nature communications

Article

https://doi.org/10.1038/s41467-024-54153-w

Probing the evolution of fault properties during the seismic cycle with deep learning

Received: 30 May 2024

Accepted: 4 November 2024

Published online: 20 November 2024

Check for updates

Laura Laurenti[1] ✉, Gabriele Paoletti[2], Elisa Tinti[2], Fabio Galasso[3], Cristiano Collettini[2] & Chris Marone[2,4]

We use seismic waves that pass through the hypocentral region of the 2016 M6.5 Norcia earthquake together with Deep Learning (DL) to distinguish between foreshocks, aftershocks and time-to-failure (TTF). Binary and N-class models defined by TTF correctly identify seismograms in test with > 90% accuracy. We use raw seismic records as input to a 7 layer CNN model to perform the classification. Here we show that DL models successfully distinguish seismic waves *pre*/*post* mainshock in accord with lab and theoretical expectations of progressive changes in crack density prior to abrupt change at failure and gradual postseismic recovery. Performance is lower for band-pass filtered seismograms (below 10 Hz) suggesting that DL models learn from the evolution of subtle changes in elastic wave attenuation. Tests to verify that our results indeed provide a proxy for fault properties included DL models trained with the wrong mainshock time and those using seismic waves far from the Norcia mainshock; both show degraded performance. Our results demonstrate that DL models have the potential to track the evolution of fault zone properties during the seismic cycle. If this result is generalizable it could improve earthquake early warning and seismic hazard analysis.

Earthquakes represent one of our greatest natural hazards. Even a modest improvement in forecasting their occurrence could have a major impact on public safety and mitigation of economic loss. Existing work suggests that fault zone properties evolve during the seismic cycle in response to stress changes and microcracking prior to rupture with subsequent post-seismic healing[1,2]. Such changes are observed commonly in lab experiments[3–15] and field data confirm these expectations in some cases, showing changes in elastic wave speed prior to earthquake fault slip, volcanic activity and landslides[16–22]. However, distinguishing subtle changes in seismic behavior or fault properties prior to and after earthquakes, even in locations with dense seismic networks, is challenging[23–31]. Active measurements of fault zone elastic properties prior to earthquake failure, or indeed throughout the seismic cycle, are rarely available and instead techniques have been developed using passive seismic noise[19,32–37]. These techniques are promising but often lack the spatiotemporal resolution required to track the evolution of seismic properties during the seismic cycle or to provide early warning. Lab experiments provide such data[38,39] but are typically conducted on homogeneous samples and simple faults. Earthquakes occur on faults with complex structures consisting of one or multiple fault cores, where most of the slip is localized, surrounded by damage zones formed by widespread fractures and subsidiary small displacement faults. Thus a major challenge in earthquake science is to identify proxies that could serve as early warning or precursory signals to impending catastrophic earthquake rupture.

Recent work in which machine learning (ML) techniques are applied to frictional stick-slip events, the laboratory equivalent to earthquakes, offers a critical opportunity for advancement. In particular these studies show: 1) clear and consistent precursors prior to repetitive earthquake-like failure and 2) that lab earthquakes can be predicted using ML[10,24,40–55]. The lab earthquakes are preceded by a cascade of micro-failure events that radiate elastic energy in a manner

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy. [2]Department of Earth Sciences, Sapienza University of Rome, Rome, Italy. [3]Department of Computer Science, Sapienza University of Rome, Rome, Italy. [4]Department of Geosciences, Pennsylvania State University, Pennsylvania, USA. ✉e-mail: laura.laurenti@uniroma1.it

that foretells impending catastrophic failure. Remarkably, ML and related Deep Learning (DL) techniques predict the lab seismic cycle in detail for hundreds of events, including the evolution of fault zone stress during the lab seismic cycle and clear precursors to failure. These techniques have been applied in many lab settings but few studies have addressed the problem of extending such ML/DL methods to probe fault zone properties directly and illuminate changes during the seismic cycle.

Here, we address this challenge using a technique in which a DL model is trained to distinguish seismic waves prior to and after an earthquake mainshock. We show that the DL method can readily distinguish seismic signals as a function of time to (and since) the 2016 M6.5 Norcia mainshock.

Results

The Amatrice-Visso-Norcia (AVN) 2016-2017 seismic sequence, Central Italy

The AVN seismic sequence began with the Mw 6.0 Amatrice earthquake on August 24th 2016 followed by the Mw 5.9 Visso and Mw 6.5 Norcia earthquakes on October 26th and 30th[56]. These three mainshocks nucleated on a set of SW-dipping normal faults with an along-strike extension of about 80 km[57]. We focus on the Vettore fault, which hosted the Norcia mainshock, because in the two months preceding 30 Oct. the Vettore fault and surrounding crust were illuminated by persistent seismicity (Fig. 1). This seismic activity increased soon after the Visso mainshock and then Norcia aftershock activity occurred for several months. One goal of our work is to distinguish between seismic events before and after the mainshock, so we use the terms 'foreshock' and 'aftershock' to refer specifically to time (*pre* and *post* Norcia mainshock), without causal prejudice.

Earthquake, seismic stations, and waveform selection

DL algorithms work best with large datasets, so we adopted a catalog created using ML[58] with approximately 900,000 events between August 15th, 2016, and August 15th, 2017. We selected all events within 3 km of the Vettore fault segment that ruptured in the Norcia mainshock, using events below 2 km to minimize anthropogenic sources and surface-based seismic noise. Our tailored catalog minimizes the impact of Amatrice aftershocks and includes equal numbers of events over the time period using an asymmetric region from 15 km NE to 4 km SW of the Norcia hypocenter (Fig. 1 and Figs. S1 and S5). In map view (Fig. 1) foreshocks appear to concentrate to the south but in cross section (Fig. S1) it is clear that *pre*/*post* events cover the same region. To ensure waveform quality, we used only Mw > 0.5 events recorded on the active local seismic stations (Table S1 and Fig. 1). We selected these thresholds because they are a good trade-off between having enough traces and being close enough to the fault. We did a variety of sensitivity tests, changing the event selection boundaries and using different stations, and different magnitude thresholds and our results are robust. In particular, we changed: the event boundaries to size 15 km × 10 km × 10 km, with Norcia mainshock located in the center of this box; the selected events have magnitude> 1.0; we found that it did not change our results. Further information on this topic is provided in Section 2.6 which summarizes the use of ambient seismic noise to assess seasonality, differences in spectral content of pre/post co-located events, and null results for stations and earthquakes far from the 2016 Norcia mainshock.

We analyzed 3 component seismic data for nine local stations (Table S1) in a window from 5 seconds before to 20 seconds after the theoretical P-wave arrival (Fig. S3). Dataset construction is provided in the Supplement. We discarded seismic traces with multiple events in catalog, but this did not eliminate all multiple events because the catalog does not contain all small events or events with unusual waveforms (Fig. S3). With these criteria our catalog includes 4694 events *pre* mainshock and 5135 events *post* mainshock (Fig. 1). The full catalog shows a gradual reduction in event rate prior to the Norcia mainshock but for our dataset the event rate is roughly constant in time, with a small decrease in rate prior to the mainshock (Figs. 2 and S2). This small decrease in event rate does not affect performance because the out-of-class split is done in a temporal fashion, by number of events, so the classes are always balanced.

Labels and CNN model

We label all events prior to the mainshock as *pre* and all events after the mainshock as *post* (Figs. 1 and 2). We trained DL models for a range of classifications from 2 classes to multiple classes based on time to (since) failure (TTF) (Figs. 2 and S2). We used a Convolutional Neural Network (CNN)[59] with 7 layers and an increasing number of filters up to a maximum of 256 in the last two convolutional blocks (Fig. S4). A batch normalization layer and ReLU activation layers were used. To enhance the strong activations (feature map) from the convolution output and eliminate the weak ones, we used a max-pooling layer after the convolution operation (Fig. S4). Dilation was applied to all but the first two convolutional blocks, and the dilation rate was increased with network depth. To regularize the network and improve generalizability in test, we also added a dropout layer with a rate of 0.2 after the fully connected layer (Supplement).

CNN model results

Figure 3 shows results for binary classification and for 4-, 8- and 9-classes based on TTF of the Norcia mainshock. We defined classes using variable time intervals that ensured an equal number of events per class. We also tested the case of defining classes based on equal time bins (i.e., for 4 classes: class 1: September; class 2: October; class 3: November; class 4: December). The results are similar in both cases, but the fixed time interval classes are less robust because they are smaller, since we have to remove events in some cases to balance the PRE/POST event numbers.

Another option is to divide the classes based on clustering features, as established in the laboratory work of Karimpouli et al.[54]. For the lab work this approach works well but it is not straightforward to use that approach with field data, both because of the need to balance event numbers, and because we do not have an independent estimate of the fault stress state, so we leave this as a goal for future work.

The train/validation/test splits were always 70/20/10%, but the splits were temporal rather than random to avoid the possibility of model shortcuts in time content (Fig. S5). Because of the temporal nature of our data and the progressive evolution in time, we selected validation and testing data from the middle part of each class. This allows a broader representation of training vs. validation and test data because each of the three is taken from the whole class rather than just one part. We explored several runs for each model to account for minor differences caused by particular seismic traces and to provide uncertainties (Table 1). The performance is measured as $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, where TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative. In each case the model performed well during training and validation (see Supplement for details) and testing was done using data that the model had not seen previously. The use of an Adam optimizer and epoch training produced good results and minimized overfitting by comparing results during training and validation[60]. For binary classification, typical performance values for train, validation, test were > 95%, with a slight reduction in accuracy from train to validation and test (Table 1).

