the two resulting nodes are as heterogeneous as possible. The RF model aggregates the outputs from all the trees to provide the final prediction values. This algorithm directly seeks the relationship between features and labels, making it well-suited for data with non-flat, monotonous distributions in the feature-label space. The conceptual illustration of an RF algorithm is provided in Fig. S14a.

One may notice that R^2 is higher in the training phase than in the testing phase (Figs. 1 and 2). While such a difference is typically indicative of overfitting, it has been well established that overfitting in Random Forest (RF) models does not negatively impact validation or testing scores⁵⁹.

Description of a forecasting method

We employ the network representation⁵³ and the classical ML technique, Random Forest (RF), implemented in scikit-learn^{59–61} as a forecasting method (see details of RF model in “Methods” and Fig. S14a). This approach, developed in a numerical study⁵³, is designed to predict the time remaining before synthetic mainshocks using only catalog information. To compute input features, we define a network as a group of earthquakes (e.g., enclosed by the green line in Fig. S14b) and statistically summarize the catalog information regarding the origin times and seismic moments. Specifically, we use 2 parameters $X_i = (\Delta t_i, M_i)$, where Δt_i is the event interval of each earthquake pair in the current network, M_i is the seismic moment transformed by $M = 10^{3(M_w + 6.067)/2.62}$, and i denotes the i th earthquake in the current network. The network abstracts them by taking the average and variance in the current network as follows:

$$\bar{X}_j = \sum_i^j \frac{X_i}{j}, \quad \text{Var}(X_j) = \frac{1}{j} \sum_i^j (X_i - \bar{X}_j)^2 \quad (3)$$

where $\bar{X}_j$ is the average, Var is the variance, and j is the number of earthquakes in the current network, which we define as the network size. The network size j is determined as percentage of the average foreshock number $j = pf$, where p is the percentage, and f is the average foreshock number in the training set. The number of networks and percentage p are optimized by cross-validation as well as hyperparameters in the RF model (see details in “Methods” and Table S2). Using the multiple networks corresponds to incorporating the temporal evolution of seismicity, which is abstracted by each network (Fig. S14c). Therefore, the network representation conceptually corresponds to adaptive time windows with no fixed duration, but with a duration that changes depending on the stage of quake cycles. For example, the notation $\bar{M}_{5f}$ refers to the seismic moment averaged over a network of

5% of the total number of foreshocks over the quake cycles. We use log-normalized values to effectively train the ML and DL models, although the RF model does not require normalization.

For labels, we compute the time remaining before the quake for each foreshock's origin time. During the training phase, the computed features are associated with the time to quake of the most recent foreshock. If the latest event is a quake, we calculate the time to the next quake. Additionally, we attempt to predict the nominal shear stress observed by the reaction force bar. Given the differences in the orders of magnitude between the labels, we use the log-normalized value of time to quake and the normalized linear value of shear stress for training and testing. Consequently, the trained ML model outputs predictions of either the time to quake or the shear stress value by considering the current and past catalog information extracted from multiple networks.

Optimization of hyperparameters of the random forest model, network sizes, and the number of networks

We first optimize the network sizes, number of networks, and hyperparameters of the Random Forest (RF) model using mini-batch training and validation to predict the time to laboratory quakes. Hyperparameter tuning is performed using a simple grid search. Subsequently, we fix the network sizes and the number of networks to ensure consistent feature extraction and then optimize hyperparameters for predicting nominal shear stress and synthetic variables. The training batch size is set to 100, while the validation batch size is 25. This results in 6 and 4 batches for the training and validation sets, respectively. Consequently, the model minimizes the loss across 24 different pairs of training and validation sets. The loss function is the mean squared error (MSE) of either the logarithmic time to quake or the linear shear stress data. The optimized features include the average and variance of Δt_i and M_i within networks of $p = (100, 70, 50, 20, \text{inst})$, where inst refers to a network containing only two earthquakes ($j=2$), capturing the instantaneous behavior. Thus, we use 20 features for each prediction task (average and variance of $(\Delta t_i, M_i) \times$ across 5 networks). The optimized hyperparameters of the RF model for each prediction task are listed in Table S2.

Additionally, we derive feature importance from the trained RF model based on the reduction of the cost function J (Eq. (1)), as shown in Fig. S15. However, since each network uses overlapping data points to compute averages and variances, features from different networks are correlated. As a result, the absolute value of the importance is less meaningful and primarily serves as an index indicating which features are most frequently referenced in the current model. To visualize the general trend of input data, we only present the features from the 20% network for each prediction target.

Evaluation of the model performance

We quantitatively evaluate the prediction score using the coefficient of determination, calculated as $\log R^2$ and $\text{lin} R^2$, following the methodology outlined in previous studies^{11,53}:

$$\log R^2 = 1 - \frac{\sum_{i=1}^n (\log_{10} y_i - \log_{10} \hat{y}_i)^2}{\sum_{i=1}^n (\log_{10} y_i - \log_{10} \bar{y})^2} \quad (4)$$

$$\text{lin} R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where n is the number of data points, y_i is the true target values, $\hat{y}_i$ is the prediction, $\bar{y}$ is the average of y_i . We use logarithmic labels to equalize the contribution of values across different orders of magnitude in predicting the time to lab quake (Eq. (4)). Since the shear stress does not evolve logarithmically, we apply the typical definition of R^2 (Eq. (5)).

Deep learning algorithms and their optimized architectures

In order to compare the performance of the Random Forest (RF) algorithm to deep learning (DL), which has also recently been used to predict laboratory earthquakes^{18,21}, we trained two classic DL architectures (a multilayer perceptron and a recurrent neural network) to predict the time to laboratory quakes and nominal shear stress. For the multilayer perceptron (referred to as DNN here), the input dimension consists of 20 features derived from the network representation, while the output dimension is 1, corresponding to either the time to a laboratory quake or the nominal shear stress at a given time. We apply batch normalization⁶³, Xavier initialization⁶⁴, and dropout regularization⁶⁵. Mish activation⁶⁶ is used as the nonlinear activation function. The model architecture, the number of neurons (n), and dropout probability are determined using a mini-batch training-validation process, similar to that used for training the RF model. For optimization, we employ AMSgrad⁶⁷ with a learning rate of $l_r = 0.001$ and hyperparameters $(\beta_1, \beta_2) = (0.9, 0.999)$. The mean squared error (MSE) loss is minimized during gradient descent. The learning process is terminated early when minimum validation loss is not updated for 100 epochs. The optimized architecture is shown in Fig. S16a.

For the recurrent neural network, we train a Long Short-Term Memory (LSTM)^{68,69} architecture to test the robustness and model independence of our results. Unlike traditional neural networks, LSTM replaces standard neurons with memory blocks that inherently capture the history of input sequences. Since LSTM can capture the temporal evolution of features, we use only 1 network (20% network, 4 features) to characterize seismicity. We apply orthogonal weight initialization⁷⁰, Xavier initialization⁶⁴, and dropout regularization. The activation function used is $\tanh(x)$. Optimization follows the same procedure as for the DNN, except that training is terminated early if the minimum validation loss is not updated for 50 epochs. The optimized architecture is shown in Fig. S16b.

The DL models yield comparable accuracy to RF models in both predicting time to laboratory quake and nominal shear stress, although unlike the RF algorithm, DL requires careful consideration of potential overfitting (avoided here as shown by similar training and testing performance). The current model is optimized for logarithmic time to laboratory quakes, prioritizing this aspect over accuracy on a linear scale.

Inter-event time model

We construct a simple inter-event time model that forecasts the timing of mainshocks based solely on the average recurrence intervals of laboratory quakes^{11,19}. At the occurrence of a laboratory earthquake, the model outputs the average recurrence interval from the training set, with the predicted time to the next event decreasing linearly. Hence, it does not utilize any foreshock information and is only a countdown from the previous main shock. Fig. S2 presents the predictions of the inter-event time model. The linear-scale accuracy is $\text{lin} R^2 = 0.73$, while the log-scale accuracy is $\log R^2 = -0.28$. When the true recurrence interval exceeds the average recurrence, the model predicts a negative value; for $\log R^2$ calculations, these values are set to 10^{-3} seconds.

Set up of a two-dimensional, fully dynamic earthquake-cycle model to replicate the laboratory data

To examine the physical mechanisms underlying precursory behavior, we develop a two-dimensional fully dynamic earthquake-cycle model to replicate laboratory quake behavior (Fig. 4a, b). The model setup is similar to those in Ito & Kaneko (2023)⁵⁴ and Norisugi et al. (2024)⁵³. The fault is embedded in an elastic continuum and subjected to tectonic loading from both edges. To prevent rupture initiation exclusively from the fault's edges, we also apply a time-independent stressing rate (Fig. 4a).

The numerical approach is based on a boundary integral method^{71,72}, adapted for a two-dimensional (in-plane, Mode II) fully dynamic model of earthquake cycles. This dynamic approach allows for the realistic simulation of dynamic ruptures, crucial for understanding earthquake rupture arrest and the resulting earthquake sizes^{54,71}. The fault length is 2400 mm, with 1200 mm regions at both ends subjected to the imposed loading rate. The entire fault is divided into 16,384 cells, each 0.15 mm in size. The fault's constitutive behavior is governed by rate-and-state friction laws with the aging form of state variable evolution^{73–75}:

$$\tau = \sigma \left[f_0 + a \ln \left(\frac{V}{V_0} \right) + b \ln \left(\frac{V_0 \theta}{D_{RS}} \right) \right], \quad \frac{d\theta}{dt} = 1 - \frac{V\theta}{D_{RS}} \quad (6)$$

where τ is the shear strength, σ is the effective normal stress, a and b are the rate-and-state constitutive parameters, V is the slip rate, f_0 is the friction coefficient at $V = V_0$, θ is the state variable, and D_{RS} is the characteristic slip distance for state variable evolution. The parameter $a - b$ primarily controls the fault's slip behavior, with positive and negative values corresponding to velocity-strengthening (VS) and velocity-weakening (VW) patches, respectively. We introduce frictional heterogeneity on the fault with alternating VS and VW patches, where $a_{VS} - b_{VS} = 0.0025$ and $a_{VW} - b_{VW} = -0.0030$, and lengths of $L_{VS} = 12$ mm and $L_{VW} = 30$ mm. The 130-mm VS regions at both ends of the fault act as permanent rupture barriers. The fault contains 50 VS patches and 51 VW patches. A uniform value of $D_{RS} = 1.5$ mm is used to generate tiny foreshocks. The parameters for this model are listed in Table S1.

