ANNUAL REVIEWS

Annual Review of Earth and Planetary Sciences

Machine Learning in Earthquake Seismology

S. Mostafa Mousavi[1,2] and Gregory C. Beroza[2]

[1]Google Inc., Mountain View, California, USA

[2]Department of Geophysics, Stanford University, Stanford, California, USA; email: beroza@stanford.edu

Annu. Rev. Earth Planet. Sci. 2023. 51:105–29

First published as a Review in Advance on November 21, 2022

The *Annual Review of Earth and Planetary Sciences* is online at earth.annualreviews.org

https://doi.org/10.1146/annurev-earth-071822-100323

Keywords

machine learning, artificial intelligence, neural networks, earthquakes, seismology

Abstract

Machine learning (ML) is a collection of methods used to develop understanding and predictive capability by learning relationships embedded in data. ML methods are becoming the dominant approaches for many tasks in seismology. ML and data mining techniques can significantly improve our capability for seismic data processing. In this review we provide a comprehensive overview of ML applications in earthquake seismology, discuss progress and challenges, and offer suggestions for future work.

- Conceptual, algorithmic, and computational advances have enabled rapid progress in the development of machine learning approaches to earthquake seismology.
- The impact of that progress is most clearly evident in earthquake monitoring and is leading to a new generation of much more comprehensive earthquake catalogs.
- Application of unsupervised approaches for exploratory analysis of these high-dimensional catalogs may reveal new understanding of seismicity.
- Machine learning methods are proving to be effective across a broad range of other seismological tasks, but systematic benchmarking through open source frameworks and benchmark data sets are important to ensure continuing progress.

1. INTRODUCTION

Machine learning (ML) is a collection of methods used to develop understanding and predictive capability by learning relationships embedded in data. ML has a decades-long history in seismology (e.g., Turhan Taner et al. 1988, Dowla et al. 1990); however, its application has been growing rapidly in the past few years. Initial efforts focused on supervised learning in seismic data processing, where a set of training data and their corresponding labels or targets are used to build (train) a model to find the connection between the data and labels. These early studies had limited impact due to conceptual, algorithmic, and computational obstacles to training effective models. The application of more complex models using deep neural networks (DNNs) became feasible due to computational and technological breakthroughs.

Earthquake seismology uses a sequence of processing steps to monitor seismic activity. These steps were initially performed manually by skilled analysts and later were performed automatically with algorithms designed to detect impulsive phase arrivals. Rapidly expanding data volumes have motivated development of efficient, robust processing workflows that ML is poised to supplant. In this review, we discuss the main processing tasks used to extract information about earthquakes and related phenomena. We organize the review by tasks, document early efforts and ensuing progress, and sample the current state of the art for each. We conclude with recommendations to accelerate progress in ML for seismological applications and speculate on future trends.

Deep neural network (DNN): an artificial neural network with multiple hidden layers between input and output layers that consist of neurons, synapses, weights, biases, and activation functions

Back propagation: a widely used algorithm for training neural networks by fine-tuning the connection weights based on the prediction errors

Convolutional neural network (CNN): a neural network that uses convolutions to extract and connect features from local regions of the input

Recurrent neural network (RNN): a neural network that runs multiple times, where parts of each run feed into the next run; particularly useful for modeling sequential data

Semisupervised learning: training a model on data where some of the training examples have labels but others do not; can be useful if labels are expensive to obtain but unlabeled examples are plentiful

2. MACHINE LEARNING FOR SEISMIC EVENT MONITORING

2.1. Event Discrimination

Event discrimination seeks to classify seismic events by type (e.g., earthquake versus explosion) or characteristics (e.g., local versus teleseismic).[1] We consider three groups of ML applications that discriminate (*a*) natural earthquakes from explosions, (*b*) various volcano-seismic events from each other, and (*c*) earthquakes from mining-related events.

2.1.1. Explosions versus earthquakes. The development of back propagation (Rumelhart et al. 1986) and the Cold War motivated discrimination of the signature of tectonic earthquakes from human-made explosions as among the first tasks for which seismologists employed neural networks (e.g., Dowla et al. 1990, Pulli & Dysart 1990). Their potential for event discrimination was immediately evident. Despite simple architectures—with one or two layers and a handful of artificial neurons trained on tens to hundreds of examples—they achieved impressive performance (usually above 90% classification accuracy). Modern counterparts of these artificial intelligence (AI)-based discriminators continue this tradition. Linville et al. (2019) showed that both convolutional neural networks (CNNs) and recurrent neural networks (RNNs) (**Figure 1**) could classify local quarry blasts and tectonic events with accuracies above 99%. Deep-learning models can also identify human errors in seismic event catalogs and be more robust than classical approaches for low signal-to-noise ratio (SNR) data (Tibi et al. 2019). Performance continues to improve with the use of advanced architectures such as attentive CNNs (Ku et al. 2020) or semisupervised learning that incorporates both labeled and unlabeled data for model building.

2.1.2. Volcano-seismic events. Classification of magmatic events as volcano-tectonic earthquakes, long-period events, volcanic tremor, or explosive earthquakes is another discrimination task. Most applications use classical techniques (Masotti et al. 2006, Beyreuther et al. 2008) or

[1] We differentiate discrimination from detection in which the goal is to identify a signal of interest among a variety of recorded events.