There are ≈ 40 to 80 events per day for both *pre* and *post*. For 2-4-8 class models we did not include events associated in time with the M5.9 Visso earthquake (from October 26th to October 30th 2016) because of the high event rates, but we relaxed that for the 9-class model that included earthquakes from the Visso sequence (Figs. 3 and S2 and Table 1). The magnitude distributions are pretty similar for *pre*/*post* events in our dataset (Fig. 2).

https://doi.org/10.1038/s41467-024-54153-w

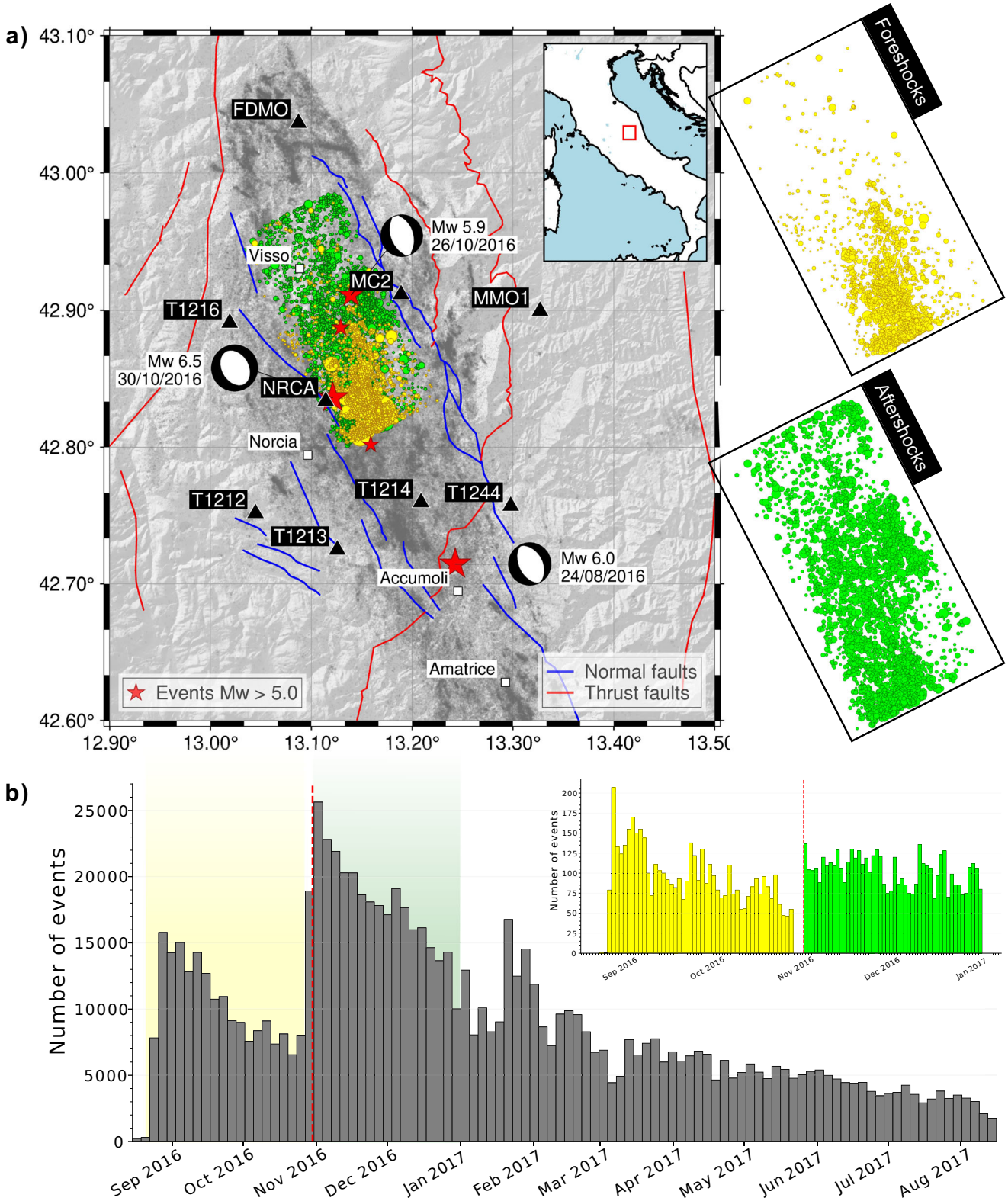

Fig. 1 | Earthquake distribution for the 2016 sequence showing the three main events (red stars and focal mechanisms). a from north to south: Visso Mw 5.9, 26/10/2016, Norcia Mw 6.5, 30/10/2016, and Amatrice Mw 6.0, 24/08/2016). Gray dots show the full catalog of 900,000 events[58]. We used earthquakes in a time window ± 2 months from 30/10/2016 and a rectangular region near the Norcia epicenter and Vettore fault. Yellow symbols show events before the Norcia mainshock, and green symbols show events afterward. Map views at right highlight foreshock and aftershock locations (See Fig. S1 for cross-section). **b** time history for all events in the selected region, and the inset (yellow and green) shows the events used in our DL analysis.

https://doi.org/10.1038/s41467-024-54153-w

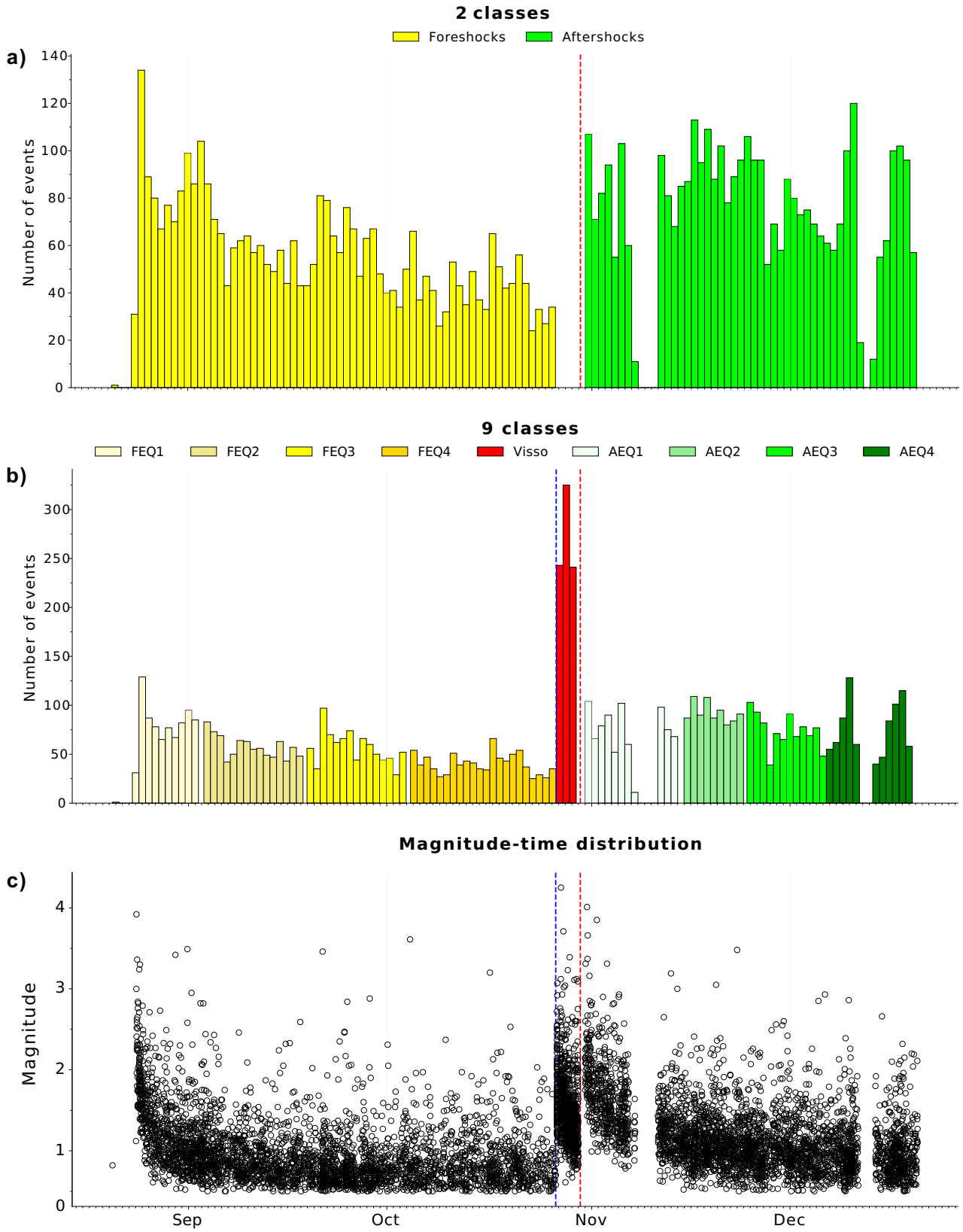

Fig. 2 | Temporal history of events in our dataset and class splits used to classify and label data. Histograms are daily event numbers. **a** shows the data split for 2 classes: *pre* or foreshocks (yellow) and *post* or aftershocks (green). **b** shows additional granularity in time for 9 classes. The 9 class model includes events of the Visso earthquake. In each case, models are trained on time periods with the same number of events in each class. **c** shows the magnitude distribution with time. Displayed events refer to NRCA station.