Under slow loading, stick-slip frictional instability develops only in the VW region, where the instability exceeds the critical nucleation length. The theoretical estimation of the nucleation length relevant to the present simulation is given by⁷⁶:

$$h_{RA}^* = \frac{2}{\pi} \frac{\mu' b_{VW} D_{RS}}{\sigma (b_{VW} - a_{VW})^2} \quad (7)$$

where σ is the effective normal stress, $\mu' = \mu/(1 - \nu)$, μ is the shear modulus, ν is Poisson's ratio, and $a_{VW} - b_{VW}$ are the frictional constitutive parameters on the velocity-weakening patch. Given the parameters listed in Table S1, the estimated critical nucleation length is $h_{RA}^* \sim 20$ mm. The actual nucleation length is influenced by the background loading rate and the manner in which the VW patch is loaded^{72,76}, so the actual nucleation size can differ from h_{RA}^* . Within the expected nucleation sizes, we confirm that enough spatial discretization is applied not to produce one-cell instability, which causes numerical artifacts.

We define the origin of a synthetic event as the moment when the fault slip rate exceeds 1 cm/s at any location on the fault, and the end of the event as when the slip rate decelerates below 0.9 cm/s. The slightly lower threshold prevents double-counting of a single event with an oscillating slip rate. The catalog contains the origin time (t), location (x), and seismic moment $M = \mu AD$, where μ is the shear modulus, A is the co-seismic slip area, and D is the co-seismic slip amount. Given that the fault is one-dimensional, we assume that A is the square of the slip length.

When the loading rate from both edges is tuned to match the value used in laboratory experiments ($V_{pl} = 0.01$ mm/s), the model produces an unrealistically fast earthquake sequence with recurrence intervals of less than one second, whereas laboratory experiments yield intervals of 20 seconds or more. Therefore, we use a much slower loading rate of $V_{pl} = 50$ nm/s and set the time-independent stressing rate to $\dot{\tau} = 0.05$ MPa/s, ensuring that the fault is primarily loaded by this stressing rate. This discrepancy may arise due to differences in fault dimensions, which affect the energy provided per unit area.

The synthetic catalog produces complex seismicity patterns (Fig. S7a), with foreshock activity sometimes coherently clustering near the fault edge or spreading toward the center. Locked areas

(without foreshocks) persist similarly to the laboratory situation³¹. However, the event size distribution (Fig. S1c)^{31,38} is not fully reproduced, and the productivity of medium-sized foreshocks ($-5 < M_w < -4$) is insufficient in this simulation (Fig. S7b). This may be due to differences in the model's dimensionality and frictional properties, as the dominant size of foreshocks is restricted by the VW patch size or nucleation size, and the behavior of rupture termination may differ in a two-dimensional fault model. While there are some qualitative differences in the catalog, our focus here is to replicate the fundamental and simplified physics of the laboratory situation rather than achieve a perfect replication. Therefore, we use the synthetic catalog and related physical quantities under the assumption that a similar physical mechanism governs the production of laboratory foreshocks and quakes.

Additionally, we vary the a - b values on the velocity-weakening (VW) patches to examine how different degrees of fault heterogeneity affect ML prediction performance. These additional cases include VW patch $a - b$ values of -0.0020 (Fig. S9), -0.0035 (Fig. S10), and -0.0015 (Fig. S11). Furthermore, to simulate a scenario that produces both partial and full ruptures, we introduce a slightly longer (54 mm) and weaker VS barrier at the fault center, with $(a-b)_{VS} = 0.0017$ (Fig. S12).

Method for Gutenberg-Richter b -value estimation

We follow⁵⁴ and apply the maximum likelihood method to estimate b -values from the synthetic catalog^{77,78}:

$$b\text{-value} = \frac{\log_{10} e}{\bar{M}_w - (M_c - \Delta M_w/2)} \quad (8)$$

where $\bar{M}_w$ is the mean magnitude above M_c , M_c is the magnitude of completeness, and ΔM_w is the bin size. The foreshock catalog is stacked across multiple mainshock cycles within the training set after preprocessing, which involves removing events smaller than the magnitude completeness threshold ($M_c = -4.6$) and events with a time to synthetic quake smaller than 10^{-3} seconds. We set $\Delta M_w = 0.05$. The standard error of the b -value is estimated as follows⁷⁹:

$$\epsilon(b\text{-value}) = 2.30(b\text{-value})^2 \epsilon(\bar{M}_w) \quad (9)$$

where

$$\epsilon(\bar{M}_w)^2 = \sum_{i=1}^n \frac{(M_w^i - \bar{M}_w)^2}{n(n-1)} \quad (10)$$

Data availability

The experimental event catalog is provided by the original study³¹. The shear stress data from the laboratory experiment are available from Futoshi Yamashita (National Research Institute for Earth Science and Disaster Resilience, Japan) upon request. The synthetic data used to reproduce all the results in this study are available in a permanent data repository⁸⁰.

Code availability

The Python scripts used to reproduce all results related to the experiments and numerical simulations are available in the same repository as the data⁸⁰.

References

- Jordan, T. H. et al. Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Ann. Geophys.* **54**, 315–391 (2011).
- Scholz, C. H., The mechanics of earthquakes and faulting. Cambridge University Press <https://doi.org/10.1017/9781316681473> (2019).

3. Nakatani, M. Evaluation of phenomena preceding earthquakes and earthquake predictability. *J. Disaster Res.* **15**, 112–143 (2020).
4. Scholz, C. H., Sykes, L. R. & Aggarwal, Y. P. Earthquake Prediction: A Physical Basis: Rock dilatancy and water diffusion may explain a large class of phenomena precursory to earthquakes. *Science* **181**, 803–810 (1973).
5. Dieterich, J. H. & Kilgore, B. Implications of fault constitutive properties for earthquake prediction. *Proc. Natl Acad. Sci.* **93**, 3787–3794 (1996).
6. Meier, M.-A., Heaton, T. & Clinton, J. Evidence for universal earthquake rupture initiation behavior. *Geophys. Res. Lett.* **43**, 7991–7996 (2016).
7. Ide, S. Frequent observations of identical onsets of large and small earthquakes. *Nature* **573**, 112–116 (2019).
8. Ito, Y. et al. Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku-Oki earthquake. *Tectonophysics* **600**, 14–26 (2013).
9. Ruiz, S. et al. Intense foreshocks and a slow slip event preceded the 2014 Iquique M w 8.1 earthquake. *Science* **345**, 1165–1169 (2014).
10. Dascher-Cousineau, K. & Bürgmann, R. Global subduction slow slip events and associated earthquakes. *Sci. Adv.* **10**, 2191 (2024).
11. Rouet-Leduc, B. et al. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **44**, 9276–9282 (2017).
12. Lubbers, N. et al. Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophys. Res. Lett.* **45**, 13–269 (2018).
13. Corbi, F. et al. Machine learning can predict the timing and size of analog earthquakes. *Geophys. Res. Lett.* **46**, 1303–1311 (2019).
14. Hulbert, C. et al. Similarity of fast and slow earthquakes illuminated by machine learning. *Nat. Geosci.* **12**, 69–74 (2019).
15. Jaspersen, H. et al. Attention network forecasts time-to-failure in laboratory shear experiments. *J. Geophys. Res.: Solid Earth* **126**, 2021–022195 (2021).
16. Johnson, P. A. et al. Laboratory earthquake forecasting: A machine learning competition. *Proc. Natl Acad. Sci.* **118**, 2011362118 (2021).
17. Pu, Y., Chen, J. & Apel, D. B. Deep and confident prediction for a laboratory earthquake. *Neural Comput. Appl.* **33**, 11691–11701 (2021).
18. Shokouhi, P. et al. Deep learning can predict laboratory quakes from active source seismic data. *Geophys. Res. Lett.* **48**, 2021–093187 (2021).
19. Shreedharan, S., Bolton, D. C., Rivière, J. & Marone, C. Machine learning predicts the timing and shear stress evolution of lab earthquakes using active seismic monitoring of fault zone processes. *J. Geophys. Res.: Solid Earth* **126**, 2020–021588 (2021).
20. Wang, K., Johnson, C. W., Bennett, K. C. & Johnson, P. A. Predicting fault slip via transfer learning. *Nat. Commun.* **12**, 7319 (2021).
21. Laurenti, L., Tinti, E., Galasso, F., Franco, L. & Marone, C. Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth Planet. Sci. Lett.* **598**, 117825 (2022).
22. Borate, P. et al. Using a physics-informed neural network and fault zone acoustic monitoring to predict lab earthquakes. *Nat. Commun.* **14**, 3693 (2023).
23. Karimpouli, S. et al. Explainable machine learning for labquake prediction using catalog-driven features. *Earth Planet. Sci. Lett.* **622**, 118383 (2023).
24. Rouet-Leduc, B. et al. Estimating fault friction from seismic signals in the laboratory. *Geophys. Res. Lett.* **45**, 1321–1329 (2018).
25. Kendall, K. & Tabor, D. An ultrasonic study of the area of contact between stationary and sliding surfaces. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **323**, 321–340 (1971).
26. Pyrak-Nolte, L. J., Myer, L. R. & Cook, N. G. Transmission of seismic waves across single natural fractures. *J. Geophys. Res.: Solid Earth* **95**, 8617–8638 (1990).
27. Hedayat, A., Pyrak-Nolte, L. J. & Bobet, A. Precursors to the shear failure of rock discontinuities. *Geophys. Res. Lett.* **41**, 5467–5475 (2014).
28. Kilgore, B., Beeler, N. M., Losos, J. & Oglesby, D. Rock friction under variable normal stress. *J. Geophys. Res.: Solid Earth* **122**, 7042–7075 (2017).
29. Shreedharan, S., Rivière, J., Bhattacharya, P. & Marone, C. Frictional state evolution during normal stress perturbations probed with ultrasonic waves. *J. Geophys. Res.: Solid Earth* **124**, 5469–5491 (2019).
30. Yamashita, F. et al. Scale dependence of rock friction at high work rate. *Nature* **528**, 254–257 (2015).
31. Yamashita, F. et al. Two end-member earthquake preparations illuminated by foreshock activity on a meter-scale laboratory fault. *Nat. Commun.* **12**, 4302 (2021).
32. Scholz, C. H. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bull. Seismol. Soc. Am.* **58**, 399–415 (1968).
33. Ohnaka, M. & Mogi, K. Frequency characteristics of acoustic emission in rocks under uniaxial compression and its relation to the fracturing process to failure. *J. Geophys. Res.: Solid Earth* **87**, 3873–3884 (1982).
34. Main, I. G., Meredith, P. G. & Jones, C. A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics. *Geophys. J. Int.* **96**, 131–138 (1989).
35. W. Goebel, T., Schorlemmer, D., Becker, T., Dresen, G. & Sammis, C. Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophys. Res. Lett.* **40**, 2049–2054 (2013).
36. Johnson, P. A. et al. Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophys. Res. Lett.* **40**, 5627–5631 (2013).
37. Kwiatek, G., Goebel, T. & Dresen, G. Seismic moment tensor and b value variations over successive seismic cycles in laboratory stick-slip experiments. *Geophys. Res. Lett.* **41**, 5838–5846 (2014).
38. McLaskey, G. C. & Lockner, D. A. Preslip and cascade processes initiating laboratory stick slip. *J. Geophys. Res.: Solid Earth* **119**, 6323–6336 (2014).
39. Lei, X., Li, S. & Liu, L. Seismic b-value for foreshock AE events preceding repeated stick-slips of pre-cut faults in granite. *Appl. Sci.* **8**, 2361 (2018).
40. Rivière, J., Lv, Z., Johnson, P. & Marone, C. Evolution of b-value during the seismic cycle: Insights from laboratory experiments on simulated faults. *Earth Planet. Sci. Lett.* **482**, 407–413 (2018).
41. Bolton, D. C., Shreedharan, S., Rivière, J. & Marone, C. Frequency-magnitude statistics of laboratory foreshocks vary with shear velocity, fault slip rate, and shear stress. *J. Geophys. Res.: Solid Earth* **126**, 2021–022175 (2021).
42. Marty, S. et al. Nucleation of laboratory earthquakes: quantitative analysis and scalings. *J. Geophys. Res.: Solid Earth* **128**, 2022–026294 (2023).
43. Goebel, T. H., Schuster, V., Kwiatek, G., Pandey, K. & Dresen, G. A laboratory perspective on accelerating preparatory processes before earthquakes and implications for foreshock detectability. *Nat. Commun.* **15**, 5588 (2024).
44. Jones, L. M. & Molnar, P. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res.: Solid Earth* **84**, 3596–3608 (1979).
45. Papadopoulos, G., Charalampakis, M., Fokaefs, A. & Minadakis, G. Strong foreshock signal preceding the L'Aquila (Italy) earthquake (M w 6.3) of 6 April 2009. *Nat. Hazards Earth Syst. Sci.* **10**, 19–24 (2010).
46. Kato, A. et al. Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake. *Science* **335**, 705–708 (2012).
47. Nanjo, K. et al. Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, earthquake: First results. *Geophys. J. Int.* **191**, 653–658 (2012).
48. Görgün, E. Analysis of the b-values before and after the 23 October 2011 Mw 7.2 Van-Erciş, Turkey earthquake. *Tectonophysics* **603**, 213–221 (2013).