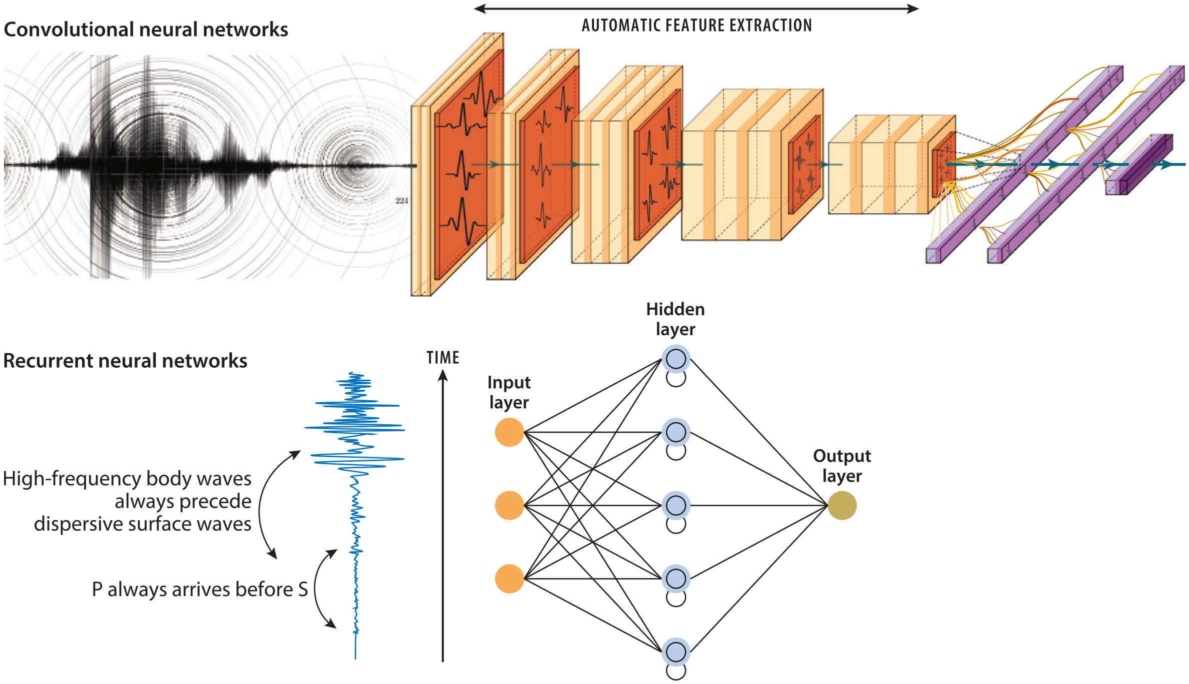

Figure 1

Schematic representations of convolutional neural networks and recurrent neural networks consisting of multiple layers. The configuration of connections between neurons and the performed operations for each layer within these networks enable specific capabilities, such as automatic feature extraction and temporal dependency modeling of seismic signals.

shallow neural networks (Esposito et al. 2006) applied to hand-engineered features. Deep learning has not seen much effort due to the scarcity of labeled data sets for volcano monitoring. Inductive transfer learning to tune pretrained CNN models by handwritten (Titos et al. 2020) or natural (Lara et al. 2021) images has been used for classification of volcano-seismic events based on their time-frequency representations. Other classification tasks such as discriminating volcano-seismic signals associated with different eruption periods, for which more training data are available, have shown encouraging results for deep Bayesian neural networks (Rodriguez et al. 2021).

Bayesian neural network: a probabilistic neural network that accounts for uncertainty in weights and outputs

Autoencoder: a neural network model whose goal is to reproduce its input, typically through encoding in a narrow bottleneck of the network

Loss function: used to measure how far a model's predictions are from its label

2.1.3. Other seismic events. Neural networks have proven useful for the important tasks of discriminating seismic events based on their source depths (Mousavi et al. 2016a), epicentral distance (e.g., Kuyuk & Ohno 2018, Mousavi et al. 2019b), and source type (e.g., Nakano et al. 2019, Peng et al. 2020, Köhler et al. 2022). Mousavi et al. (2019b) presented an unsupervised deep-learning approach that used a self-supervised convolutional autoencoder for feature learning and dimensionality reduction to discriminate local from teleseismic waveforms. Simultaneous optimization of two loss functions—one for feature learning and the other for clustering—resulted in a small feature set from spectrograms that are optimized for this classification problem. This unsupervised approach can match the performance of supervised models that use much larger labeled data sets. Nakano et al. (2019) trained a CNN to discriminate tectonic tremor from local earthquakes based on their time-frequency representation. Peng et al. (2020) used a capsule neural network to classify microearthquake records in underground mines into five classes: microearthquakes, blasts, ore extraction, mechanical noise, and electromagnetic interference. Köhler et al. (2022) combined the empirical matched field method and a CNN to differentiate iceberg calving events from earthquakes and to classify them based on their locations.

Supervised ML techniques are well suited for event discrimination. The limited size of discrimination data sets—from a few thousand to several tens of thousands of samples—is an impediment to applying deep-learning methods. In addition to transfer learning (Titos et al. 2020, Lara et al. 2021), semisupervised and unsupervised learning (Mousavi et al. 2019b) approaches using synthetic data or data augmentation might help address this limitation. In the current literature, demonstrations of model generalization to other regions or cases are quite rare, and their limitations are unclear. While the performance metrics for the classification task are standardized and commonly used by researchers, the lack of common benchmark data sets renders comparisons among different approaches difficult. Compiling a unified benchmark data set and introducing baseline methods would help accelerate improvements in ML applications for this task.

2.2. Earthquake Signal Detection

Earthquake signal detection is the task of identifying the signature of earthquakes within background noise and a wide variety of non-earthquake signals in continuous seismic data. Minimizing false negatives (missing events) and false positives (misidentifying noise as earthquake signals) is the main objective. Detection is often defined as a binary classification of each sample point as noise or signal and is often performed for either long durations or a large number of stations, such that efficiency is important. There are two main categories of conventional methods: characteristic function methods [e.g., short-term average over long-term average (STA/LTA), wavelet-based methods] and similarity search (e.g., template matching, Fingerprint And Similarity Thresholding). Characteristic function methods are efficient and generalizable, but they are sensitive to background noise and susceptible to high error rates. Similarity search is robust, but it is computationally less efficient and insensitive to sources in new locations.

Wang & Teng (1995) developed an early AI-based detector in which 50 waveforms were used to train two neural networks—one with recursive STA/LTA time series and the other with moving window spectrograms as input. This detector outperformed conventional STA/LTA, especially for low SNR data or spike-like non-earthquake signals. Other AI detectors were developed using a variety of techniques, including support vector machine (e.g., Madureira et al. 2013), hidden Markov models (HMMs) (e.g., Beyreuther & Wassermann 2008), and more recently CNNs (Perol et al. 2018), RNNs (Mousavi et al. 2019c), multifeature fusion networks (G. Kim et al. 2020), capsule neural networks (Saad & Chen 2022), and attentive models (Mousavi et al. 2020). Supervised approaches prevail for this classification problem. The variety of earthquake signal shapes and the existence of instrumental, human-made, and natural non-earthquake signals pose challenges to detection as a clustering (unsupervised learning) problem that would generalize to all regions; however, Yoon et al. (2015) used set-based similarity search to generalize from template matching to unsupervised data mining.

The success of ML techniques for earthquake signal detection has expanded beyond conventional seismic data/events to earthquake detection directly from scanned images of analog seismograms (Wang et al. 2019), smartphone data (Kong et al. 2019, Chin et al. 2020), distributed acoustic sensing (Hernandez et al. 2022), and low-frequency earthquakes (Thomas et al. 2021) and tremor (Rouet-Leduc et al. 2020).

ML earthquake detectors can achieve similar sensitivity to match filtering, but with greater computational efficiency and ability to generalize to signals. They can result in fewer false positives than STA/LTA and be more robust to low SNR. This combination allows efficient real-time processing in diverse environments. The high efficiency of AI detectors makes them viable for real-time applications including volcano monitoring (Retailleau et al. 2022) and earthquake early

Capsule neural network: a type of neural network that can be used to model hierarchical relationships

Unsupervised learning: ML model that finds patterns in an unlabeled data set

Short-term average over long-term average (STA/LTA): the most widely used algorithm for earthquake signal detection and onset arrival time measurement; it continuously calculates the average values of the absolute amplitude of a seismic signal in two consecutive moving-time windows to search for impulsive arrivals

Fingerprint And Similarity Thresholding: an earthquake detection algorithm based on locality-sensitive hashing for set-based waveform similarity search

Support vector machine: a supervised ML algorithm that analyzes data for classification and regression

Hidden Markov model (HMM): a statistical model in which the system being modeled is assumed to be a Markov process warning (EEW) (e.g., Chin et al. 2020). The effectiveness of deep-learning-based detectors is largely due to convolutional layers, which perform cross correlation of the input waveforms and short template functions such that template features are learned from data. Successive convolutions eliminate the need for manually computing features and ensure that the extracted features are optimized.