https://doi.org/10.1038/s41467-024-54153-w

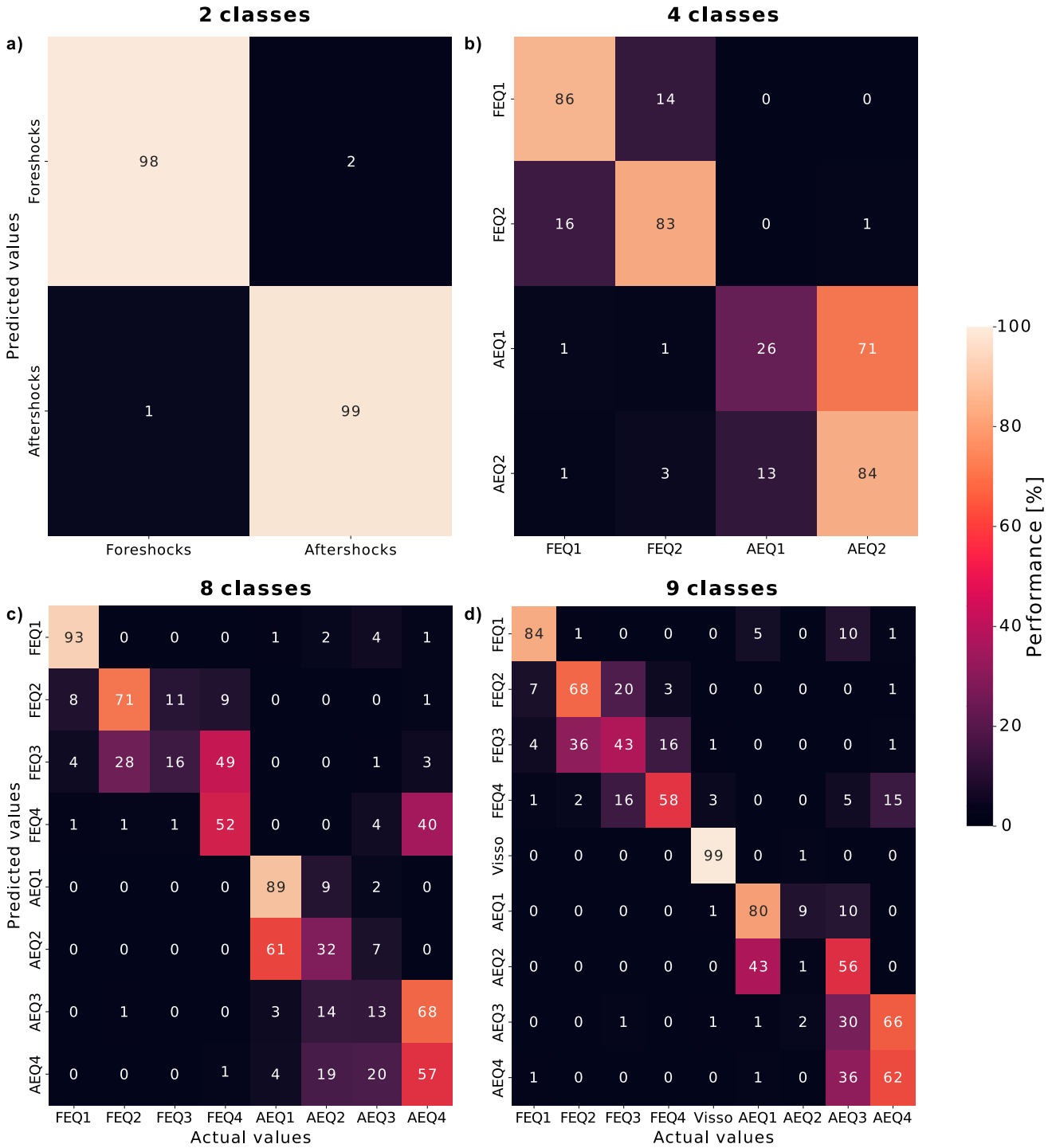

Fig. 3 | Confusion matrices for each class split (see Fig. 2 and Fig. S2 for split details). a is the simplest case of 2 classes: 98% of the *pre* and 99% of the *post* in the test set are classified correctly. **b** shows the results for 4 classes, (**c**) shows the results for 8 classes, and panel **d**) shows the results for 9 classes. Note that *pre* and *post* are well identified in all cases and that DL models can resolve time to failure, distinguishing between events close in time to the mainshock versus those well before/after the mainshock. See Table 1 for data. Displayed events refer to NRCA station. Note that the behavior isn't the same for each run (seed); we are showing random examples, not ad-hoc best cases.

The binary classification results are remarkably good for all 9 stations (Table 1 and Fig. 3), with performance ratios ranging from 79 to 99%. The CNN models are very good at detecting which seismograms come from earthquakes before/after the Norcia mainshock. Station T1213, located south of the Norcia mainshock, shows the best performance for *pre*-events, with 100% of the foreshocks identified correctly. In contrast, only 72% and 74% of the seismic traces for *pre*-Norcia events recorded at Stations MMO1 and T1212 were identified correctly (Table 1). These stations are farther from the mainshock than NRCA, which is essentially at the epicenter (Fig. 1), but not farther than station FDMO, which showed very good performance (Table S1). The model performance does not vary significantly with distance or direction from station to hypocenter (see Fig. S6 and station T1214). We did not study possible differences in source characteristics with subsurface

Table 1 | Performance results for 2, 4 and 9 class models

Station	2 (%)	2 PRE (%)	2 POST (%)	4 (%)	8 (%)	9 (%)
FDMO	98.84 ± 0.22	99.19 ± 0.19	98.49 ± 0.52	68.81 ± 1.86	45.81 ± 3.29	47.05 ± 0.83
MC2	99.16 ± 0.43	98.93 ± 0.25	99.40 ± 0.72	63.90 ± 2.96	55.03 ± 5.96	53.04 ± 2.73
MMO1	84.61 ± 0.69	79.45 ± 2.47	89.80 ± 2.68	54.92 ± 2.46	39.04 ± 2.99	39.33 ± 1.55
NRCA	99.19 ± 0.12	99.16 ± 0.19	99.22 ± 0.07	68.21 ± 4.85	44.69 ± 1.76	57.94 ± 4.42
T1212	77.52 ± 3.37	62.37 ± 7.36	92.71 ± 1.61	69.67 ± 7.22	46.41 ± 3.99	41.02 ± 6.98
T1213	99.69 ± 0.13	99.84 ± 0.13	99.53 ± 0.38	80.57 ± 0.60	67.68 ± 4.31	64.57 ± 1.64
T1214	84.47 ± 7.03	77.62 ± 17.23	91.31 ± 4.43	61.48 ± 5.81	36.76 ± 2.35	40.96 ± 7.64
T1216	92.65 ± 0.83	96.45 ± 0.90	88.85 ± 1.51	59.53 ± 3.40	40.93 ± 2.70	31.83 ± 1.97
T1244	96.83 ± 0.55	98.85 ± 0.42	94.81 ± 0.95	71.25 ± 1.11	49.92 ± 1.61	46.66 ± 1.43

Values are the percentage of correctly classified events to total events. Pre- and post-mainshock values are given for the 2-class model. Fig. 3 gives offers details for each class, for station NRCA. Note that the results for station T1213 are best for each case. The mean and standard deviation reported are computed for 3 separate training, validation and testing runs.

lithology[61] nor variations in performance with back azimuth; both would be interesting future works.

n-class models

The remarkable performance of the DL models in test suggests further study of how seismic traces change with TTF. Table 1 provides results for *n*-class models defined by TTF for the Norcia mainshock (Fig. 2). Details of the time splits and results for all models are given in Figs. S2 and S6 b) and the Supplement (Section 8.2).

The performance results for 4-, 8- and 9-class models are consistent with the 2-class results (Table 1). Keeping in mind that the chance level is 50, 25, 12.5 and 11.11% for 2, 4, 8 and 9-class models, the results are very good for all nine stations. For each data set and split (Fig. S6 b)) we ran three separate models to assess model variability. Performance results in test are significantly above chance levels for each station (Table 1).

***n*-class results, Station NRCA.** We did additional *n*-class studies using data for station NRCA which is the closest to the Norcia mainshock and a few kilometers north-east Norcia city (Fig. 1). The high average accuracy (98.8%) for binary classification shows that DL can accurately identify and classify most of the earthquakes in the test set (Fig. 2). The test set used for binary models contains a total of ~1500 events and only ~1% were classified incorrectly. We evaluated spatiotemporal data for the mislabelled traces, which are not clustered in time or location.

For the 4-class and 8-class models, we followed the same procedure used above and created balanced subdivisions with equal numbers of earthquakes. The results for TTF classification are quite good (Table 1 and Fig. 3). The models can correctly identify whether seismic events occurred well before the Norcia mainshock, in late Aug., early Sep. or after in Nov./Dec. (Fig. 3). Results in the confusion matrix show that most of the misidentified results are off by only one class, such that foreshocks in class 2 (FEQ2) are sometimes placed in class 3 (FEQ3). For example, only 16% of the events in FEQ3 were identified correctly (Fig. 3) but 93% of the events in the FEQ3 test set were placed in FEQ2, FEQ3 or FEQ4. An exception is foreshocks in class 4 (FEQ4), where 40% of them were misclassified as aftershocks in category AEQ4 (Fig. 3). The DL algorithm performance level was 81% for the 4-class test, 57% for the 8-class test, and 62% for the 8-class test for seismograms recorded at station NRCA. We point out that Fig. 3 shows results for one case and there are slight differences between each model run; here and in other figures, we show random examples rather than ad-hoc best cases.

We also trained models with 9-classes in time to include events associated with the October 26th M5.9 Visso mainshock, from October 26th to October 30th 2016 (Fig. 2). This classification is particularly interesting because it includes events that precede Norcia and *post*-date Visso. Figure 2 shows the division of the dataset into 9 classes with the corresponding magnitude distribution for the entire data set of 9829 events. Here again the test performance is very good, with an average of 58% of the events categorized correctly (Table 1, Fig. 3), compared to the chance level of ≈ 11%.

The confusion matrix provides a clear visual representation of the algorithm's performance (Fig. 3). We observe a division similar to the one for 8 classes, with a performance peak corresponding to the Visso events (99%). One interpretation of these results is that the algorithm is able to identify changes in the seismograms associated with changes in fault zone properties between the Visso and Norcia mainshocks.

Origin of model performance

We want to understand why our DL models trained on local data can detect differences in seismograms as a function of TTF for the Norcia mainshock. One hypothesis is that stress changes during the seismic cycle cause micro-fracturing, fault zone damage, and crack sealing. Other possibilities include subtle changes in earthquake source properties or seasonal effects and changes in susceptibility associated with stress and microfracturing.