49. Schurr, B. et al. Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake. *Nature* **512**, 299–302 (2014).
50. Gulia, L., Tormann, T., Wiemer, S., Herrmann, M. & Seif, S. Short-term probabilistic earthquake risk assessment considering time-dependent b values. *Geophys. Res. Lett.* **43**, 1100–1108 (2016).
51. Shimbaru, T. & Yoshida, A. The b value in the seismic activity around foci of large crustal earthquakes before and after their occurrence. *Zisin 2nd Collect.* **74**, 77–86 (2021).
52. Simon, J. D., Simons, F. J. & Irving, J. C. A MERMAID miscellany: Seismoacoustic signals beyond the P wave. *Seismol. Res. Lett.* **92**, 3657–3667 (2021).
53. Norisugi, R., Kaneko, Y. & Rouet-Leduc, B. Machine learning predicts earthquakes in the continuum model of a rate-and-state fault with frictional heterogeneities. *Geophys. Res. Lett.* **51**, 2024–108655 (2024).
54. Ito, R. & Kaneko, Y. Physical mechanism for a temporal decrease of the Gutenberg-Richter b -value prior to a large earthquake. *J. Geophys. Res. Solid Earth* **128**, 2023–027413 (2023).
55. Yabe, S. & Ide, S. Variations in precursory slip behavior resulting from frictional heterogeneity. *Prog. Earth Planet. Sci.* **5**, 1–11 (2018).
56. Ide, S. & Beroza, G. C. Slow earthquake scaling reconsidered as a boundary between distinct modes of rupture propagation. *Proc. Natl Acad. Sci.* **120**, 2222102120 (2023).
57. McLaskey, G. C., Lockner, D. A., Kilgore, B. D. & Beeler, N. M. A robust calibration technique for acoustic emission systems based on momentum transfer from a ball drop. *Bull. Seismol. Soc. Am.* **105**, 257–271 (2015).
58. Kagan, Y. Y. Seismic moment distribution revisited: I. Statistical results. *Geophys. J. Int.* **148**, 520–541 (2002).
59. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
60. Louppe, G., Wehenkel, L., Sutter, A., Geurts, P. Understanding variable importances in forests of randomized trees. *Adv. Neural Inf. Process. Syst.* **26** (2013).
61. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
62. Hanks, T. C. & Kanamori, H. A moment magnitude scale. *J. Geophys. Res. Solid Earth* **84**, 2348–2350 (1979).
63. Ioffe, S. & Christian S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning. PMLR.* **37**, 448–456 (2015).
64. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256 (2010). *JMLR Workshop and Conference Proceedings*
65. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
66. Misra, D.: Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681* (2019)
67. Reddi, S.J., Kale, S., Kumar, S.: On the convergence of Adam and beyond. *arXiv preprint arXiv:1904.09237* (2019)
68. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
69. Gers, F. A., Schmidhuber, J. & Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **12**, 2451–2471 (2000).
70. Saxe, A.M., McClelland, J.L., Ganguli, S.: Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv preprint arXiv:1312.6120* (2013).
71. Lapusta, N., Rice, J. R., Ben-Zion, Y. & Zheng, G. Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate-and state-dependent friction. *J. Geophys. Res.: Solid Earth* **105**, 23765–23789 (2000).
72. Kaneko, Y., Nielsen, S. B. & Carpenter, B. M. The onset of laboratory earthquakes explained by nucleating rupture on a rate-and-state fault. *J. Geophys. Res. Solid Earth* **121**, 6071–6091 (2016).
73. Dieterich, J.H. Time-dependent friction and the mechanics of stick-slip. *Rock friction and earthquake prediction*, 790–806 https://doi.org/10.1007/978-3-0348-7182-2_15 (1978).
74. Dieterich, J. H. Modeling of rock friction: 1. Experimental results and constitutive equations. *J. Geophys. Res. Solid Earth* **84**, 2161–2168 (1979).
75. Ruina, A. Slip instability and state variable friction laws. *J. Geophys. Res.: Solid Earth* **88**, 10359–10370 (1983).
76. Rubin, A.M. & Ampuero, J.-P. Earthquake nucleation on (aging) rate and state faults. *J. Geophys. Res.: Solid Earth* **110**, <https://doi.org/10.1029/2005JB003686> (2005).
77. Aki, K. Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bull. Earthq. Res. Inst., Tokyo Univ.* **43**, 237–239 (1965).
78. Utsu, T. A statistical significance test of the difference in b -value between two earthquake groups. *J. Phys. Earth* **14**, 37–40 (1966).
79. Shi, Y. & Bolt, B. A. The standard error of the magnitude-frequency b value. *Bull. Seismol. Soc. Am.* **72**, 1677–1687 (1982).
80. Norisugi, R., Kaneko, Y. & Rouet-Leduc, B. Numerical data for the research paper titled “Machine learning predicts meter-scale laboratory earthquakes” *Zenodo*. <https://doi.org/10.5281/zenodo.14925543> (2025).

Acknowledgements

We thank Futoshi Yamashita for kindly giving us the data of the laboratory experiment and helpful discussion. This work is supported by the JSPS KAKENHI (21H05206, Y.K.). B.R.L. acknowledges support from the JSPS Hakubi program (B.R.L.) and NASA NSPIRES (80NSSC22K1282, B.R.L.). We gratefully acknowledge the computational resources for our numerical simulation on Wisteria/BDEC-01 provided by ERI JURP 2024 -S-B103 and 2024-S-B301 at the Earthquake Research Institute, the University of Tokyo.

Author contributions

Y.K. and R.N. designed the study. R.N. analyzed the experimental data, conducted numerical simulations, and applied the ML/DL methods to the experimental and numerical data. R.N., Y.K. and B.R.L. analyzed and discussed the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64542-4>.

Correspondence and requests for materials should be addressed to Reiju Norisugi.

Peer review information *Nature Communications Materials* thanks Samson Marty, Parisa Shokouhi, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

UNCORRECTED PROOF

Supplementary information for “Machine learning predicts meter-scale laboratory earthquakes”

Reiju Norisugi^{1*}, Yoshihiro Kaneko¹ and Bertrand Rouet-Leduc²

¹Department of Geophysics, Kyoto University, Japan.

²Disaster Prevention Research Institute, Kyoto University, Japan.