Deep-learning-based preprocessing (denoising) or post-processing approaches can improve detection performance for low SNR data. Xiao et al. (2021) proposed post-processing for single-station deep-learning detectors based on the idea of template matching in the feature domain. In this approach, after application of a pretrained deep-learning model to continuous data, the extracted features at different layers of the model are used for multiscale (i.e., multiple layers of a neural network with different kernel sizes) representation of template P and S waves in the feature domain. A normalized version of these feature-domain templates is used to search for initially missed picks on the surrounding stations around the detection time by estimating their feature similarities using a Siamese neural network. This strategy improves the performance for low SNR waveforms and highlights the potential of more robust similarity measurements in a new domain (feature space).

Siamese neural network: a neural network that uses the same weights while working in tandem on two different input vectors to compute comparable output vectors

Encoder: any ML system that converts from a raw, sparse, or external representation into a more processed, denser, internal representation

Decoder: any ML system that converts from a processed, dense, or internal representation to a more raw, sparse, or external representation

2.3. Seismic Phase Picking

Once an earthquake is detected, the arrival times of distinct seismic phases (P wave and/or S wave) need to be measured, or picked, for subsequent analyses. Phase picking is similar to event detection, but on a smaller (narrower) scale. In addition to minimizing false positive and false negative rates, the temporal resolution and precision of arrival time picks are crucial for effective phase picking. Phase picking for important events is still often performed manually by human analysts, and it is traditionally performed on the waveform of detected events, so shorter lengths of seismic data are processed compared to detection. It is a time-consuming task that requires experienced analysts, and the amount of work can be overwhelming during swarms and aftershock sequences when information flux is high. Statistical and wavelet-based approaches are classical approaches for phase pickers (Lomax et al. 2012, Ross & Ben-Zion 2014, Mousavi et al. 2016b); however, sensitivity to noise, systematic bias, and difficulties in picking emerging S waves are common issues in conventional algorithms.

Pioneering AI phase picking methods used HMMs (Hammer et al. 2012) and shallow neural networks (Dai & MacBeth 1995, Wang & Teng 1997), but recent methods employ deep hybrid neural networks of up to several tens of layers (Ross et al. 2018b, Zhu & Beroza 2019, Mousavi et al. 2020, Liao et al. 2021). Multistation approaches (e.g., Yang et al. 2021, Yano et al. 2021, Zhu et al. 2022b) could reduce false positives but may be less sensitive to smaller events recorded by only a few stations within a network, and their ability to generalize is unclear.

Single-station models (Ross et al. 2018b, Zhu & Beroza 2019, Mousavi et al. 2020, Liao et al. 2021) can achieve high generalization due to the availability of large training data volumes. Models pretrained by data in a specific setting perform well in other settings such that deep-learning pickers can be used without retraining. The availability of huge archives of seismic data and labels from earthquake catalogs is particularly beneficial. The neural network architecture, neural network type, and size of a training set are not the only defining factors for good generalization (Mousavi et al. 2020). The hyperparameter tuning (e.g., Soto & Bernd 2020), training procedure, and data augmentation (e.g., Zhu et al. 2020) also play important roles. Transfer learning approaches are a viable option in cases with low generalization. For example, Lapins et al. (2021) used the convolutional part of the generalized phase detection model (Ross et al. 2018b) as the encoder of a new network and trained a new decoder using limited data to achieve good performance in a volcanic setting. There will always be a limit to generalization of deep-learning models

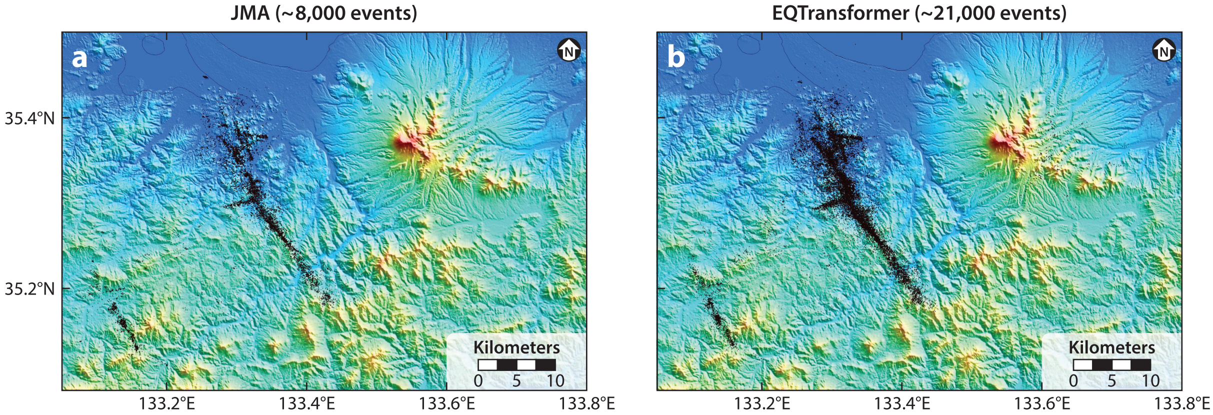

Figure 2

Precision seismic catalog of a Tottori, Japan, sequence, including ∼8,000 events manually picked by Japan Meteorological Agency (JMA) analysts (*a*) and ∼21,000 artificial intelligence–based events detected by EQTransformer (*b*).

to out-of-distribution cases. Thus, the question of which picker to use can best be addressed through conducting systematic benchmarking (e.g., Münchmeyer et al. 2022b). Open source frameworks such as SeisBench (Woollam et al. 2022) play a key role by facilitating benchmarking and cross-domain evaluations for model developers and users.

The AI pickers referenced above outperform traditional algorithms and achieve a similar picking accuracy to skilled analysts while picking far more phases in much less time. The power of AI pickers is most evident for the challenging tasks of picking S waves and making picks at low SNRs. Deep-learning detectors/pickers are already widely used in practice and have shown great performance in diverse environments. Due to their computational efficiency, these new-generation phase pickers are usually deployed without the event detection step and applied directly to continuous data.