Mainshock date and seasonality. DL models are well known for detecting shortcuts in inference, so we tested a series of models to assess the possibility that our results are a product of time of the year in addition to time relative to the Norcia mainshock. Seasonal variations are expected for summer vs. winter times or rainy vs. dry conditions and may arise from local fluid content in the upper crust or variations in fracture characteristics with atmospheric temperature[19,29,32,35,62–64]. Previous results suggest that these changes can occur over weeks to months so we began by testing models with small shifts in the mainshock date.

Models trained with the wrong mainshock date perform worse than those trained with the correct date (Fig. 4). Here, we focused on Station NRCA and trained models for binary classification using several dates for the mainshock. To conduct these tests, we retrain the entire model on the same events, but we modify the labels used to define *pre*/*post* using false mainshock dates. In each case we then retested with the correct mainshock date. As before, of the ≈1500 events in the test set, fewer than 1% are misidentified for our standard model (Fig. 3, Table 1). For a model trained with 21 Oct. as the Norcia mainshock date, the overall performance drops only slightly, to 96%, as expected given the total number of foreshocks, but for the test events that occurred between 21 and 30 Oct., the correct mainshock date, the performance drops to 11% (Fig. 4). Of the 43 events in the test set between 21 and 30 Oct., 38 of them are classified 'incorrectly' for 21 Oct. as the mainshock (Table 2). That is, even though events during this period are labeled as aftershocks during training, during testing the model identified 38 of them as foreshocks. The colors in Fig. 4a, highlight the events that are identified correctly/incorrectly based on the fake mainshock date.

https://doi.org/10.1038/s41467-024-54153-w

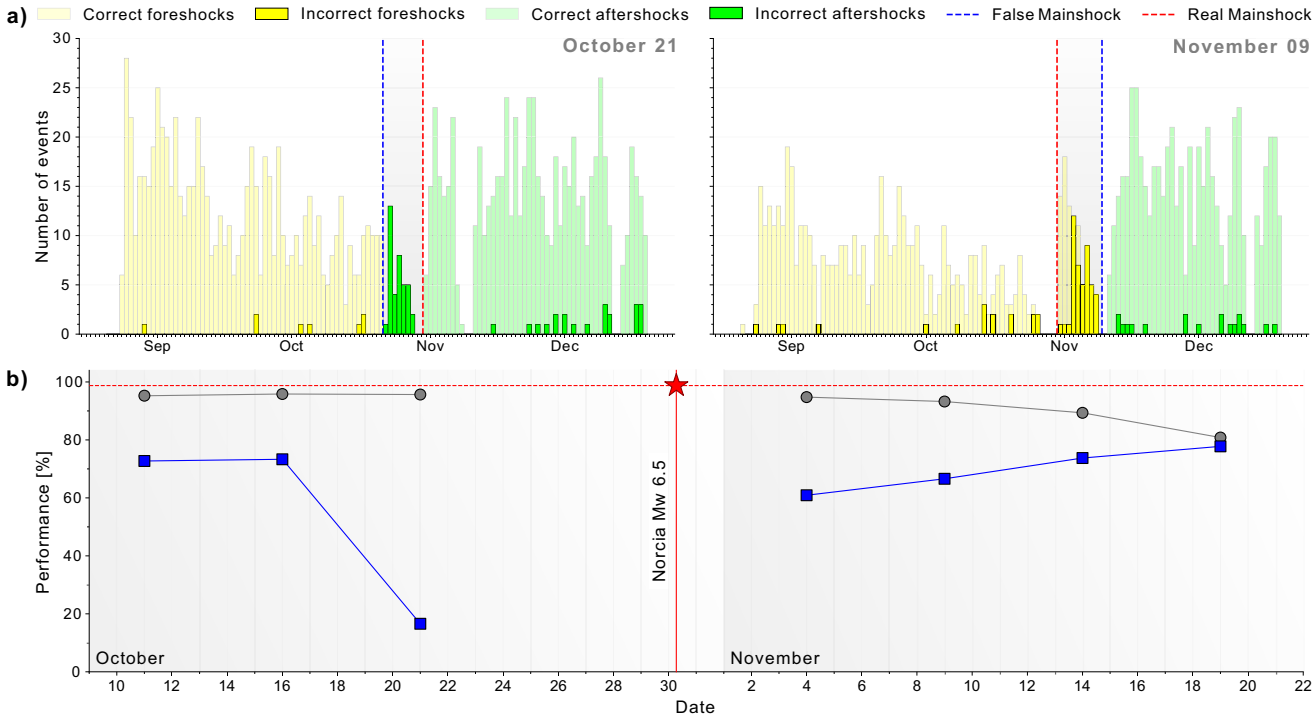

Fig. 4 | Model validation test data for a series of wrong mainshock dates. **a** shows two examples. The shaded regions highlight the time period between the selected false date and the actual mainshock on October 30th. **b** shows the performance for models trained with the correct mainshock date and seven false dates. The dashed red line shows the average performance for the correct mainshock date. Gray dots show average performance for the test set when trained using a model that includes incorrect labels (see Table 2 for details). Note that the average test performance decreases with time away from the mainshock date, as more are labeled incorrectly. Blue symbols show test performance for events during the period where incorrect labels were used in training. Note that the percentage of these events is always lower than the performance using the correct mainshock date and that performance is worse closer to the mainshock time.

Comparing model performance for the whole test set with that for mislabeled events (Table 2) shows that most events are labeled correctly as *pre*/*post* even though the training was done with incorrect labels.

The overall model performance decreases as the time between the fake and actual mainshock date increases (Table 2) but the decrease is more pronounced for aftershocks compared to foreshocks (Fig. 4). This is seen as a drop in the gray circles and a rise in the blue squares with increasing time since mainshock (Fig. 4b). These data suggest that the model is struggling to accurately classify the events in the time period between the fake and real mainshock dates, highlighting hidden patterns common to all the traces of *pre* and *post*, despite the labels used to train the model. These indicate that our results are not simply a function of seasonal variations or time of year.

Table 2 | Results for models trained with data for Station NRCA and a range of incorrect mainshock dates

Date	Performance (%)	NM/NC	FP (%)
Oct 11	95.88	149/42	71.8
Oct 16	95.63	107/38	64.5
Oct 21	95.55	43/38	11.6
Oct 30	98.84	-	-
Nov 04	95.00	64/36	43.7
Nov 09	93.42	121/45	62.8
Nov 14	90.07	171/41	76.0
Nov 19	79.58	261/61	76.6

Models are retrained using the whole data set in each case, with foreshock/aftershock labels based on the date given. Performance is the ratio of correctly classified events to total events for the test set. Column 3 is the number of mislabeled events (NM) and the number of those that are identified with the correct (NC), rather than fake label. Note that the test set has 43 events in the period 21 to 30 Oct. and 38 of them are labeled as foreshocks even though events during this time in the training set were labeled as aftershocks. FP is the test performance during the 'fake' time interval, when training labels were wrong. For the period 21 to 30 Oct. 11.6% of the test events were identified as aftershocks while the others, 88.4%, were classified, correctly, as foreshocks despite the wrong training labels.

To assess longer-term seasonal variations we also evaluated seismic noise recorded at Station NRCA from 2013 to 2022. We verified that noise waveforms did not contain earthquakes using an STA/LTA detector and divided successive 4-month intervals into two classes: Before and After the beginning of each month (2 months per class). We trained our model with the standard train/validation/test ratios and evaluated data from 2013 to 2022, excluding 2016 and 2017, when there were too many earthquakes to find sufficient numbers of clean noise traces. These 2-Class models show test performances that ranged dramatically with levels below 50% and as high as 80%. Performance levels were consistently lower than our results for *pre* and *post* events for the Norcia mainshock.

We also addressed the role of seasonality by comparing results for seismic noise in September and October of three years (2014, 2016, and 2022) and by looking at earthquake data for foreshocks in 2016. In each case, we trained a model to distinguish noise (or earthquake traces) in Sep./Oct using 2- and 4-class models. The results for 2-class models using seismic noise show that performances are lower in 2014 (68%) and 2022 (71%) compared to 2016 (83%) and that pre-Norcia earthquake traces in 2016 show even higher performance (91%). Results for 4-class models show similar trends (Fig. S7). The elastic waves from earthquakes are more able to distinguish times in Sep. 2016 from those in Oct. 2016 compared to noise. For details, see Supplementary section 8.3 and Fig. S7.

Taken together, the tests show that seasonality and variations with time of year do not explain our observations of model test performance for foreshocks, aftershocks and time to failure for the M6.5 Norcia mainshock. However it is clear that seasonality plays some role