*Corresponding author(s). E-mail(s): norisugi.reiju.77e@st.kyoto-u.ac.jp;

Contents of this file

1. Table S1: Physical parameters used in the fully-dynamic earthquake-cycle simulation.
2. Table S2: Optimized hyperparameters of the RF model.
3. Figure S1: Event catalog from the meter-scale rock-friction laboratory experiment.
4. Figure S2: Prediction by the inter-event time model.
5. Figure S3: Effect of the magnitude extrapolation on the prediction score.
6. Figure S4: Detrending process of nominal shear stress.
7. Figure S5: Prediction of time to laboratory quakes by DNN and LSTM.
8. Figure S6: Prediction of nominal shear stresses by DNN and LSTM.
9. Figure S7: Event catalog generated by numerical simulation.
10. Figure S8: The evolution of slip and shear stress before and after synthetic events.
11. Figure S9: Prediction of time to synthetic quakes with weaker VW patches of $(a-b)_{\text{VW}} = -0.0020$.
12. Figure S10: Prediction of time to synthetic quakes with stronger VW patches of $(a-b)_{\text{VW}} = -0.0035$.
13. Figure S11: Prediction of time to synthetic quakes with weaker VW patches of $(a-b)_{\text{VW}} = -0.0015$.
14. Figure S12: Prediction of time to synthetic quakes in the partial rupture case.
15. Figure S13: Effect of the data size on the prediction of time to laboratory quakes.
16. Figure S14: Schematic illustration of RF algorithm.
17. Figure S15: Examples of feature importance when training the RF model.
18. Figure S16: Schematic illustration of DNN and LSTM architectures.

Table S1 Parameters assumed in the numerical model of earthquake cycles presented in the main text.

Parameters		
ρ	Density	2.980 g/cm ³
μ	Rigidity	41.2 GPa
ν	Poisson's ratio	0.25
V_{pl}	Loading rate	50 nm/s
$\dot{\tau}$	Shear stress rate uniformly added to the entire fault	0.050 MPa/s
σ	Effective normal stress	6.7 MPa
f_0	Reference frictional coefficient	0.75
D_{RS}	Characteristic slip distance for slip evolution	1.5 nm
c_p	P-wave speed	6.92 km/s
c_s	S-wave speed	3.63 km/s
a_{VS}	Frictional constitutive parameter on VS patches	0.020
b_{VS}	Frictional constitutive parameter on VS patches	0.0175
a_{VW}	Frictional constitutive parameter on VW patches	0.020
b_{VW}	Frictional constitutive parameter on VW patches	0.023
L_{VS}	Length of VS patches	12 mm
L_{VW}	Length of VW patches	30 mm

Table S2 Hyperparameters of Random Forest optimized for each prediction task.

Tasks	max_depth	min_samples_split	max_features	n_estimators
Time to lab quake	15	10	5	100
Nominal shear stress	10	10	5	100
Time to synthetic quake	20	30	5	200
Average shear stress	20	10	5	100
VS shear stress	20	10	5	100
Case in Figure S9	10	10	5	50
Case in Figure S10	25	30	5	100
Case in Figure S11	5	40	5	200
Case in Figure S12	20	10	5	200

max_depth: the number of splits in a single tree.

min_samples_split: the minimum number of data in a node.

max_features: the number of features used for a single split.

n_estimators: the number of trees in an RF model.

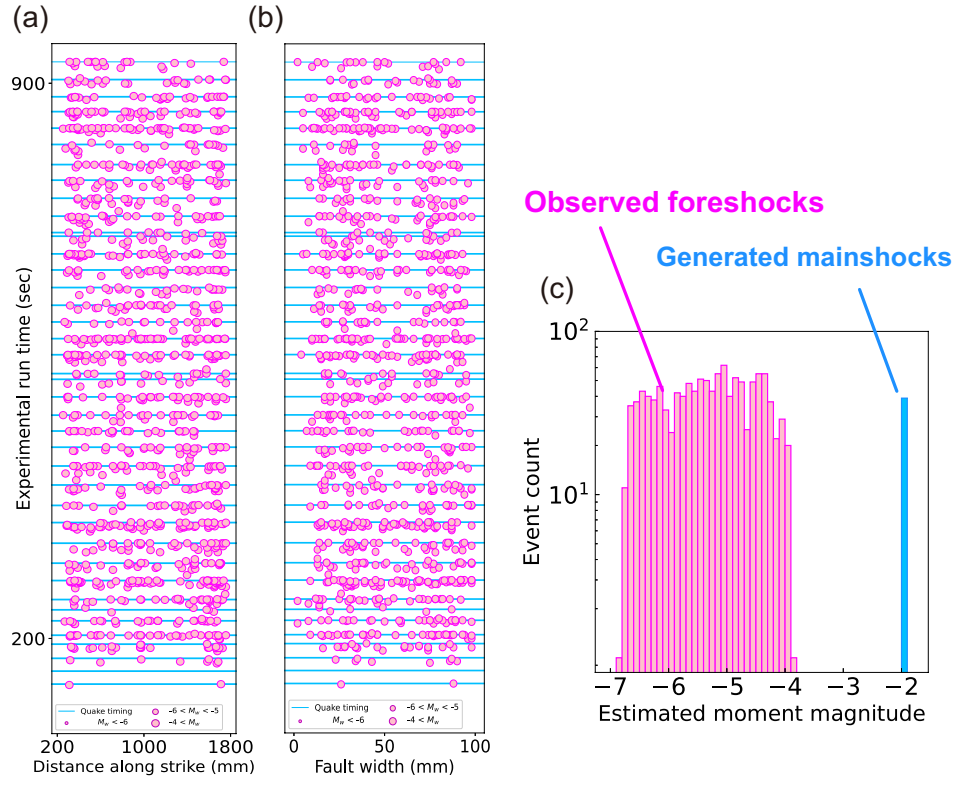


Fig. S1 The catalog data of meter-scale laboratory experiment. (a-b) Spatio-temporal distribution of laboratory earthquakes observed in LB12-011, presented as a time series of seismicity along the fault distance (a) and width direction (b). Blue lines represent the timings of laboratory quakes, while pink circles indicate the occurrences of foreshocks. (c) Magnitude-frequency plot observed in LB12-011. Pink and blue represent foreshocks and laboratory quakes, respectively. Note that the moment magnitudes of the laboratory quakes are estimated with rough approximations and include random errors (see Methods for details).

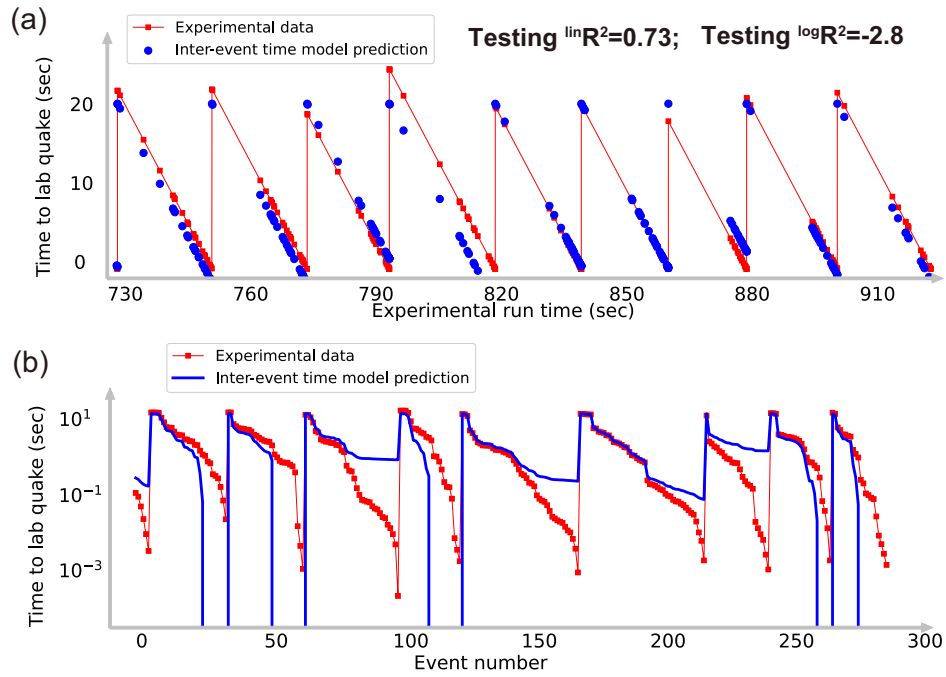


Fig. S2 (a–b) Prediction of time to laboratory quakes by the inter-event time model on timescales of tens of seconds (a) and milliseconds (b). When the prediction score is negative, the $\text{log}R^2$ is computed by setting the value to 10^{-3} seconds. Since the prediction relies solely on the average recurrence interval of laboratory quakes, it does not perform well on the logarithmic scale.

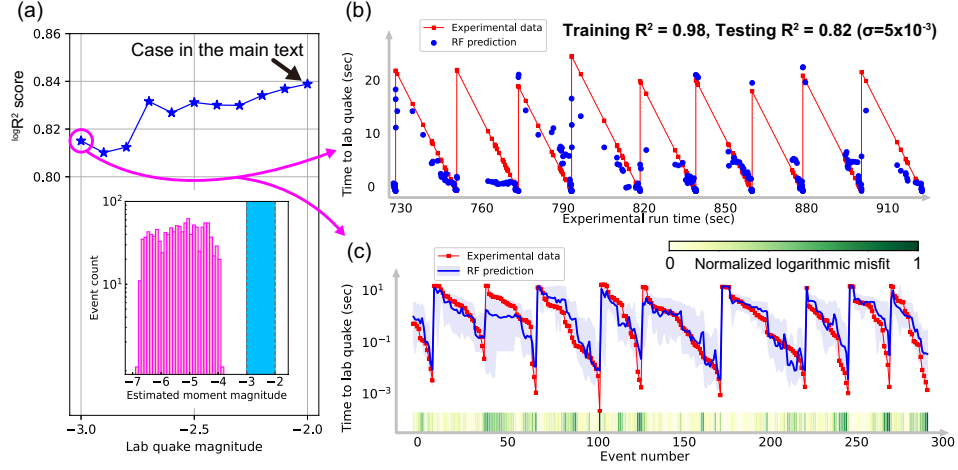


Fig. S3 The effect of assuming different magnitudes of laboratory quakes. (a) Prediction score as a function of the extrapolated moment magnitude for laboratory quakes. The blue histogram indicates the range of moment magnitudes assumed as the quake magnitudes. (b-c) Prediction of time to laboratory quakes in the test set when the quake magnitude is set to $M_w = -3.0 + \text{error}$, the minimum value used in this validation, on a scale of tens of seconds (b) and milliseconds (c). In (c), the blue-shaded area represents the 5 to 95% percentile range of the prediction, and the green color bar indicates the normalized logarithmic misfit. The standard deviation of score is sampled from 100 different Random Forest model randomly trained. While the prediction score slightly decreases with smaller laboratory quake magnitudes, it remains above $\log R^2 = 0.80$, demonstrating that the prediction is satisfactory across timescales from tens of seconds to milliseconds, regardless of the assumed moment magnitudes.