Improving the detection/phase picking step in earthquake monitoring pipelines creates more complete earthquake catalogs and higher-resolution maps of seismicity than classical methods (**Figure 2**) due to detecting many more smaller events with weaker signals (e.g., Park et al. 2020, Gong et al. 2022, Li et al. 2021, Tan et al. 2021, Jiang et al. 2022). For example, Tan et al. (2021) were able to increase the number of events in the Amatrice-Visso-Norcia Italy earthquake sequence by a factor of 13. Recent packages use ML phase pickers extensively (e.g., Shi et al. 2022, Walter et al. 2021, Hao & Pascal 2022). To date, applications have focused on earthquake monitoring. These algorithms have yet to see much application for seismic velocity modeling via travel time tomography (e.g., Wei & Zhao 2022).

Despite this progress, the full potential of ML for improving earthquake monitoring has not yet been realized. AI phase associators, event locators, magnitude estimators, etc., could further improve the quality of catalogs by providing the higher sensitivity/accuracy required for processing larger numbers of small events near the detection threshold. Extending ML approaches to regional and teleseismic ranges by including arrivals beyond direct phases (e.g., head waves, reflections, etc.) is an obvious direction for future development.

The high efficiency of ML-based detectors makes them viable for real-time monitoring, including EEW; however, more research is needed to establish their accuracy and reliability. Incorporating data uncertainty into the modeling procedure and quantifying uncertainties in the output estimates are additional important factors for their operational use. Compared to discrimination, baseline methods for detection and phase picking are clear and widely used; however, adapting more unified performance reporting and benchmarking would help assess progress in ML applications for these tasks. Compared to the discrimination task, the baseline methods for detection and phase picking are more clear and widely used. There are, however, still some discrepancies among researchers calculating true-positive and false-positive picks. Adapting to more unified performance reporting and benchmarking would help progress in deep-learning applications for these tasks.

Graph theory–based clustering: the data of a clustering problem represented as a graph where each element to be clustered is represented as a node and the distance between two elements is modeled by a certain weight on the edge linking the nodes

Gaussian mixture model: a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters

2.4. Polarity Determination

First arrival polarities are used to determine focal mechanisms and can also be used in diffraction-stack-based location for microseismic monitoring based on surface-array data. Application of ML for automatic first motion polarity determination has been relatively limited. Ross et al. (2018a) and Hara et al. (2019) used a supervised approach to train a CNN using manually determined polarities. Mousavi et al. (2019b) approached the polarity determination problem as a clustering task and showed that similar performance could be achieved via unsupervised deep learning. Tian et al. (2020) expanded the supervised method to multiple stations and showed that using polarity information of neighboring sensors improves performance. This is a relatively straightforward task for DNNs, and these studies showed that deep-learning models can classify more polarities than analysts and without sacrificing quality, resulting in more, and more accurate, focal mechanisms. Uchide (2020) demonstrated the potential of this approach by using ML-picked polarities to determine the focal mechanisms of approximately 110,000 microearthquakes in Japan.

2.5. Phase Association

Phase association links arrival times at different stations to a common origin. It is a critical step in earthquake monitoring and can be challenging because of near-simultaneous occurrence of events in different parts of a network and because the number of contributing events is a priori unknown. ML approaches for association include both travel time–based and waveform-based methods.

2.5.1. Travel time–based methods. Travel time–based methods rely on travel time information for association. Ross et al. (2019) developed PhaseLink as a deep-learning-based associator using RNNs trained on synthetic phase arrival times. McBrearty et al. (2019) performed association using a pattern detection metric based on the principle of back-projection followed by graph theory–based clustering and an integer linear optimization routine to associate arrivals with the minimum number of required sources. This method solves for all sources and phase assignments simultaneously, rather than sequentially, as is common in other association methods. Yu & Wang (2022) used a neural network to accelerate grid search–based association. Zhu et al. (2022a) took an unsupervised approach by treating phase association as a clustering problem in a probabilistic framework, where each earthquake corresponds to a cluster of P and S phases with hyperbolic arrival time moveout and a decay of amplitude with distance. Incorporating the amplitude into the association problem enables an estimate of earthquake location, origin time, and magnitude using the Gaussian mixture model.

2.5.2. Waveform-based methods. Waveform-based methods search for waveform similarity from proximal earthquake sources in a low-dimensional feature space. These methods are similar to Siamese neural networks used in computer vision for facial recognition. McBrearty et al. (2022) trained a CNN to take pairs of earthquake waveforms around phase arrival times at two stations to predict whether they are from a common source. Dickey et al. (2020) employed a temporal CNN to map raw seismograms to a low-dimensional embedding space, where nearness in the space

Classification: the task of predicting a discrete class of labels

Regression: the task of predicting a continuous quantity

corresponds to similarity of the source function regardless of path or recording instrumentation. This path-agnostic embedding space forms a new representation for seismograms, characterized by robust, source-specific features, which is useful for performing both pairwise event association and template-based source discrimination.

Both approaches show promising results. While travel time–based approaches have superior performance, models may not generalize to other regions or network geometries without retraining. Waveform-based models may benefit from a higher generalization and can be used for tasks beyond association.

2.6. Earthquake Source Parameterization

Classical methods for hypocentral estimation can be divided into efficient travel time inversion methods that require precise arrival time picks on multiple stations and more computationally expensive grid search–based methods that utilize the full waveform. Both approaches rely on wave propagation through a known velocity model. Similarly, focal mechanism/moment tensor inversion approaches based on full waveforms are computationally more complex than amplitude- or polarity-based methods and are often limited to $M > 4.0$ earthquakes that have longer period waves that can be simulated reliably. Traditional magnitude estimation (e.g., local magnitude) requires known earthquake location, attenuation, and site amplification models. When applied to EEW these methods require rapid parameter estimation from a limited amount of observed data, which makes source estimation more challenging. Despite these challenges, the potential of ML for providing timely and reliable earthquake source parameters has attracted researchers' attention.

Early work includes that of Bose et al. (2008), who trained a two-layer neural network to estimate the hypocenter, moment magnitude, and rupture expansion. This multistation approach provides a rapid estimate of source characteristics for complex ruptures, which can play a key role in EEW. Käufl et al. (2016) used an ensemble of neural networks to learn the mapping from strong motion observations to marginal posterior probability density functions of earthquake source parameters (location, moment tensor components, and source half duration), which are parameterized by mixtures of Gaussians. In this approach, the inverse relationship is learned from synthetic 3D Green's functions. By applying their method to an event in the Los Angeles Basin, they showed that the location and magnitude can be estimated as early as 14 s after the origin time, while more data and time (40 s) are required to obtain a reliable estimate of fault strike and source depth. Modern counterparts of these pioneering ML approaches can be categorized into several groups.