https://doi.org/10.1038/s41467-024-54153-w

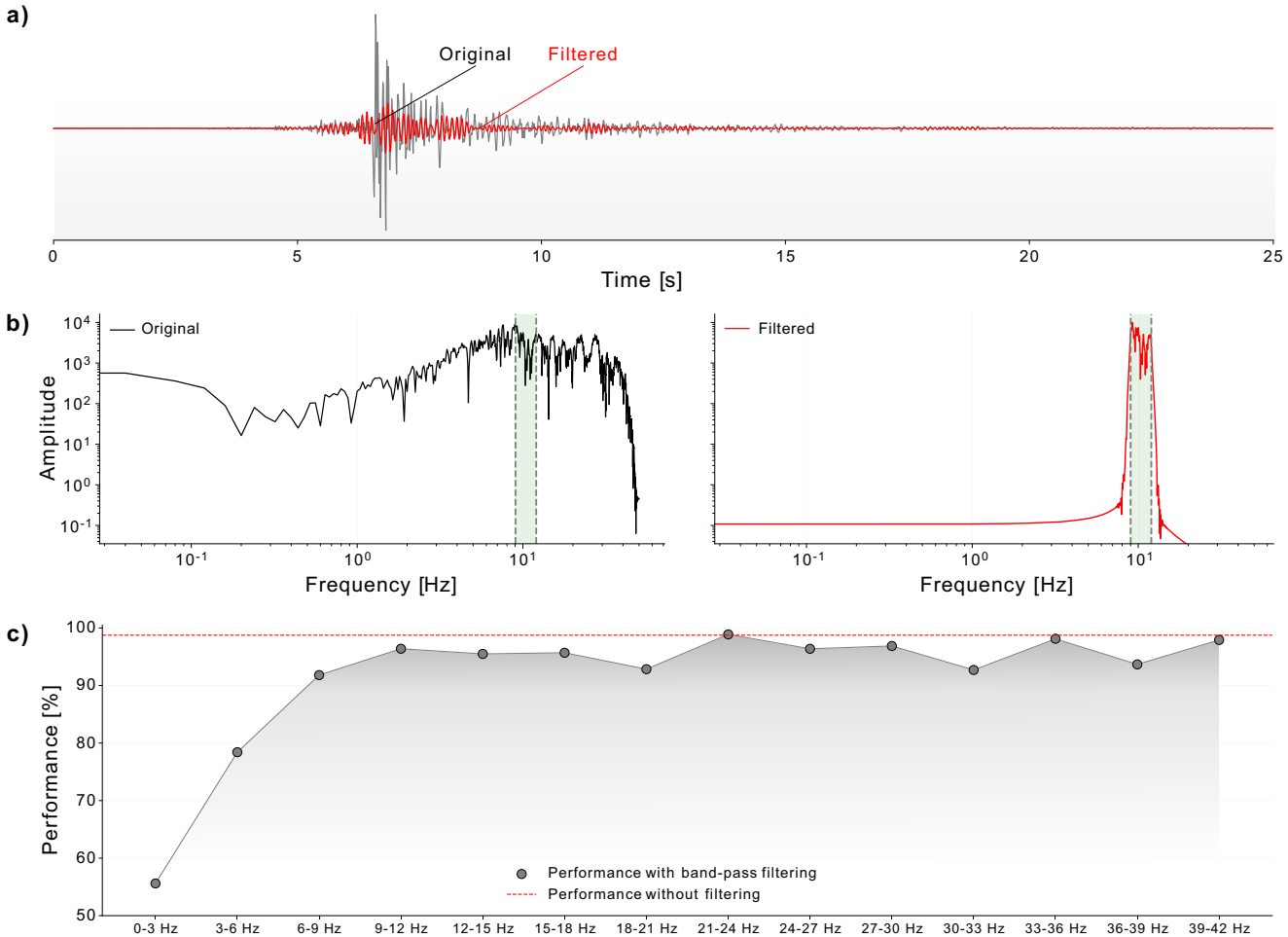

Fig. 5 | Frequency and attenuation analysis for events recorded at station NRCA. a Gray trace shows an example raw seismogram with 5 seconds before and 20 seconds after the P-wave. The red trace is the 9-12 Hz band-pass filtered seismogram at the same scale. **b** shows amplitude spectra for both traces with the band-pass highlighted. **c** Model performance for band-pass filtered seismograms at different frequency ranges. The red line at the top is the baseline performance of raw traces. Note that model performance drops off significantly for models using seismograms with only frequencies below ≈ 10 Hz. The DL models distinguish seismograms for *pre*/*post* and time to failure (TTF) using high frequency content of the elastic waves.

and thus we posit that high performance values in test (> 90%) are partially due to seasonality, but primarily a result of fault stress and damage state. The tests with seismic noise suggest the value of further work to assess its utility for monitoring earthquake-related signals, as has been shown by previous studies using seismic noise to monitor changes in crustal seismic velocity[19,32–36]

Frequency analysis and nearby similar events. One expects that microfracture development and changes in crack aperture with stress would attenuate seismic waves preferentially in specific frequency ranges, for example as demonstrated by recent work in the Geysers geothermal field[65]. We conducted binary classification experiments using filtered waveforms to test the hypothesis that our DL models are sensitive to subtle changes in frequency content (Fig. 5). We also directly compared waveforms for nearby similar events to assess differences in the *pre*/*post* waveform characteristics (Fig. 6).

Figure 5 shows results for band-pass filtered seismic traces recorded at Station NRCA. The sampling rate is 100 Hz, so the seismograms contain frequencies up to 50 Hz. We trained and tested CNN models on our complete data set with seismic traces that were band-pass filtered (see Supplement for details) in steps of 3 Hz. Performance results are lower for models trained with seismograms that do not contain high frequencies (Fig. 5). The dashed red line shows binary classification performance for the raw seismograms at NRCA, which is 98.8%. Band-pass filtered traces show similar results for 3-Hz windows above 10 Hz, within the scatter expected for different realizations (Table 1). Models trained with frequencies lower than 10 Hz show a significant drop in performance. These results show that the evolution of elastic properties during the seismic cycle is more apparent in frequency bands above 10 Hz and that lower frequency waves do not illuminate differences between *pre*/*post* seismic waves.

To further verify that our DL models are detecting TTF and distinguishing foreshocks from aftershocks using subtle changes in seismic waves, we compared waveforms for nearby *pre*/*post* Norcia mainshock (Fig. 6). Our catalog contains 918 *pre*/*post* event pairs with similar hypocenter locations (± 500 m) and local magnitude (± 0.1M). Station NRCA has data for 253 of the pairs (Fig. S9) and we compared spectra (Fig. 6) using the full three-component seismograms. Figure 6 shows an example for two Mw 0.9 events at 6.0 km depth, and Fig. S10 shows additional examples. The spectra and spectral ratios (derived from a multitaper approach[66]), show clear differences, particularly at high frequencies. Aftershock seismograms have a clear peak between 30 and 40 Hz that is not present for foreshocks and these differences extend to lower frequencies, naturally and because the filters are not perfect box-car functions. These differences are common for all similar events (Fig. 9Sd) and presumably relate to differences in fracture density or fault zone damage that appear to increase with time between the foreshock and aftershock (Figs. S10 and S9).

https://doi.org/10.1038/s41467-024-54153-w

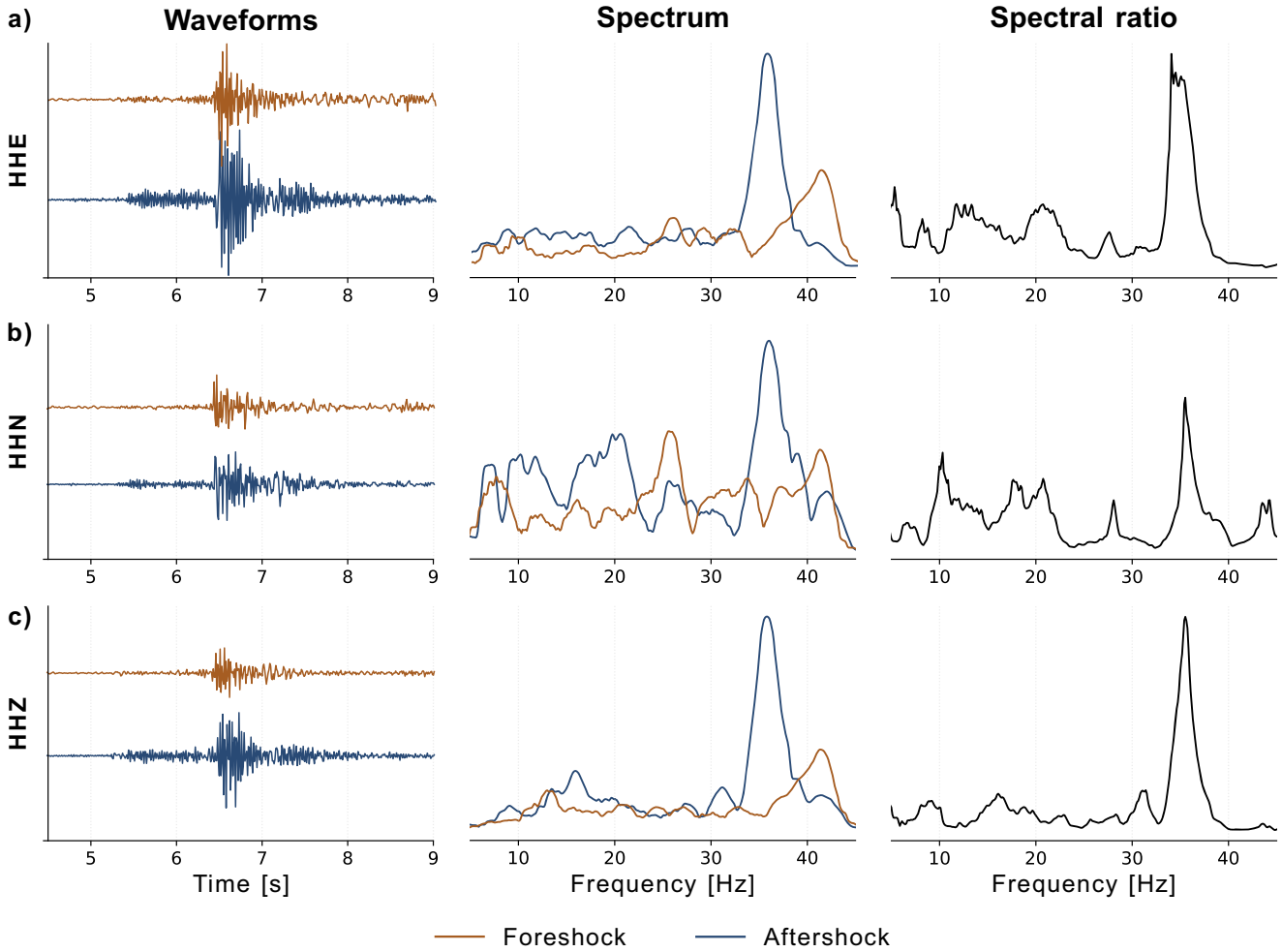

Fig. 6 | Comparison of spectra for co-located foreshocks and aftershocks to assess time dependent changes in crustal properties. a, **b**, and **c** display three component seismograms for two co-located Mw 0.9 events that occurred on the Vettore fault plane (see Fig. S9). Note that the foreshock (orange) which occurred on 2016-08-29 has less energy in the frequency band from 30–40 Hz than the aftershock (blue) which occurred on 2016-11-12. The differences are apparent in each component and clear in the raw data, amplitude spectrum and spectral ratios. The spectral ratio shows aftershock over foreshock and these differences are consistent for all of the 253 event pairs (see also Fig. S10).