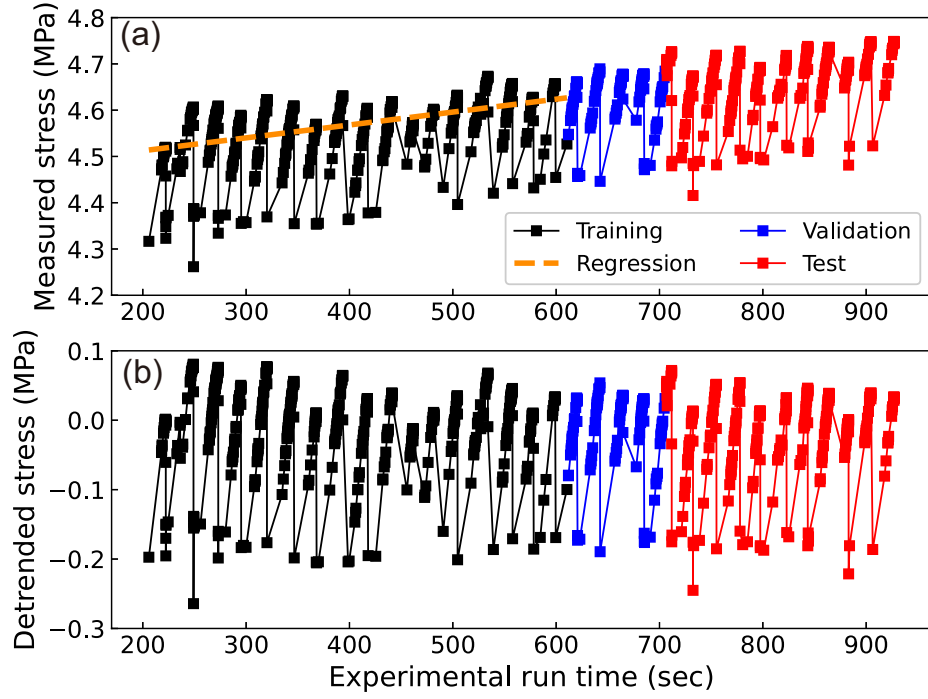


Fig. S4 Nominal shear stress and detrended nominal shear stress in a laboratory experiment. (a) Raw shear stress data before detrending. The black curve represents the nominal shear stress in the training set, while the orange dashed line indicates the linear regression of the training set. The blue and red curves are shear stress data in the validation and test sets, respectively. (b) The nominal shear stress after detrending. The linear trend obtained from the training set is removed from training, validation, and test sets. Shear stress values (represented by square dots) are sampled at the occurrences of foreshocks and laboratory quakes.

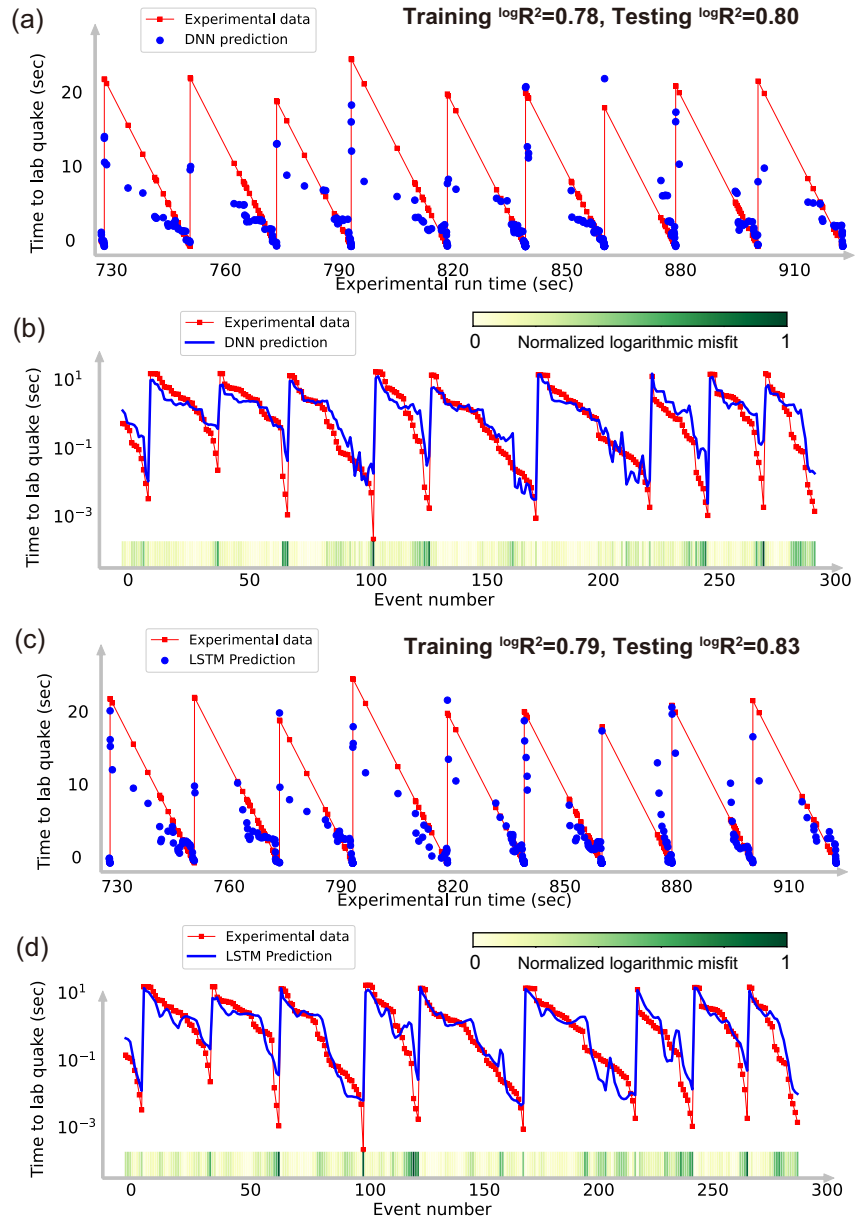


Fig. S5 (a–b) Prediction of time to laboratory quakes using a Deep Neural Network (DNN) on a linear time scale (a) and event number scale (b). (c–d) Prediction using a Long Short-Term Memory (LSTM) on a linear time scale (c) and event number scale (d). The red lines represent the actual time to laboratory quakes, while the blue lines indicate the model predictions. The green color bar denotes the logarithmic misfit. Both models appear to mitigate overfitting, and LSTM exhibits surprising accuracy that testing accuracy surpasses the training accuracy. The LSTM score is comparable to the machine learning model (Figure 1c-1d).

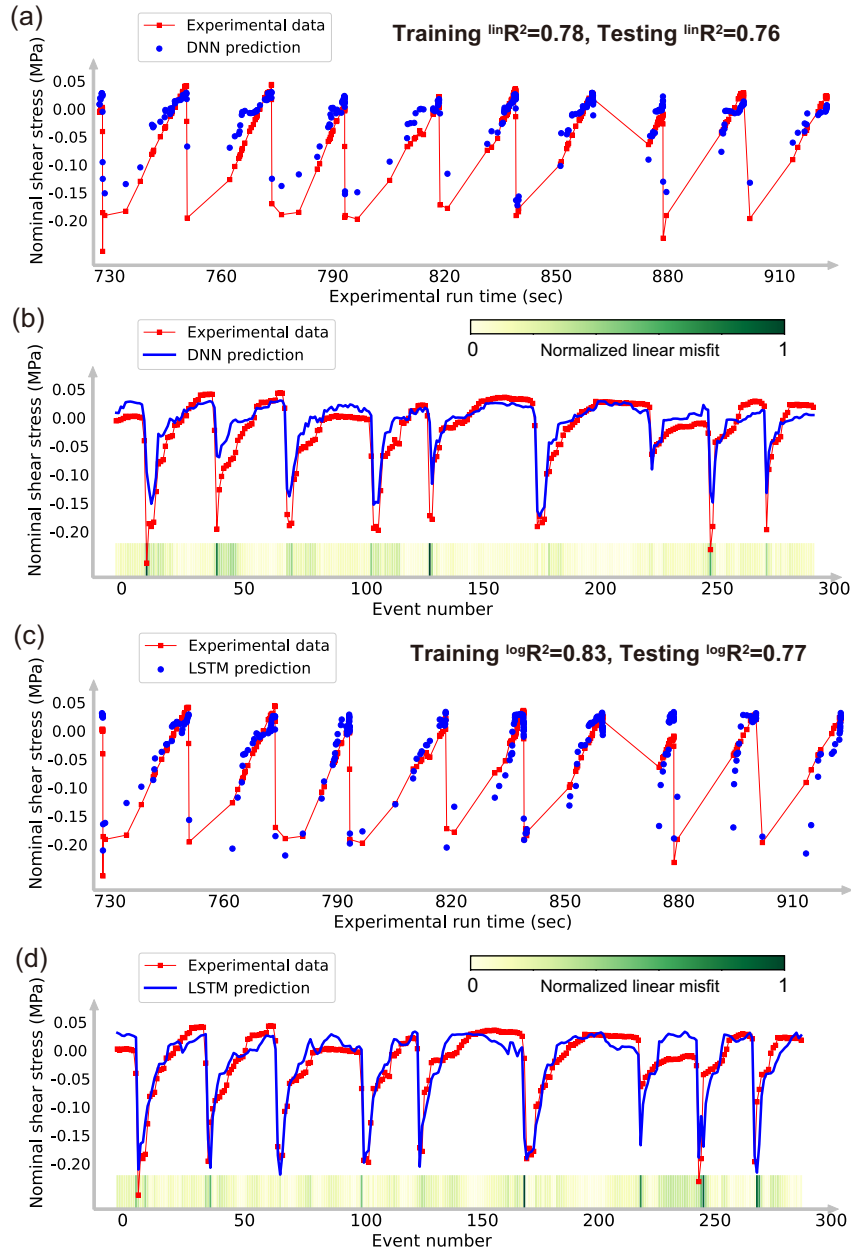


Fig. S6 (a–b) Prediction of nominal shear stress using a Deep Neural Network (DNN) on a linear time scale (a) and event number scale (b). (c–d) Prediction using a Long Short-Term Memory (LSTM) on a linear time scale (c) and event number scale (d). The red lines represent the nominal shear stress, while the blue lines indicate the model predictions. The green color bar denotes the linear misfit. The DNN appears to mitigate overfitting, whereas the LSTM tends to overfit, although the difference of scores in training and test set is relatively small.