2.6.1. Single-station methods. Lomax et al. (2019) investigated the feasibility of rapid earthquake characterization from single-station data using DNNs. They formulated it as a classification problem and used a multi-input CNN to classify three-component (3C), seismic waveforms into a larger number of discrete classes including event noise, station-event distance, station-event azimuth, event magnitude, and focal depth; however, they reported a high prediction error rate on a test set. Mousavi & Beroza (2020a) treated single-station location estimation as a set of smaller regression problems and approached each using a specialized Bayesian CNN. One network is used to learn the relationship between 1 min 3C full waveforms with epicentral distance and P travel time, and another network learns to estimate back azimuth from 1.5 s of P wave and the corresponding covariance matrix and eigenvectors. These predictions along with estimated epistemic and aleatory uncertainties are used to compute the epicenter, origin time, and depth, and their associated confidence intervals. Ristea & Radoi (2021) improved single-station-based estimates of epicentral distance, depth, and magnitude by using time-frequency representations of 3C waveforms as input data and utilizing a CNN capable of handling complex numbers. Mousavi & Beroza (2020b) used a simple network architecture based on a combination of CNNs and RNNs to overcome the common issue of amplitude normalization for deep-learning-based magnitude estimation and found that individual 3C waveforms contain enough information to learn a universal model of attenuation and site amplification to estimate magnitude.

Graph neural network: neural model that captures the dependence of graphs via a message passing between the nodes of graphs

2.6.2. Multistation methods. Another group of ML approaches for earthquake source parameterization rely on seismic array observations. The pattern of seismic energy propagation across a seismic network provides useful information for source characterization. Lin et al. (2021) used this to address the challenge of underestimating the magnitude of large earthquakes ($M_w > 8$) in EEW systems. They proposed a multistation approach that was trained on simulated ground displacement for a set of rupture scenarios in a particular region (e.g., the Chilean subduction zone). The neural network was used to estimate earthquake magnitude based on the patterns of crustal deformation in real time. Applying their model to five historical events in Chile, they were able to estimate the magnitudes of three events with reasonable accuracy within 60 s of the origin time. Licciardi et al. (2022) presented a novel approach for EEW in which a DNN is trained to estimate earthquake location and time-dependent magnitude from the speed-of-light prompt elasto-gravity signals. They found that the model trained on synthetic data could estimate the 2011 Tohoku earthquake's magnitude in about 50 s, which is faster than state-of-the-art EEW systems.

Kriegerowski et al. (2019) introduced a multistation regression for end-to-end learning of earthquake locations in which seismic waveforms recorded on a fixed number of ordered stations, and aligned based on the P-wave arrival time at a reference station, are used for hypocentral estimation. Approximate knowledge of location of the stations is implicit in the input data structure, and the trained neural network is not directly transportable to different station configurations. Zhang et al. (2020) proposed a similar multistation approach, but in a classification framework for which a 3D grid of a region of interest is represented by 3D kernels at the output layer of a CNN in which the closest grid point to the ground truth hypocenter is labeled by a 3D Gaussian function. The observed waveforms are directly classified into a predefined grid space based on the output probabilities. A downside of this approach is that the resolution of the location estimates is bounded by the dimension of the output layer of the neural network (the number of grid points or classes) and the total number of layers in the neural network. In such a many-class approach, the large input and output dimensions increase the nonlinearity of the learning system, which poses challenges for training with a limited number of multistation labels.

2.6.3. Dynamic multistation methods. To overcome these limitations, Van Den Ende & Ampuero (2020) proposed a multistation approach using an edgeless graph neural network. They used a CNN to encode 3C waveforms at each station and the geographic coordinates of the station into a feature vector. Hence, the absolute locations of seismic stations are explicitly incorporated into the modeling. These feature vectors represent station locations and their observed data on a seismic network as a graph. They are concatenated by maximum pooling to form a single feature representation of the wave propagation from each source across the seismic network. This network-level feature vector is mapped to the latitude, longitude, depth, and magnitude of the event by another neural network. This framework provides a flexible multistation approach for analyzing the data from seismic networks, as it is invariant to the number and ordering of the seismic stations. Münchmeyer et al. (2021a) proposed a dynamic framework based on attentive neural networks to expand this approach to varying input data window length and to incorporate the interstation dependency in the feature aggregation step. In this approach the representations are

Transformer: a neural network architecture that relies on self-attention mechanisms to transform a sequence of input embeddings into a sequence of output embeddings without relying on convolutions or recurrent neural networks

Physics-informed neural network (PINN): a neural network that is trained to solve supervised learning tasks while respecting any physical laws described by general nonlinear partial differential equations

Reinforcement learning: a family of algorithms that learn an optimal policy, whose goal is to maximize return when interacting with an environment; reinforcement learning systems can become expert at playing complex games by evaluating sequences of previous game moves that ultimately led to wins and sequences that ultimately led to losses

Fully connected layer: a hidden layer in which each node is connected to every node in the subsequent hidden layer

aggregated using the transformer to form a network-level feature representation by incorporating contextual information and interstation dependencies. This sparse representation of the entire seismic array and observations across it is used by another neural network to predict the parameters and weights of a mixture of Gaussians for the hypocentral location and magnitude. Another attentive multistation approach for earthquake location is presented by Chin et al. (2021) based on arrival time picks. In this travel time–based approach, the transformer architecture is used to generate a new feature representation of the seismic network and P-wave detection times with respect to a reference station. This representation is enriched by the mutual relationships among station-observation pairs in a high-dimensional feature space that is used to learn the relationship between the arrival time observations and the hypocentral location using an interpolation network. The sequential nature of seismic wave propagation in time and space is the main motivation behind the application of a transformer network in these studies.

2.6.4. Physics-informed methods. In the studies mentioned above, DNNs were used to learn the nonlinear inverse relation between the observed seismic data and the source parameters directly from data. Smith et al. (2022) developed a physics-informed neural network (PINN) to infer earthquake hypocenters from observed arrival times in a Bayesian framework. A DNN is used as a differentiable reparameterization of the Earth model for forward modeling of a fast and mesh-free calculation of the theoretical seismic wave travel times between any two points. This efficient forward modeling is paired with Stein variational inference to approximate a target posterior. Wu et al. (2019) used reinforcement learning to search for hypocenters by reconstructing a 3D seismic wavefield based on sparse observed 1D ground motions and searching over source locations to explain the reconstructed wavefield.

2.6.5. Earthquake source mechanism. Applications of ML for focal mechanism estimation are relatively new compared to other source parameter inversion tasks. Kuang et al. (2021) presented a multistation approach similar to that by Kriegerowski et al. (2019) and Zhang et al. (2020) for rapid earthquake focal mechanism estimation using a CNN. They generated a large number (790,000) of synthetic full waveforms based on source parameters with a variety of focal mechanisms for the Ridgecrest region in Southern California that they used to build a model that they subsequently applied to $M_w > 5.4$ events in the Ridgecrest sequence. Steinberg et al. (2021) proposed a probabilistic deep-learning approach for near real-time estimation of an ensemble of moment tensors and their uncertainties from short time windows of seismic waveforms. In this approach, individual Bayesian neural networks are trained for each grid point in a 3D regional mesh using synthetic seismic waveforms and moment tensor parameters. They used variational inference to quantify model uncertainties in location and time, as well as in the Earth model. Zhang et al. (2021) developed a physics-guided neural network for estimation of shear-tensile focal mechanisms for multiple microearthquakes using the displacement amplitudes of P waves on shallow-borehole arrays. In this approach, the learning is unsupervised. A neural network composed of fully connected layers and a scaling-shifting layer is used to learn the mapping from observed amplitudes to shear-tensile focal parameters (strike, dip, rake, and slope). Next, a radiation pattern and normalization layer are used for forward modeling and mapping the inferred shear-tensile parameters to synthetic P-wave amplitudes.