Null result for a station far from the Norcia mainshock. As a further test of our DL models, we evaluated seismic traces recorded outside the main study area and away from the Norcia mainshock (Fig. S8). We downloaded data for Station SMA1 for all events in our catalog (4694 *pre* events and 5135 *post* events) and trained our CNN model for binary classification. The new performance ratio was 69.48 ±1.80% on average with an accuracy of 63.53 ±4.07% for *pre* and 75.43 ±2.84% for *post*. These numbers are much lower than those for seismic stations close to the Norcia mainshock (Table 1), but they are above the 50% chance level. We performed a seasonality test for this station using seismic noise from 2013. As before, we selected 4 months at a time and did binary classification for each set (with 2 months for each class), using seismic noise. The average performance for 2013 is 78.4% and for the months September to December it is 72.8%. These data suggest that the results for station SMA1 are primarily seasonal changes rather than changes related to the Norcia earthquake.

To extend this assessment of a null result and better characterize the propagation of seismic waves that do not pass near or through the Vettore fault and the M6.5 Norcia hypocenter, we evaluated seismic data from earthquakes in two additional regions: one located just south of the Norcia region, near Accumoli, and another farther south, near Amatrice (see Fig. S8). The northern region includes the M6 Amatrice epicenter (August 24th 2016) and the southern region includes the town of Amatrice. We included earthquakes that occurred before/after the Norcia mainshock, from September to December 2016, and built a dataset for Station SMA1 using our standard procedure. We trained our CNN model for binary classification to determine whether the model could distinguish *pre* and *post* events using seismograms that did not pass near or through the Vettore fault.

For the earthquakes in the northern region (Fig. S8) the average model performance ratio for *pre*/*post* classification is 77.3% and for southern region the average was 78.2%. Both of these results are similar to the seasonality results presented above for station SMA1. The lower performance values, presumably due primarily to seasonality, are consistent with the idea that seismic waves that do not pass near or through the Vettore fault and the M6.5 Norcia hypocenter are not capable of detecting TTF for the Norcia mainshock.

Discussion

Laboratory and theoretical models of the seismic cycle suggest that the earthquake preparatory phase involves progressive cracking and damage as stresses increase to failure, followed by abrupt changes in elastic properties during dynamic rupture and gradual healing as cracks close. For lab earthquakes, these processes produce clear changes in elastic wave properties that provide precursory signals of

https://doi.org/10.1038/s41467-024-54153-w

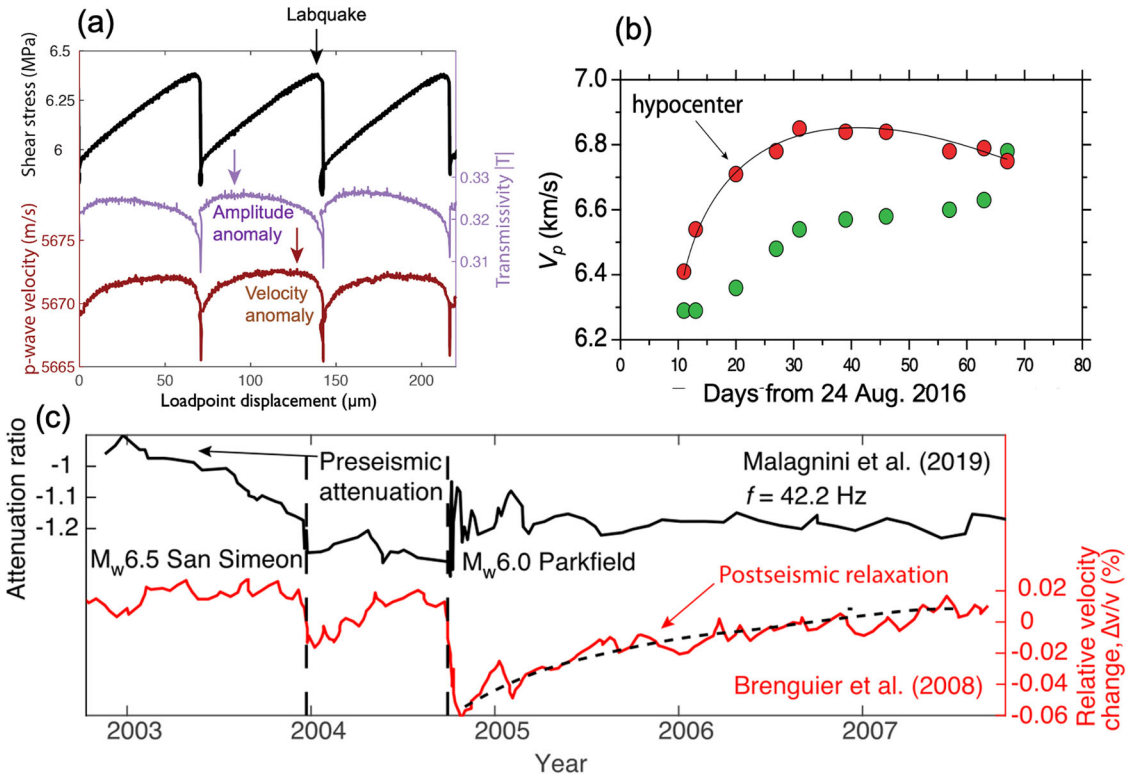

Fig. 7 | Lab and field data showing changes in fault zone elastic properties during the seismic cycle. a swohs data for three lab seismic cycles showing shear stress, Vp, and elastic wave amplitude transmitted through the fault zone. Note that elastic properties evolve systematically and show clear precursors to failure. Data from[50]. **b** shows P-wave velocity inferred from seismic tomography in central Italy around the 2016 M6.5 Norcia earthquake on 30 Oct. 2016. Red dots show the area around the mainshock hypocenter, and green dots show Vp data for the area north of the hypocenter. Modified from[21], (**c**) shows data for (black) seismic attenuation based on repeating earthquakes during 5-yr period (Oct 2002–Sept 2007) and (red) changes in Vp based on ambient seismic noise. Time period includes the 2003 Mw 6.5 San Simeon and 2004 Mw 6.0 Parkfield earthquakes. Note that seismic attenuation data and Vp show clear changes for the 2003 San Simeon and 2004 Parkfield earthquakes. Data from[17,69].

impending failure (Fig. 7a). However, while similar observations have been made in nature in special cases, there is still an open question about why precursors are common in lab experiments but rare for natural faults. Our results provide an important opportunity to address this question using DL performance as a proxy for fault zone damage/stress state.

We leverage the power of CNNs, which excel at extracting high-level features. The CNN convolution kernels scan the input data, enabling the network to capture intricate and informative details. Pooling layers reduce the spatial dimensions of the data while preserving essential features, enhancing the network's ability to recognize patterns, subtle differences, and hierarchies at multiple levels of abstraction (e.g., ref. 67). We train a CNN model to differentiate between foreshocks, aftershocks and TTF of the 2016 M6.5 Norcia mainshock, and the model is remarkably successful. Our model can distinguish between seismic waves prior to and after the mainshock with greater than 90% accuracy (Figs. 1 and 3). The performance is equally good for models trained on TTF, based on *n*-class models trained to distinguish between different time periods before/after the mainshock (Fig. 2 and Table 1). The results are robust across the full range of magnitudes down to M0.5 and for multiple seismic stations around the Norcia mainshock. These results are consistent with lab data and with the field observations available in a few special cases (Fig. 7).

In the lab, fault zone elastic properties vary systematically during the seismic cycle (Fig. 7). Elastic wave speed and transmitted amplitude drop abruptly with stress drop as the fault slips and frictional contact junctions are rejuvenated. Wave speed and amplitude then increase progressively when the fault is quasi-stationary and frictional contacts grow and increase in stiffness. These changes in fault elastic properties continue until stress increases enough for the fault to creep at a rate sufficient to re-open cracks and reduce asperity contact stiffness. These data are consistent with our findings for the performance of DL models in distinguishing foreshocks from aftershocks. Our DL results are also in accord with work showing that the evolution of fault zone elastic properties can be used to predict lab earthquakes[49,68]. We posit that seismic waves traveling through and near the M6.5 Norcia hypocenter and Vettore fault record subtle changes in elastic wave properties similar to those observed in the lab (Fig. 7). For the lab seismic cycle, the evolution of wave speed and amplitude during the seismic cycle provide clear precursors to impending failure. While our DL results for the 2016 central Italy seismic sequence do not have the same resolution, they suggest a similar pattern of changes before and after the M6.5 Norcia event.

Additional field observations confirm the interpretation that our DL models detect changes in seismic wave properties during the seismic cycle. The field data are from dense monitoring networks or detailed seismic tomographic studies done using local earthquakes or ambient noise[17,21,69]. Chiarabba and co-workers[21] studied the 2016 AVN seismic sequence of central Italy (Fig. 7b) and found changes in P-wave velocity and Vp/Vs in the region around the hypocenter preceding the Norcia mainshock. The changes persist for several weeks and are more significant around the hypocenter than to the north and south (Fig. 7b). We posit that our DL models illuminate the same changes in elastic properties prior to and after the M6.5 Norcia mainshock. The performance of our DL models is greatest for stations near the epicenter and for seismic waves that travel near the hypocenter, in concert with the results of seismic tomography[21].