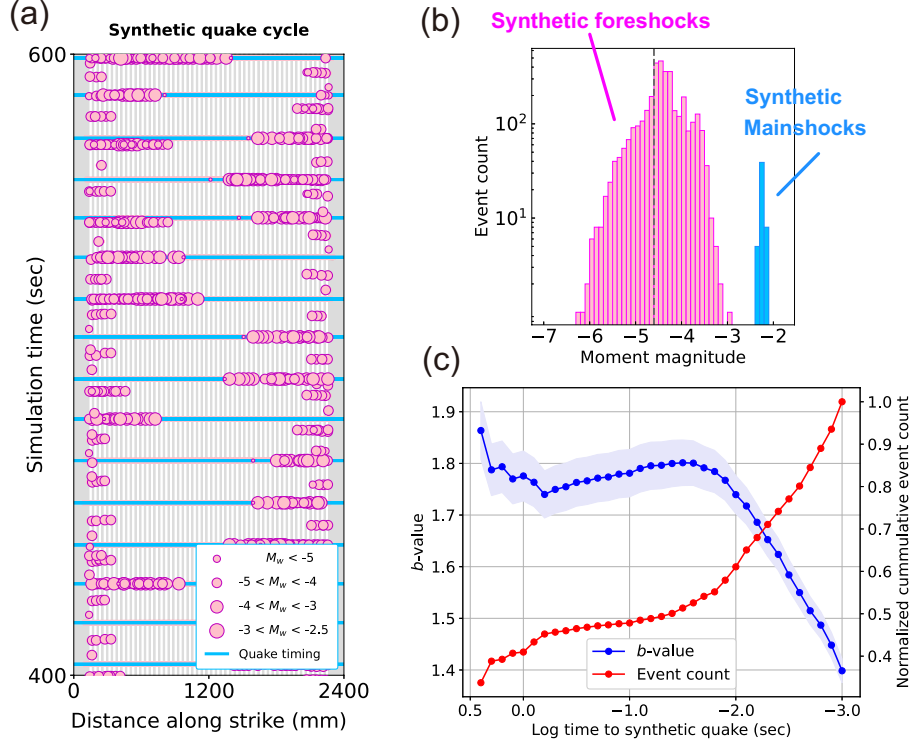


Fig. S7 Synthetic catalog generated by the numerical model, replicating the laboratory experimental data. (a) Spatio-temporal distribution of synthetic quakes along the fault distance. Blue lines indicate the timings of synthetic quakes, while pink circles represent the occurrences of synthetic foreshocks. The background shading shows the fault's characteristics, with gray indicating velocity-strengthening patches and white indicating velocity-weakening patches. (b) Magnitude-frequency plot of the synthetic catalog. Pink and blue represent synthetic foreshocks and quakes, respectively. The histogram bins are set to match those in Figure S1. (c) Estimated b -values (blue line) and cumulative event count (red line) as a function of logarithmic time to synthetic quake. The shaded region represents the estimation error of the b -value (see Methods for details). The estimation is performed at every 0.1 log time step for the cumulative, stacked catalog. The relatively large b -values compared to natural observations may be attributed to the 1-D fault assumption.

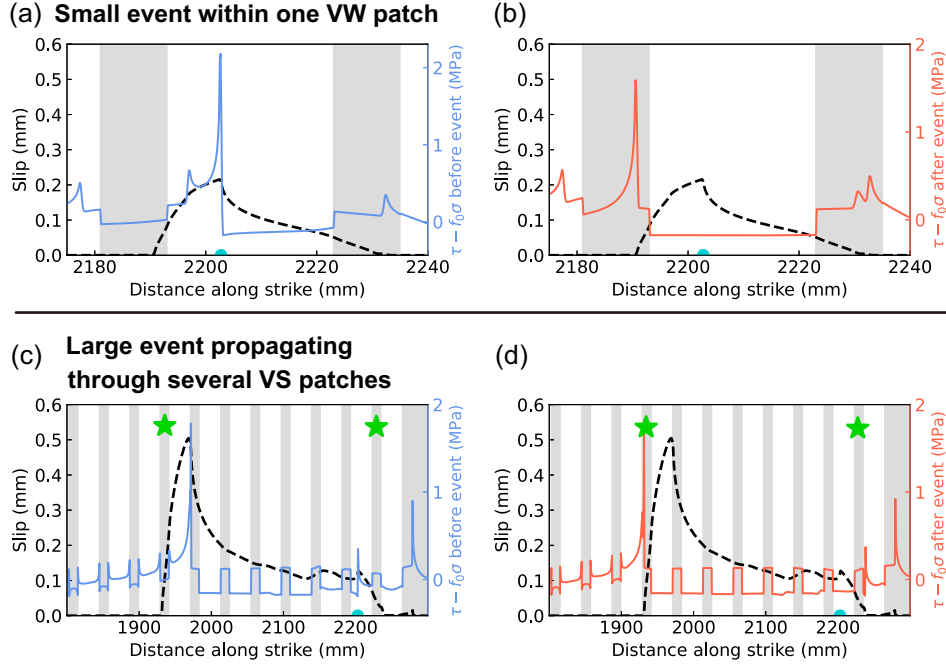


Fig. S8 Coseismic slip and shear stress distributions before and after synthetic events on a one-dimensional fault. The gray/white areas represent velocity-strengthening (VS) and velocity-weakening (VW) patches, respectively. (a-b) Total slip and shear stress evolution before (blue curve) and after (red curve) a small event arrested within the VW patch. (c-d) Total slip and shear stress before (blue curve) and after (red curve) a larger event (not the mainshock) that propagates through several VS patches. The light blue dots at the bottom in (a-d) indicates the hypocenter of the events (i.e., the points that the velocity first exceeds 1 cm/s). The green stars in (c-d) indicate the VS patches that halt the rupture. These rupture-arresting VS patches show significant stress concentration (i.e., slip velocity increases logarithmically). Prior to rupture propagation in (c-d), the patches that allow rupture propagation have slightly higher shear stress levels (i.e., faster creeping velocity) compared to those that stop the rupture (starred patches).

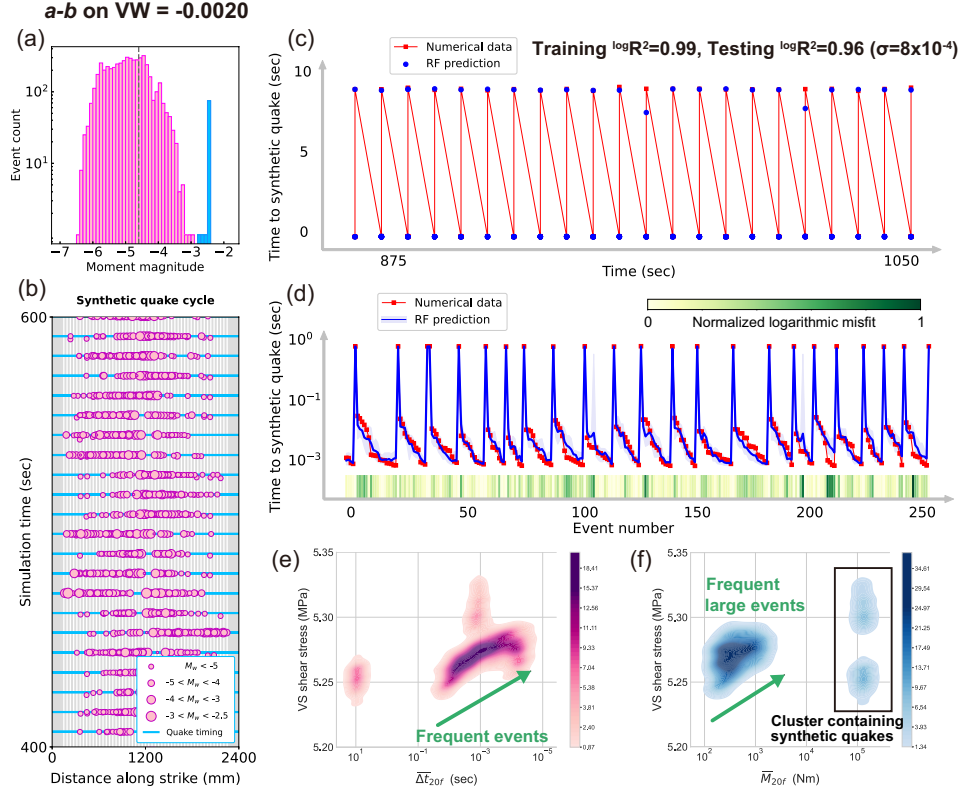


Fig. S9 Results of the Random Forest (RF) prediction for a synthetic catalog generated by the numerical model with weaker velocity-weakening (VW) $(a - b)_{VW} = -0.0020$ patches. (a) Magnitude-frequency relation. Pink bars represent foreshocks, while blue bars indicate defined mainshocks. Small events below the gray line are removed during catalog processing. (b) Space-time plot of synthetic seismicity. Pink dots denote foreshocks, and the blue line marks the timing of synthetic quakes. (c–d) Prediction of time to synthetic quakes at timescales of tens of seconds (c) and milliseconds (d). Foreshocks occur only immediately before quakes, and no dots appear in the linear-scale plot. (e–f) Probability density of velocity-strengthening (VS) shear stress as a function of Δt_{20f} (e) and $\bar{M}_{20f}$ (f). During time periods when the log-scale prediction is accurate, a strong correlation between seismicity and VS shear stress emerges.