ML applications for source characterization are in a more formative stage than those for other aspects of seismic monitoring. While multistation approaches are similar to conventional methods, they suffer from smaller training data (usually a few thousand events) and lower generalization. Single-station approaches are in an experimental phase of exploring the possibilities of maximizing information from small units of observations, which has the potential to improve the monitoring of smaller events and the rapid characterization of larger events, while using a limited number of observations and benefiting from more abundant training data.

In the above studies the true source properties for field data are unknown, and the accuracy of estimates in earthquake catalogs depends on the quantity and quality of observations and algorithms, such that the accuracy of the catalogs used for the labeling and training of DNNs and the distribution of the events used for testing the model play a key role in performance. ML detectors trained on high-precision catalogs and tested on a local cluster of events (Kriegerowski et al. 2019) can result in much lower prediction error (on the order of 10 m for epicenter estimation) compared to those trained on standard catalogs and tested over wider regions (e.g., Van Den Ende & Ampuero 2020). Thus, it is important to account for labeling uncertainties and the width of test distributions when comparing different models. Incorporating aleatory and epistemic uncertainties into the modeling procedure and employing benchmark data sets for performance testing could accelerate progress. Data-driven approaches for deep learning of magnitude estimation suffer from underrepresentation of large ($M > 6.0$) events in the training sets arising from the frequency-magnitude distribution of earthquakes. This can manifest as magnitude underestimation by ML models. Synthetic data generation for particular regions and rupture scenarios is an option; however, generalization of such models beyond their training distribution remains unclear due to difficulties in generating realistic realizations that span a complex, multivariate space. Moreover, the limited number of large events in field data makes it difficult to evaluate the performance of these models. Incorporating physical constraints into the model building procedure through PINNs is a promising path to overcome these challenges.

Generative adversarial network (GAN): a system to create new data in which a generator creates data and a discriminator determines whether the created data are valid or invalid

2.7. Seismogram Simulation

Theoretical seismograms are widely used to interpret data—and to train deep-learning models. Discretizing and iteratively solving the wave equation using numerical methods such as by finite differences or spectral elements is the most common approach for seismic wave simulation. The computational cost of these methods grows very rapidly with increasing frequency, which is problematic when the solution domain is large. Although applications of deep learning for synthetic seismic signal generation are at an early stage, they may offer an effective alternative for overcoming some limitations of numerical methods.

Neural networks have been used to generate theoretical ground motion accelerograms and response spectra (Lee & Han 2002) and to extend ground motion simulations to short periods (Paolucci et al. 2018). Initial attempts to use DNNs for seismic wave simulation were end-to-end (e.g., Moseley et al. 2020b), but PINNs and generative adversarial networks (GANs) subsequently emerged for this application.

2.7.1. Physics-informed models. PINNs are an effective tool to learn the solution of the wave equation in both time (Moseley et al. 2020a) and frequency (Song et al. 2020) domains. For PINNs, a neural network is trained to solve the wave equation for a medium that is implicitly defined in a loss function composed of either a form of the wave equation (Song et al. 2020) or a combination of a boundary condition and the wave equation (Moseley et al. 2020a). Moseley et al. (2020a) designed a PINN that takes a source location as its input and outputs an approximation of the pressure wavefield at another location. By training the network for many different source locations, it is possible to generalize to new source locations without retraining. Their model learned to solve the time-domain wave equation in complex 2D media and generalized well beyond the timestamps of their training data set. Song et al. (2020) used a variant of the wave equation for transversely isotropic media with vertical symmetry as the PINN's loss function and solved for the scattered pressure wavefield rather than the wavefield itself. They used their model to obtain frequency-domain acoustic solutions for a fixed source location in 2D and 3D, and in the presence of irregular topography. PINNs are much more efficient in computing the wavefield compared to full numerical simulations and can reduce the computational time required by several orders of magnitude.

Random forest: an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time

2.7.2. Generative models. The second group of deep-learning approaches for seismogram simulation uses GANs. These models are composed of two networks: a generator that generates fake data and competes against a discriminator that discriminates real from fake samples. This is based on the idea of using the universal function approximation ability of a neural network to learn the probability distributions of attributes of a training data set by optimizing a generator network. This generator model can then be used for reparameterization of distributions to generate new samples drawn from the learned distributions. GANs have been used to generate synthetic earthquake and non-earthquake seismograms as a data augmentation tool in training of deep-learning earthquake detectors (Wang et al. 2021); to generate broadband seismic signals by blending the low-frequency outcome of physics-based numerical simulations with sparsely sampled broadband observations (Gatti & Clouteau 2020); to generate 3C strong motion time series for different magnitude, distance, and VS30 [the time-averaged shear-wave velocity (VS) to a depth of 30 meters] values (Florez et al. 2022); and to refine computer simulations of DAS data for noise ground motions (Shiloh et al. 2019). GANs provide an efficient framework for generating large-scale synthetic training data that can improve the performance of deep-learning classifiers. Both generative and physics-informed approaches seem promising for efficient synthetic seismic data generation.

2.8. Ground Motion Characterization

ML has been used to estimate ground shaking intensity measures such as peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement, cumulative absolute velocity (CAV), pseudo-spectral acceleration (PSA), and Arias intensity, as well as seismic wave travel times.

2.8.1. Conventional machine learning–based models. Strong ground shaking intensity (e.g., PGA) is usually modeled based on a predefined functional relationship of a set of discrete attributes such as source magnitude, source-to-site distance, and local site conditions, in a parametric regression. These ground motion models (GMMs) are used for seismic hazard assessment, EEW, and structural engineering. ML methods have been shown to be effective for developing nonparametric GMMs, especially using neural networks where the network weights represent the model coefficients, using a fixed set of similar input parameters (Kerh & Ting 2005, Derras et al. 2014, Trugman & Shearer 2018, Wiszniowski 2019, Withers et al. 2020, Wang et al. 2020).

Derras et al. (2014) used neural networks to develop a ground motion prediction equation (GMPE) for Europe in a model with only five inputs—magnitude, Joyner-Boore distance, focal mechanism, depth, and VS30—and outputs of PGA, PGV, and 5% damped PSA at periods from 0.01 to 4 s. They showed that neural network models are competitive in accuracy with conventional GMPEs. A random forest approach was shown to be an effective ML method for building GMPE models (e.g., Trugman & Shearer 2018, Khosravikia & Clayton 2021). Hu et al. (2022) successfully applied an SVM to develop GMMs of Arias intensity, CAV, and significant duration. ML-based GMPE models work best when plentiful high-quality data are available, which presents a challenge because near-source records of large-magnitude earthquakes, which pose the greatest hazard, are sparse. Withers et al. (2020) used neural networks to build a GMPE from a database of ground motion simulations. They found that for the Ridgecrest event, the neural network–based GMPE has similar bias and variability to empirical GMPEs. ML-based GMPEs may prove to be powerful new tools in the next generation of seismic hazard models. The ML models described above have similar input-output and similar performance as traditional models, although they have a greater capacity to uncover nonlinear relationships.