Our results for DL models trained on bandpass filtered seismic waves (Fig. 5) and the comparison of event pairs *pre*/*post* the Norcia mainshock (Fig. 6) suggest that differences in elastic properties are most apparent from 30–40 Hz. For a fault zone with Vp of ≈ 4.5–6 km/s this represents wavelengths of 150–250 m, which corresponds to events of M2-3. Therefore one explanation of our results is that DL models track the fault damage and fracturing associated with Mw 2–3 earthquakes. Prior to the Norcia mainshock the rate of such events is low. This increases abruptly with the Visso-Norcia mainshocks and then progressively reduces with time. Following the mainshock, fluid flow into these fractures promotes the activation of chemical and mechanical healing, which facilitates crack closure and healing. We note that carbonates are the dominant lithology of the seismogenic zone in the Apennines, and lab experiments show rapid healing rates for these fault rocks[70].

The performance of our DL models is also consistent with previous work[17,69] showing changes in elastic wave properties prior to and after earthquake rupture (Fig. 7). These studies documented the evolution of crustal wave attenuation using repeating earthquakes and passive noise seismology for the 2003 Mw 6.5 San Simeon earthquake and the M6 2004 Parkfield earthquake (Fig. 7c). Both studies document progressive changes prior to the mainshock followed by postseismic relaxation. The performance of our DL models is consistent with these data. In addition, our comparison of seismic wave spectra for co-located foreshock/aftershock pairs shows clear differences (Figs. 6 and S10) that are consistent with changes in transmitted amplitude and attenuation.

We challenged the DL models in several ways to assess reliability and robustness and to test the hypothesis that DL models can track the evolution of fault properties during the seismic cycle. We achieve very good model accuracy with high precision and few examples of false negatives. In particular, we test the hypothesis that DL models recognize the evolution of seismic wave attenuation during the seismic cycle by training/testing models on bandpass-filtered waveforms, finding that the model is more confident in the classification when using seismic waves that contain higher frequencies. We get the same results by comparing waveforms for co-located similar events occurring *pre*/*post* mainshock, finding clear differences in spectral content. The null tests were also helpful, showing that stations far from the hypocenter and seismic waves that do not pass near or through the mainshock do not successfully distinguish between foreshocks or aftershocks.

A critical limitation of our work is that we have studied only one location. A clear goal for future works will be to assess the efficacy of this method in other tectonic settings and to investigate the possibility of using real-time data to identify precursors prior to a mainshock rather than working with post-event data.

In conclusion, we note that monitoring the evolution of tectonic faults during the seismic cycle could inform earthquake forecasting and early warning, but distinguishing changes in fault properties is challenging. We leverage a CNN classifier to investigate the variations in elastic properties of a fault zone during a seismic sequence, following previous findings from the laboratory. Our results demonstrate that the CNN model can accurately classify the fault zone's state and TTF using raw seismic waveforms, achieving over 90% accuracy across various test scenarios. The model's performance was consistent across multiple local stations. Our validation approach uses multiple verification tests, such as using incorrect mainshock dates and distant seismic waves. These tests underscored the model's sensitivity to subtle changes in elastic wave properties.

Our findings support the hypothesis that the CNN model can detect progressive damage and microfracture development within the fault zone and surrounding rock prior to an earthquake, as well as gradual postseismic recovery. This aligns with theoretical and laboratory expectations for seismic cycles and suggests that DL models can enhance existing techniques for earthquake early warning and seismic hazard assessment.

This study employes DL methods to track the evolution of fault zone properties during the seismic cycle. If our result is generalizable to other tectonic settings it could improve earthquake early warning and seismic hazard analysis and ultimately provide a pathway to identify earthquake precursors.

Methods

We gathered the data necessary for our study using ObsPy, an open-source Python framework specifically designed for processing seismological data (more information available at https://docs.obspy.org/). With ObsPy, we were able to download waveforms and metadata for each event and station from the INGV network and assemble the dataset (Fig. 1).

The datasets were organized separately for each station, with two `.hdf5` files: one containing waveforms for earthquakes that occurred *pre* Norcia mainshock and one containing waveforms *post*. The Hierarchical Data Format version 5 (HDF5) is an open-source file format designed to handle large, complex, and heterogeneous data sets. It uses a file directory structure that enables the organization of data within the file in a variety of structured ways, similar to how files are organized on a computer. In addition, we have two `.csv` files containing metadata for each earthquake, which helps us to identify the type of receiver used for recording as well as the location and characteristics of the earthquake. This metadata allows us to understand better the context in which each earthquake occurred and how it was recorded.

Other information and further details of our methods are provided in the Supplement Section 8.

Data availability

The seismicity catalog used in this work is published by Tan et al., 2021 and access can be obtained at the Zenodo dataset repository (https://doi.org/10.5281/zenodo.4662870[71]). The dataset coming from our selected sub-region, as well as the sub-catalog, can be obtained at Zenodo dataset repository (10.5281/zenodo.12795621[72]). The preprocessed dataset, ready to be used with our code, can be obtained at the Zenodo dataset repository (10.5281/zenodo.12806081[73]). Correspondence and requests for materials should be addressed to Laura Laurenti or Chris Marone.

Code availability

All codes are accessible on GitHub at https://github.com/lauralaurenti/CNN_Norcia_sequence_evolution, which is linked to Zenodo: https://doi.org/10.5281/zenodo.13909498.

References

1. 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).
2. Sibson, R. Fault rock s and fault mechanisms. *Geol. Soc. Lond. J.* **133**, 191–231 (1977).
3. Chen, W. Y., Lovell, C. W., Haley, G. M. & Pyrak-Nolte, L. J. Variation of shear-wave amplitude during frictional sliding. *Int. J. Rock. Mech. Min. Sci. Geomech. Abstr.* **30**, 779–784 (1993).
4. Goebel, T. H., Schorlemmer, D., Becker, T. W., Dresen, G. & Sammis, C. G. Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophys. Res. Lett.* **40**, 2049–2054 (2013).
5. Kwiatek, G. et al. Intermittent criticality multi-scale processes leading to large slip events on rough laboratory faults. *J. Geophys. Res.* **129**, e2023JB028411(2024).

6. Lei, X. & Ma, S. Laboratory acoustic emission study for earthquake generation process. *Earthq. Sci.* **27**, 627–646 (2014).
7. Stanchits, S., Vinciguerra, S. & Dresen, G. Ultrasonic velocities, acoustic emission characteristics and crack damage of basalt and granite. *Pure Appl. Geophysics* **163**, 975–994 (2006).
8. Guérin-Marthe, S. Preparatory slip in laboratory faults: Effects of roughness and load point velocity. *J. Geophys. Res.: Solid Earth* **128**, e2022JB025511 (2023).
9. Main, I. & Meredith, P. G. Classification of earthquake precursors from a fracture mechanics model. *Tectonophysics* **167**, 273–283 (1989).
10. McBeck, J., Ben-Zion, Y. & Renard, F. Fracture network localization preceding catastrophic failure in triaxial compression experiments on rocks. *Frontiers in Earth Sci.*, **9**, 7788112021.
11. Renard, F. et al. Microscale characterization of rupture nucleation unravels precursors to faulting in rocks. *Earth Planet. Sci. Lett.* **476**, 69–78 (2017).
12. Renard, F. et al. Critical evolution of damage toward system-size failure in crystalline rock. *J. Geophys. Res.: Solid Earth* **123**, 1969–1986 (2018).
13. Scuderi, M., Marone, C., Tinti, E., Di Stefano, G. & Collettini, C. Precursory changes in seismic velocity for the spectrum of earthquake failure modes. *Nat. Geosci.*, https://doi.org/10.1038/ngeo2775 (2016).
14. McLaskey, G. C. Earthquake initiation from laboratory observations and implications for foreshocks. *J. Geophys. Res.: Solid Earth* **124**, 12882–12904 (2019).
15. Shreedharan, S., Bolton, D. C., Rivière, J. & Marone, C. Competition between preslip and deviatoric stress modulates precursors for laboratory earthquakes. *Earth and Planetary Sci. Lett.*, **553**, 116623 (2021).
16. Niu, F., Silver, P., Daley, T., Cheng, X. & Majer, E. Preseismic velocity changes observed for active source monitoring at the Parkfield SAFOD drill site. *Nature* **454**, 204–8 (2008).
17. Brenguier, F. et al. Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations. *Science* **321**, 1478–1481 (2008).
18. Lucente, F. P. et al. Temporal variation of seismic velocity and anisotropy before the 2009 MW 6. *Geology* **38**, 1015–1018 (2010).
19. Steinmann, R., Seydoux, L., Journeau, C., Shapiro, N. M. & Campillo, M. Machine learning analysis of seismograms reveals a continuous plumbing system evolution beneath the Klyuchevskoy volcano in Kamchatka, Russia. *J. Geophys. Res. Solid Earth*, **129**, e2023JB027167 (2024).
20. Ghimire, P. S., Guéguen, P., Giffard-Roisin, S. & Schorlemmer, D. Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 gorkha nepal earthquake. *Earthq. Spectra* **38**, 2970–2993 (2022).
21. Chiarabba, C., De Gori, P., Segou, M. & Cattaneo, M. Seismic velocity precursors to the 2016 mw 6. *Geology* **48**, 924–928 (2020).
22. Poli, P. Creep and slip: Seismic precursors to the Nuugaatsiaq landslide (Greenland). *Geophys. Res. Lett.* **44**, 8832–8836 (2017).
23. Schaff, D. P. Placing an upper bound on preseismic velocity changes measured by ambient noise monitoring for the 2004 MW 6. *Bull. Seismological Soc. Am.* **102**, 1400–1414 (2012).
24. Elizabeth Li, Y., O'Malley, D., Beroza, G., Curtis, A. & Johnson, P. Machine learning developments and applications in solid-earth geosciences: fad or future? *J. Geophys. Res.: Solid Earth* **128**, e2022JB026310 (2023).
25. Rouet-Leduc, B., Hulbert, C. & Johnson, P. A. Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nat. Geosci.* **12**, 75–79 (2019).
26. Goebel, T. H. W., Sammis, C. G., Becker, T. W., Dresen, G. & Schorlemmer, D. A comparison of seismicity characteristics and fault structure between stick-slip experiments and nature. *Pure Appl. Geophysics* **172**, 2247–2264 (2015).
27. Kwiatek, G. et al. Months-long preparation of the 2023 MW 7.8 kahramanmaraş earthquake, türkiye. https://doi.org/10.21203/rs.3.rs-2657873/v1 (2023).
28. Martínez-Garzón, P. & Poli, P. Cascade and pre-slip models oversimplify the complexity of earthquake preparation in nature. *Commun. Earth Environ.* **5**, 120 (2024).
29. Obermann, A., Planès, T., Larose, E., Sens-Schönfelder, C. & Campillo, M. Depth sensitivity of seismic coda waves to velocity perturbations in an elastic heterogeneous medium. *Geophys. J. Intl*, **195**, 372–382 (2013).
30. Bouchon, M., Durand, V., Marsan, D., Karabulut, H. & Schmittbuhl, J. The long precursory phase of most large interplate earthquakes. *Nat. Geosci.* **6**, 299–302 (2013).
31. Beaucé, E., Poli, P., Waldhauser, F., Holtzman, B. & Scholz, C. Enhanced tidal sensitivity of seismicity before the 2019 magnitude 7. *Geophys. Res. Lett.* **50**, e2023GL104375 (2022).
32. Shapiro, N., Campillo, M., Stehly, L. & Ritzwoller, M. High-resolution surface-wave tomography from ambient seismic noise. *Sci. (N. Y., N. Y.)* **307**, 1615–8 (2005).
33. Seydoux, L. et al. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nat. Commun.* **11**, 3972 (2020).
34. Steinmann, R., Seydoux, L. éonard, Beaucé, E. & Campillo, M. Hierarchical exploration of continuous seismograms with unsupervised learning. *J. Geophys. Res.: Solid Earth* **127**, e2021JB022455 (2022).
35. Steinmann, R., Seydoux, L. & Campillo, M. Ai-based unmixing of medium and source signatures from seismograms: ground freezing patterns. *Geophys. Res. Lett.* **49**, e2022GL098854 (2022).
36. Brenguier, F. et al. Train traffic as a powerful noise source for monitoring active faults with seismic interferometry. *Geophys. Res. Lett.* **46**, 9529–9536 (2019).
37. Mikhael, N., Poli, P. & Garambois, S. Non-linear seismic velocity variations observed during a seismic swarm in the alto tiberina low angle normal fault from ambient noise correlation measurements. *J.Geophys. Res.: Solid Earth*, **129**, e2023JB028232 (2024).
38. Goebel, T. H. W., Brodsky, E. E. & Dresen, G. Fault roughness promotes earthquake-like aftershock clustering in the lab. *Geophys. Res. Lett.* **50**, e2022GL101241 (2023).
39. Pandey, K., Taira, T., Dresen, G. & Goebel, T. H. Inferring damage state and evolution with increasing stress using direct and coda wave velocity measurements in faulted and intact granite samples. *Geophys. J. Int.* **235**, 2846–2861 (2023).
40. Rouet-Leduc, B. et al. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **44**, 9276–9282 (2017).
41. Rouet-Leduc, B. et al. Estimating fault friction from seismic signals in the laboratory. *Geophys. Res. Lett.* **45**, 1321–1329 (2018).
42. Hulbert, C. et al. Similarity of fast and slow earthquakes illuminated by machine learning. *Nat. Geosci.* **12**, 69–74 (2019).
43. Corbi, F. et al. Machine learning can predict the timing and size of analog earthquakes. *Geophys. Res. Lett.* **46**, 1303–1311 (2019).
44. Bolton, D. C. et al. Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Res. Lett.* **90**, 1088–1098 (2019).
45. Bolton, D., C., Shreedharan, S., Rivière, J. & Marone, C. Acoustic energy release during the laboratory seismic cycle: Insights on laboratory earthquake precursors and prediction. *Journal of Geophysical Research: Solid Earth*, **125**, https://doi.org/10.1029/2019jb018975 (2020).
46. Trugman, D. et al. The spatio-temporal evolution of granular microslip precursors to laboratory earthquakes. *Geophys. Res. Lett.* https://doi.org/10.1029/2020GL088404 (2020).