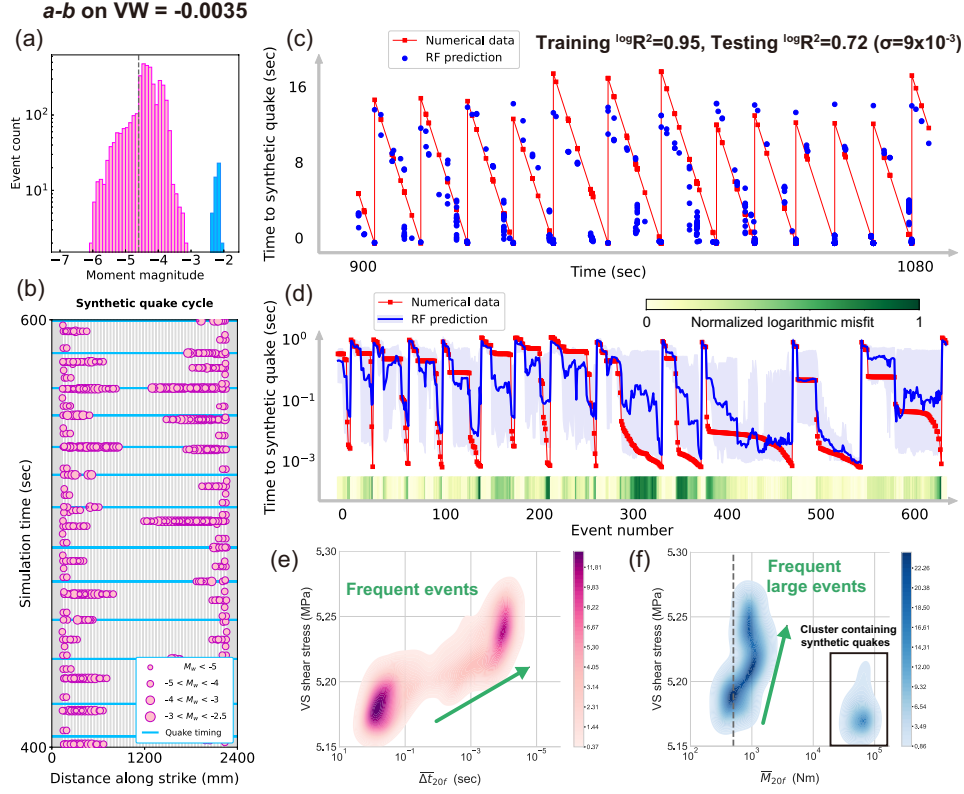


Fig. S10 Results of the Random Forest (RF) prediction for a synthetic catalog generated by the numerical model with stronger velocity-weakening (VW) ($a - b$)_{VW} = -0.0035 patches. (a) Magnitude-frequency relation. Pink bars represent foreshocks, while blue bars indicate defined mainshocks. Small events below the gray line are removed during catalog processing. (b) Space-time plot of synthetic seismicity. Pink dots denote foreshocks, and the blue line marks the timing of synthetic quakes. (c–d) Prediction of time to synthetic quakes at timescales of tens of seconds (c) and milliseconds (d). Compared to the default case shown in Figure 4, the prediction scores deteriorate slightly, likely due to more unstable nature of the fault slip caused by stronger VW patches. (e–f) Probability density of velocity-strengthening (VS) shear stress as a function of Δt_{20f} (e) and $\bar{M}_{20f}$ (f). The correlation between VS shear stress and seismicity characteristics appears slightly weaker than in the default case.

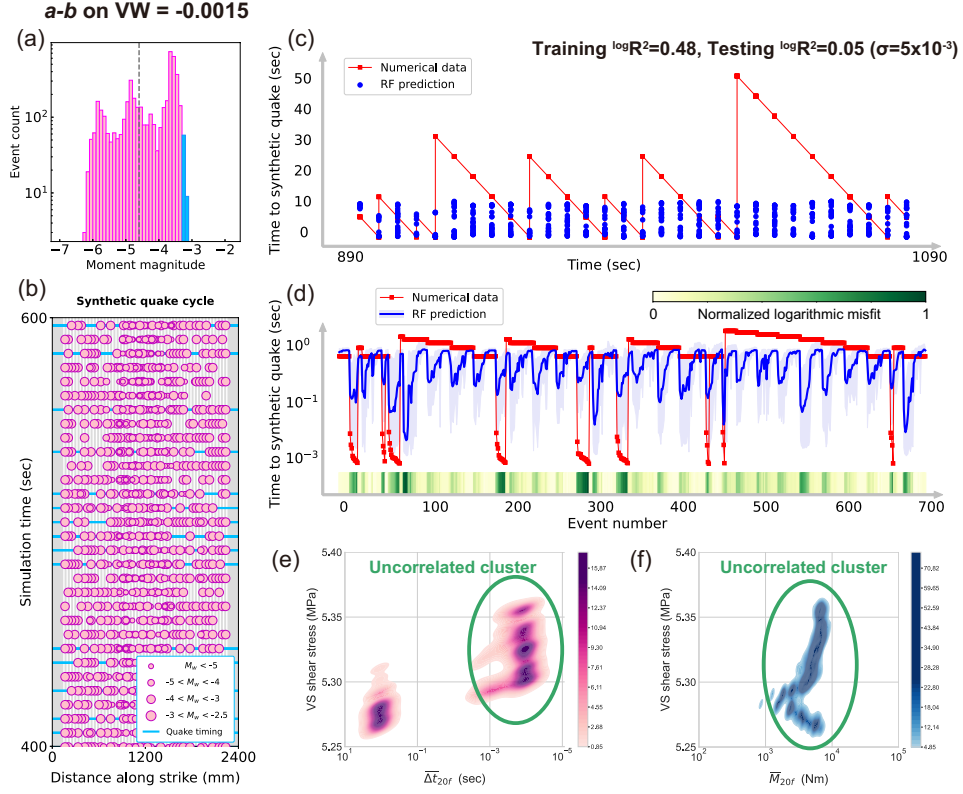


Fig. S11 Results of the Random Forest (RF) prediction for a synthetic catalog generated by the numerical model with weaker velocity-weakening (VW) $(a - b)_{VW} = -0.0015$ patches. (a) Magnitude-frequency relation. Pink bars represent foreshocks, while blue bars indicate defined mainshocks. Small events below the gray line are removed during catalog processing. In this case, distinguishable mainshocks are absent, making the machine learning task ill-defined. (b) Space-time plot of synthetic seismicity. Pink dots denote foreshocks, and the blue line marks the timing of synthetic quakes. (c–d) Prediction of time to synthetic quakes at timescales of tens of seconds (c) and milliseconds (d). Due to the ill-defined nature of the task, predictions perform poorly at both timescales. (e–f) Probability density of velocity-strengthening (VS) shear stress as a function of Δt_{20f} (e) and $\bar{M}_{20f}$ (f). Uncorrelated clusters emerge, indicating weak relationships between seismicity characteristics and VS shear stress.

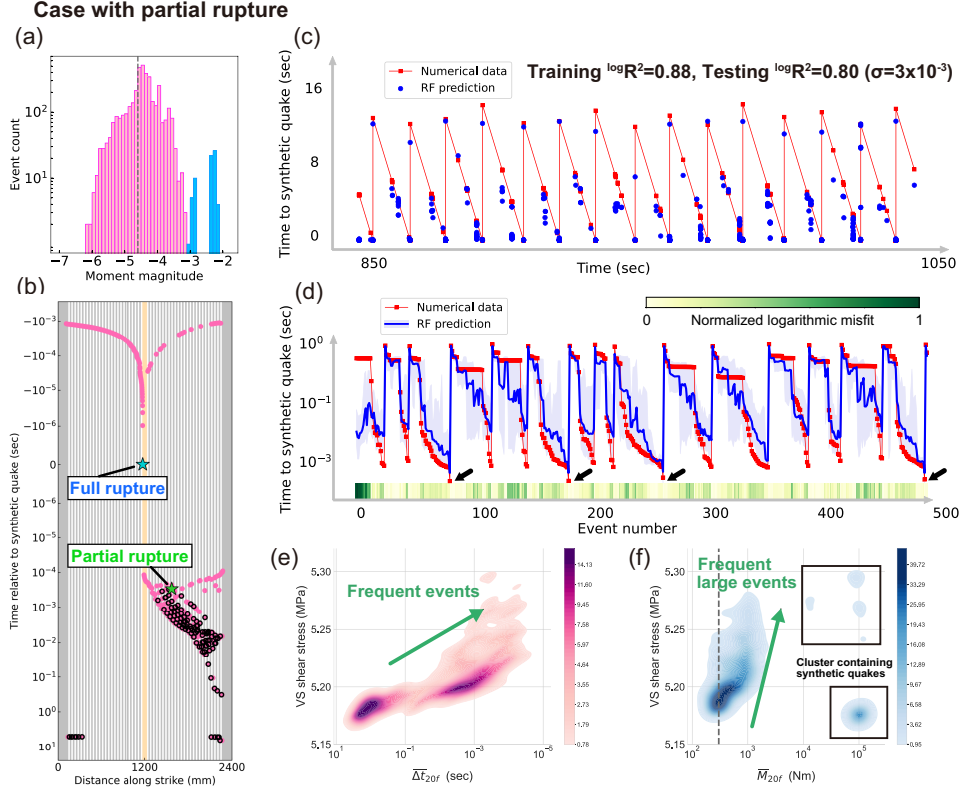


Fig. S12 Results of the Random Forest (RF) prediction for a synthetic catalog generated by the numerical model that produces partial ruptures. (a) Magnitude-frequency relation. Pink bars represent foreshocks, while blue bars indicate defined mainshocks. Small events below the gray line are removed during catalog processing. The peak around $M_w = -3$ corresponds to partial ruptures. (b) Space-time plot of synthetic seismicity preceding both full and partial ruptures. Black dots represent foreshocks, the green star denotes the hypocenter of a partial rupture, and the blue star marks the hypocenter of a full rupture. The light-colored region highlights a slightly long, weak velocity-strengthening (VS) barrier that facilitates partial ruptures. The pink region corresponds to high slip velocity, serving as a proxy for the coseismic rupture area. (c–d) Prediction of time to synthetic quakes at timescales of tens of seconds (c) and milliseconds (d). The black arrows indicate the data points of partial ruptures. Despite the occurrence of partial ruptures, the RF model accurately predicts the timing of both partial and full ruptures. (e–f) Probability density of VS shear stress as a function of Δt_{20f} (e) and $\bar{M}_{20f}$ (f). The correlation between VS shear stress and seismicity characteristics emerges, similar to the default case (Figure 5).