2.8.2. Novel approaches. AI enables new models that provide rapid estimates of ground shaking directly from the observed wavefield without earthquake source parameter or site condition estimates. These end-to-end ML models are particularly useful for EEW.

Hsu & Huang (2021) used a CNN to predict the expected peak shaking at a station based on the first 3 s of waveform data observed at the station. This approach accurately predicted PGA for earthquakes that did not have a long and complex rupture process. Jozinović et al. (2020) proposed a deep-learning approach for a rapid PGA and PGV estimation at a fixed set of stations directly from a short window (10 s) of a 3C acceleration waveform. They suggested that the network-level pattern of ground motion propagation contains enough information for a deep CNN to nowcast ground shaking levels at more distant stations that have not been shaken based on those stations that have been. The proposed CNN model shows similar variance but smaller bias compared to traditional models. Münchmeyer et al. (2021b) developed a similar end-to-end approach to estimate PGA probability densities at locations of interest based on the observed waveforms at a set of seismic stations (**Figure 3**). Their approach is flexible enough to work on an arbitrary number of stations at arbitrary locations by implementing station-level feature extraction and PGA prediction. They used a transformer architecture to integrate multistation observations by explicitly encoding station locations and learning network-level dependencies between observation and target stations in the feature domain. They found that their method outperformed existing EEW methods and that it could handle large events that are normally underrepresented in the training set through domain adaptation.

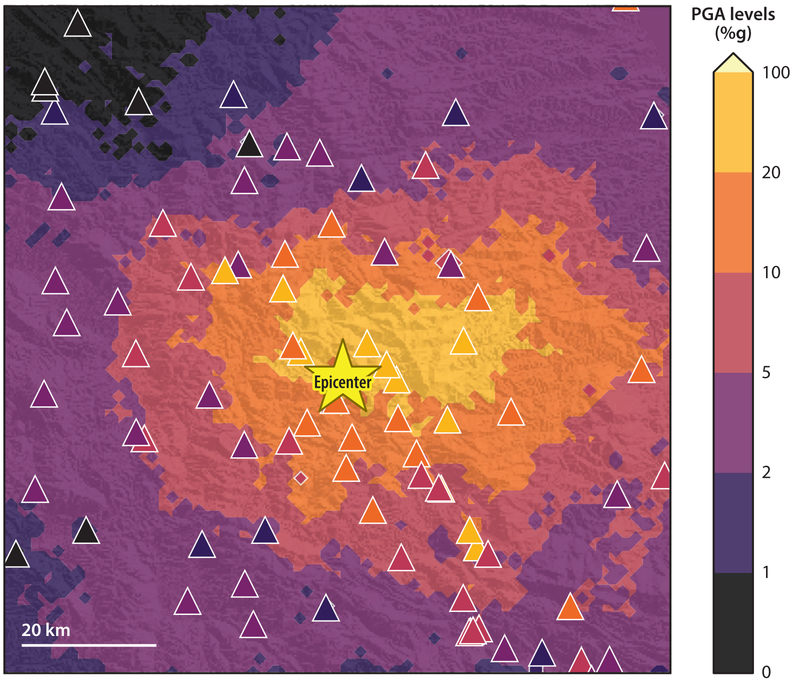

Figure 3

Predicted peak ground acceleration (PGA) levels for 3.5 s after P arrival (4.5 s after origin time) by the transformer earthquake alerting model (Münchmeyer et al. 2021b).

The power of RNNs for modeling time-series data makes them appealing for nowcasting ground motion intensity. Otake et al. (2021) used a deep RNN for real-time estimation of seismic intensity at target locations based on the observed ground motions at a sparse set of nearby stations. The ability of RNNs to deliver fast and robust alerts has been demonstrated by Wang et al. (2022) for onsite EEW (estimating upcoming destructive S waves based solely on observed initial P waves). Similarly, Datta et al. (2022) developed a network-based model for forecasting future shaking intensity at a set of stations based on observed ground shaking in a previous time window.

2.8.3. Eikonal solver models. The information in complex waveforms is often reduced to a set of arrival times for direct geometric rays. The Eikonal equation is used to estimate the travel time of high-frequency waves that propagate through a known medium and has many applications in seismology, and PINNs have shown great promise. Smith et al. (2020) proposed a PINN framework to learn the solution of the Eikonal equation for a 3D velocity model. The trained network can be used for rapid, grid-free estimation of the travel time between source and receiver locations in the continuous 3D domain for which the model has been built. A loss function based on a factored form of the Eikonal equation is used implicitly to reparameterize the 3D slowness field by the neural network weights and learn the travel time field. A similar approach was developed by bin Waheed et al. (2021) but for 2D velocity grids and a fixed source location. Overall, AI Eikonal solvers could have much lower space and time complexity compared to the traditional approaches (especially for complex structure), as they eliminate the need for travel time lookup tables and benefit from parallel computing on deep-learning platforms.

2.9. Earthquake Forecasting

Forecasting (days/months/years in advance) of future mainshock characteristics is a challenging task. While there are debates on its feasibility (Geller et al. 1997), earthquake forecasting represents a compelling subject for ML applications. ML models for earthquake forecasting usually take space-time discretized seismicity as their input and future mainshock characteristics, such as event magnitude in a time or space-time window, as their output.

Panakkat & Adeli (2007) used neural networks for predicting the magnitude of the largest seismic event in the following month by analyzing eight seismicity indicators. Adeli & Panakkat (2009) used similar indicators to train a neural network to predict the magnitude of the largest earthquake in a predefined future time period. The structured and sparse nature of seismicity indicators may not warrant deep-learning approaches; hence, recent neural network–based forecasting approaches have concentrated on more direct incorporation of seismic catalogs and their spatiotemporal structures in the model building procedure (Feng & Fox 2020, Wang et al. 2017). Although still in an experimental stage, this approach utilizes the feature extraction ability of deep learning architectures, and their ability to learn the dynamics of data could shed new light on this long-standing problem.