47. Jasperson, H. et al. Attention network forecasts time-to-failure in laboratory shear experiments. *J. Geophys. Res.: Solid Earth*, **126**, https://doi.org/10.1029/2021JB022195 (2021).
48. Johnson, P. A. et al. Laboratory earthquake forecasting: A machine learning competition. *Proc. Natl. Acad. Sci.* **118**, e2011362118 (2021).
49. Shokouhi, P. et al. Deep learning can predict laboratory quakes from active source seismic data. *Geophys. Res. Lett.* **48**, e2021GL093187 (2021).
50. 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**, e2020JB021588 (2021).
51. Blanke, A., Kwiatek, G., Goebel, T. H. W., Bohnhoff, M. & Dresen, G. Stress drop–magnitude dependence of acoustic emissions during laboratory stick-slip. *Geophys. J. Int.* **224**, 1371–1380 (2021).
52. McBeck, J., Aiken, J. M., Ben-Zion, Y. & Renard, F. Predicting the proximity to macroscopic failure using local strain populations from dynamic in situ x-ray tomography triaxial compression experiments on rocks. *Earth Planet. Sci. Lett.* **543**, 116344 (2020).
53. 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).
54. Karimpouli, S., Kwiatek, G., Martínez-Garzón, P., Dresen, G. & Bohnhoff, M. Unsupervised clustering of catalogue-driven features for characterizing temporal evolution of labquake stress. *Geophys. J. Int.* **237**, 755–771 (2024).
55. 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**, https://doi.org/10.1029/2024GL108655 (2024).
56. Chiaraluce, L. et al. The 2016 Central Italy Seismic Sequence: A First Look at the Mainshocks, Aftershocks, and Source Models. *Seismological Res. Lett.* **88**, 757–771 (2017).
57. Scognamiglio, L. et al. Complex fault geometry and rupture dynamics of the MW 6. *J. Geophys. Res.: Solid Earth* **123**, 2943–2964 (2018).
58. Tan, Y. J. et al. Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence. *The Seismic Record*, 11–19, https://doi.org/10.1785/0320210001 (2021).
59. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
60. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980 (2014).
61. Collettini, C., Barchi, M. R., De Paola, N., Trippetta, F. & Tinti, E. Rock and fault rheology explain differences between on fault and distributed seismicity. *Nat. Commun.* **13**, 1–11 (2022).
62. Poli, P., Marguin, V., Wang, Q., D'Agostino, N. & Johnson, P. Seasonal and coseismic velocity variation in the region of L'Aquila from single station measurements and implications for crustal rheology. *J. Geophys. Res.: Solid Earth* **125**, e2019JB019316 (2020).
63. Hillers, G.Y., Ben-Zion, Y., Campillo, M. & Zigone, D. Seasonal variations of seismic velocities in the San Jacinto fault area observed with ambient seismic noise. *Geophysi. J. Intl*, **202**, 2 (2015).
64. Li, J. et al. Strong seasonal variations of seismic velocity in eastern margin of Tibetan plateau and Sichuan basin from ambient noise interferometry. *J. Geophys. Res. : Solid Earth*, **126**, e2021JB022600 (2021).
65. Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F. & Repetto, D. Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Sci. Adv.* **4**, eaao2929 (2018).
66. Prieto, G. The multitaper spectrum analysis package in Python. *Seismological Res. Lett.* **93**, 1922–1929 (2022).
67. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
68. Shreedharan, S., Bolton, D. C. Rivière, J. & Marone, C. Competition between preslip and deviatoric stress modulates precursors for laboratory earthquakes. *Earth Planet. Sci. Lett.* **553**, 116623 (2021).
69. Malagnini, L., Dreger, D. S., Bürgmann, R., Munafò, I. & Sebastiani, G. Modulation of seismic attenuation at Parkfield, before and after the 2004 m6 earthquake. *J. Geophys. Res.: Solid Earth* **124**, 5836–5853 (2019).
70. Carpenter, B. M., Scuderi, M. M., Collettini, C. & Marone, C. Frictional heterogeneities on carbonate-bearing normal faults: Insights from the Monte Maggio fault, Italy. *J. Geophys. Res. : Solid Earth*, **119**, https://doi.org/10.1002/2014JB011337 (2014).
71. Tan, Y. J., Waldhauser, F. & Ellsworth, W. Machine-learning-based high-resolution earthquake catalog for the 2016-2017 central Italy sequence, https://doi.org/10.5281/zenodo.4662870 (2021).
72. Paoletti, G. D-set: Probing the evolution of fault properties during the seismic cycle with deep learning (1.0.0), https://doi.org/10.5281/zenodo.12795621 (2024).
73. Laurenti, L. Probing the evolution of fault properties during the seismic cycle with deep learning - dataset, https://doi.org/10.5281/zenodo.12806081 (2024).

Acknowledgements

We would like to thank Chris Johnson, Paul Johnson, Massimo Cocco, and Lauro Chiaraluce for helpful discussions. This work was supported by European Research Council Advance grant 835012 (TECTONIC) to CM and by the RETURN Extended Partnership, which provided funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan - NRRP, Mission 4, Component 2, Investment 1.3 - D.D. 1243 2/8/2022, PE0000005).

Author contributions

L.L.: Conceptualization, methods, software, formal analysis, investigation, writing - original draft, writing - review & editing. G.P.: methods, software, formal analysis, investigation, data curation, writing - original draft, writing - review & editing, visualization. E.T.: Conceptualization, investigation, validation, writing - original draft, writing - review & editing. F.G.: Methods, data curation, validation, supervision, writing - review & editing C.C.: Conceptualization, investigation, writing - original draft, writing - review & editing. C.M.: Conceptualization, original draft, writing, focusing the investigation, methods, supervision, resources, funding acquisition.

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/s41467-024-54153-w.

Correspondence and requests for materials should be addressed to Laura Laurenti.

Peer review information *Nature Communications* thanks Andrew Delorey, Patricia Martinez-Garzon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.