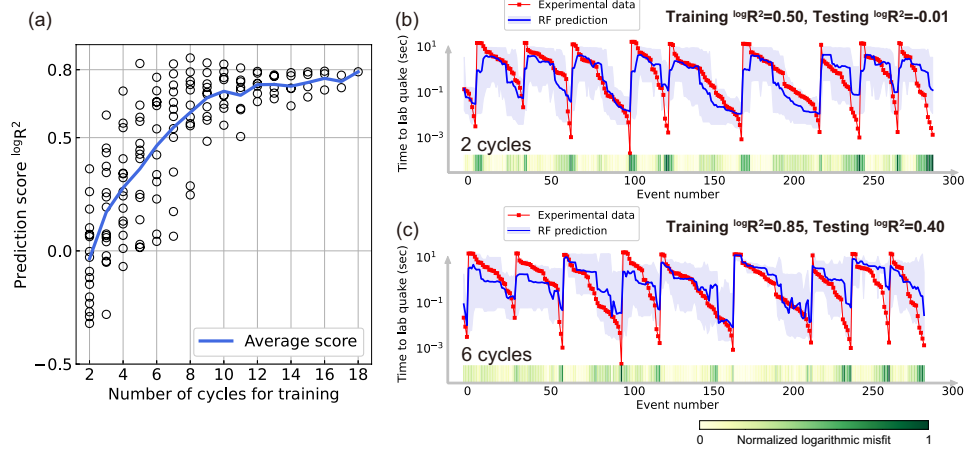


Fig. S13 Effect of training data size on laboratory quake prediction using the Random Forest (RF) model. (a) The $\log R^2$ scores as a function of the number of cycles used for training. Each dot represents a segment of n continuous cycles from the training set. The blue line indicates the average score. (b) Example of prediction results using a 2-cycle training set. (c) Example of prediction results using a 6-cycle training set. The $\log R^2$ score becomes comparable to the default case when the training dataset size exceeds the half of the default case shown in Figure 1.

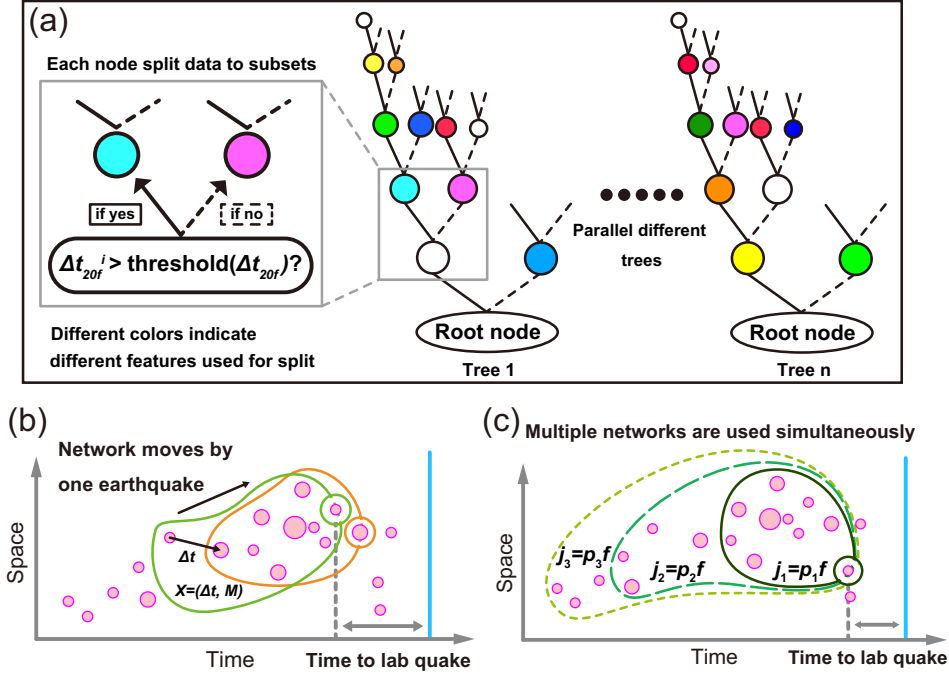


Fig. S14 Conceptual illustrations of (a) the Random Forest algorithm and (b-c) network representation used for forecasting the time to lab quakes or shear stress. In (b-c), pink circles represent foreshock data, while the blue line denotes the timing of laboratory quakes. In (b), the green and orange groups illustrate the current and next network for the individually circled earthquake. Statistically abstracted catalog information is associated with the time to lab quake of the most recent earthquake within the current network (individually circled). In (c), the multiple network representation uses several networks of different sizes (represented by differently styled circles) simultaneously to capture the temporal evolution of earthquake patterns.

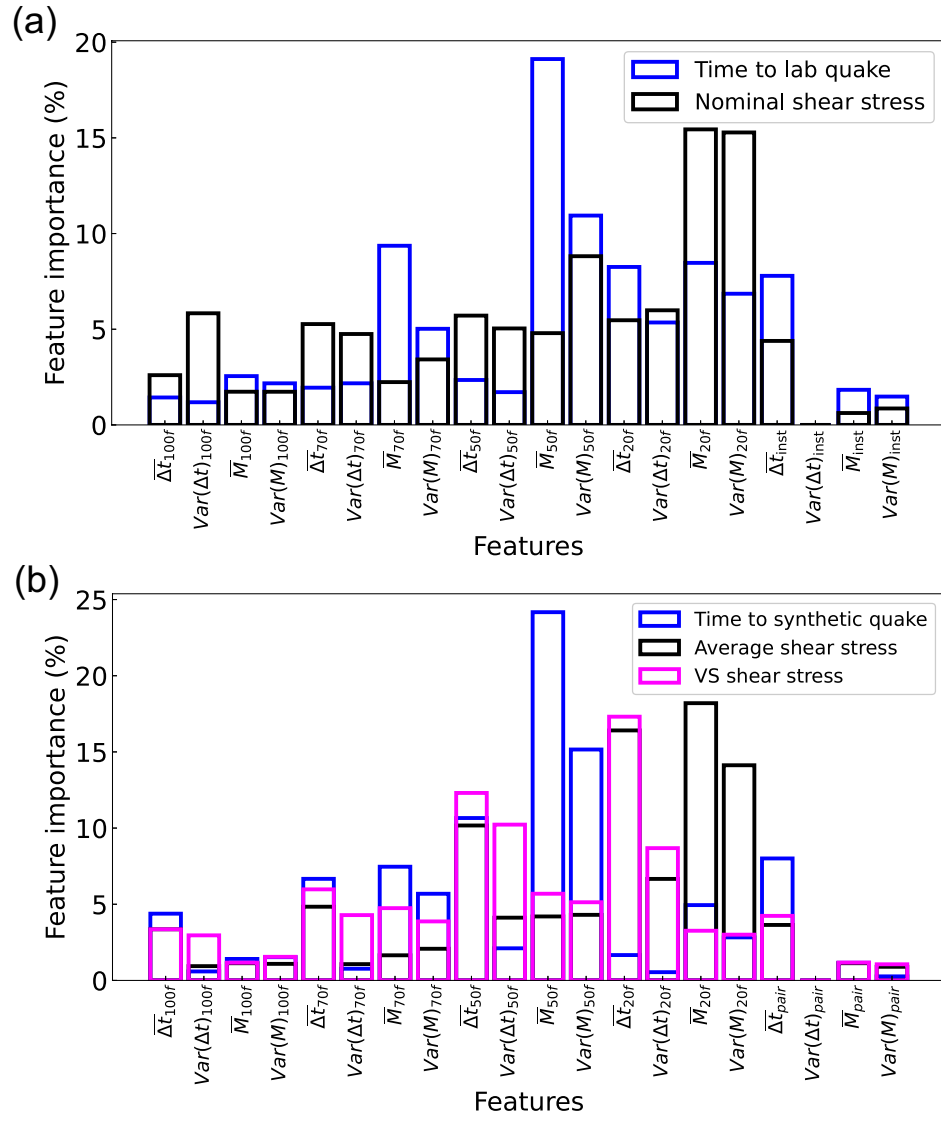


Fig. S15 Feature importance output from the trained Random Forest (RF) model. (a) Laboratory quake predictions. (b) Synthetic quake predictions documented in the main text. Higher importance means that the feature is frequently used to split the input data when making trees in the RF model.

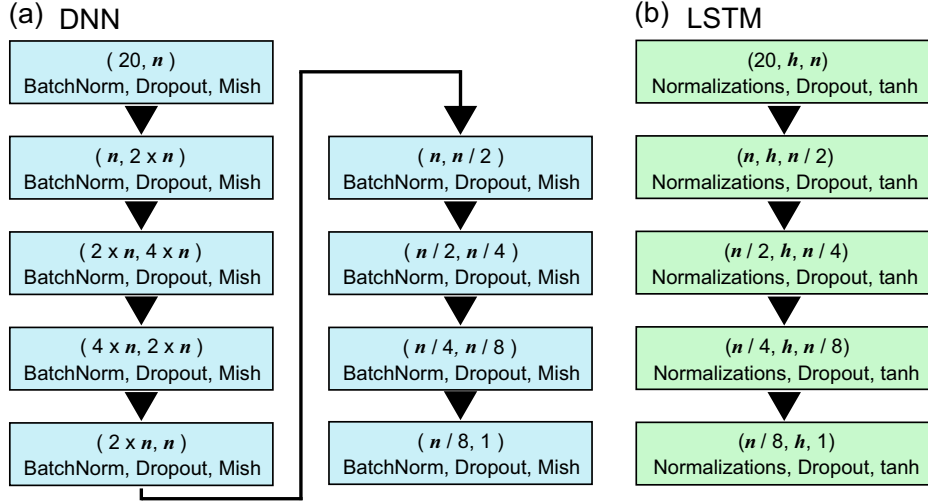


Fig. S16 Conceptual illustration of Deep Learning architectures. (a) The architecture of Deep Neural Network (DNN). Each box represents a single layer, where n denotes the number of neurons. The mean squared error (MSE) loss reaches its minimum when $n = 256$ for predicting time to laboratory quakes and $n = 32$ for nominal shear stress. The dropout probability is set to $p = 0.3$. (b) The architecture of Long Short-Term Memory (LSTM). Each box represents a single LSTM block, where h denotes the history length of input features. The MSE loss reaches its minimum at $h = 5$ and $n = 128$ for predicting both the time to laboratory quakes and nominal shear stress. The dropout probability is set to $p = 0.5$.