The infrequency of large earthquakes (on the order of a century between them even on active plate boundaries), limited history of instrumental recording (also on the order of a century), and limited instrumental coverage conspire to pose severe challenges to ML approaches. In the laboratory setting, however, fault-slip prediction has seen significant progress. Rouet-Leduc et al. (2017) applied a random forest to data sets from shear experiments with the goal of identifying hidden signals preceding laboratory earthquakes. They showed that the acoustic signal emitted by a laboratory fault zone that was previously thought to be low-amplitude noise contains information that can be used to forecast the time to the failure. Rouet-Leduc et al. (2019) applied this approach to seismic data recorded in the Cascadia subduction zone and showed that low-amplitude signals are continuously being emitted that inform the occurrence of slow-slip events. They suggested that this provides indirect real-time access to fault physics on the down-dip portion of the megathrust and may prove useful in determining if and how slow-slip may evolve into a large earthquake.

Recent advances in ML-based earthquake monitoring and development of more complete earthquake catalogs with unprecedented spatiotemporal resolution could lead to an improvement in both statistical and physics-based forecasting, as well as AI-based approaches. The existence of a preparatory phase preceding earthquakes has been suggested by laboratory, field, and theoretical studies, which suggests more complete seismicity catalogs enriched by smaller events should be helpful (e.g., Picozzi & Iaccarino 2021). Mignan (2014) argued that preparatory phases are potentially identifiable in real-world seismicity if seismic events more than three orders of magnitude lower than the mainshock are detected and characterized in a seismicity catalog. Hence, the predictive value of ML-based approaches using newly uncovered small earthquakes in ML-based catalogs deserves attention (Beroza et al. 2021). A key challenge is effective exploration of these complex catalogs and extraction of new relationships within them. This is a task for which deep learning should be well suited. Picozzi & Iaccarino (2021) modeled the preparatory phase of moderate induced earthquakes ($M \sim 4$) at the Geysers geothermal field in California using an RNN and a set of seismicity features computed from an earthquake catalog. They found that ML revealed the preparation of $M \sim 4$ earthquakes through learning the spatiotemporal evolution of microseismicity. ML offers new opportunities in earthquake forecasting by blending different potential precursory data and known earthquake physics with seismicity data and underlying physics that remain unknown.

Self-organized mapping: an unsupervised ML technique used to produce a low-dimensional representation of a higher-dimensional data set while preserving the topological structure of the data

K-means: a simple and popular clustering algorithm that groups examples in unsupervised learning

Manifold learning: an approach to nonlinear dimensionality reduction

Scattering neural network: a class of designed CNNs with fixed weights

2.10. Machine Learning for Exploratory Research

2.10.1. Direct investigation of seismic waveforms. ML is a powerful tool for exploratory data analyses and finding patterns directly from seismic data. Not surprisingly, unsupervised methods predominate in this domain. Köhler et al. (2009) used an unsupervised feature learning approach based on a self-organized mapping neural network to investigate temporal patterns in seismic single-station or network recordings. Chamarczuk et al. (2020) used a hybrid approach of array-processing, k-means, and feature extraction for unsupervised processing and clustering of seismic ambient noise data. This helps sort detected events into an optimal number of classes and facilitates the selection of desired parts of the wavefield for imaging and monitoring. Holtzman et al. (2018) used a combination of supervised (HMM) and unsupervised (k-means) ML techniques to reveal patterns in time-dependent spectral properties of seismic signals indicating changes in faulting processes. ML provides a powerful tool to identify and characterize such subtle changes in source properties. An unsupervised graph-based manifold learning was used to order diffracted seismograms along the core-mantle boundary and to obtain a panoptic view of scattered seismic waves across the Pacific region by D. Kim et al. (2020).

Unsupervised/self-supervised DNNs have shown promise for similar tasks. A combination of a deep scattering neural network and a Gaussian mixture model was used to reveal repeating seismic features preceding a landslide in Greenland (Seydoux et al. 2020) and to cluster daily quasi-periodicity and transient microevents on continuous seismic data recorded on Mars (Barkaoui et al. 2021). A similar approach has been used to cluster and identify the sources of urban seismic noise (Snover et al. 2021) and to identify dominant types of impulsive seismic signals observed in an ice shelf (Jenkins et al. 2022). In this approach, features that represent distinct clusters most effectively are automatically extracted from input data using a deep CNN through simultaneous optimization of clustering and feature-learning tasks. Other unsupervised approaches (e.g., Yoon et al. 2015, Mousavi et al. 2016b) are well suited for exploratory analyses of high-dimensional seismic data.

Logistic regression: a classification model that uses a sigmoid function to convert a linear model's raw prediction into a value between 0 and 1

Bayesian network: a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph

2.10.2. As a post-processing tool.

Aden-Antoniów et al. (2022) presented an ML approach for effective declustering of earthquake catalogs by classifying background seismicity and aftershocks. In this approach a random forest model was trained on various synthetic catalogs generated by an epidemic-type aftershock sequence model and used to decluster real seismicity catalogs. Unsupervised ML has been used for clustering of events in earthquake catalogs and unveiling the 3D rupture morphology of faults (Brunsvik et al. 2021, Gong et al. 2022). McKean et al. (2019) used a Gaussian mixture model with some physical constraints (fracture length based on an analytical diffusion-based model) as a probabilistic clustering approach to identify patterns and fracture networks from microseismic point clouds. By applying their approach to a hydraulic fracturing site in western Canada, they were able to distinguish hydraulically fractured events from those triggered on preexisting faults in the region.

2.10.3. As a research tool.

ML can be used as an analytic framework to study and explore hypotheses or relationships. Pawley et al. (2018) trained a logistic regression model to evaluate tectonic, geomechanical, and hydrological proxies that influence induced seismicity. Hincks et al. (2018) investigated the role of injection depth on induced seismicity using a Bayesian network that jointly models dependencies among spatial, operational, and seismicity parameters. They found that injection depth relative to basement most strongly correlates with seismic moment release. Larson et al. (2020) modeled the frequency of induced seismicity in Oklahoma using Euclidean distance from earthquakes to the nearest disposal well, the nearest fault, and average fluid injection rate to develop nonparametric regressions. They found that proximity to wastewater disposal sites, fluid injection rates, and adjacency to subsurface faults were sufficient to model seismicity in north-central Oklahoma.

Neural networks can be used to study eventual rupture predictability and magnitude of large events based on the initial stages of rupture in a probabilistic framework (Münchmeyer et al. 2022a). Their results suggested universal initiation behavior for small and large ruptures. Corradini et al. (2022) used a DNN trained on synthetic data to estimate the rise time and rupture velocity along a fault from back-projection images. ML was used by McLellan & Audet (2020) to explore the causal relationships between geophysical data reflecting physical properties of subducting plates and the occurrence of slow-slip events at subduction zones. They used multiple supervised ML algorithms to test the predictive power of features such as subducting plate age, sediment thickness, relative plate velocity, slab dip, and plate surface roughness for predicting the occurrence of slow-slip events within 25-km-wide, trench-parallel segments for circum-Pacific subduction zones. They found strong negative correlations between subducting plate age and relative velocity with the short-term slow-slip events, while sediment thickness correlates positively.

3. DISCUSSION

When neural networks were first applied to seismology in the late 1980s, the focus was on classification tasks in automatic data processing with supervised methods. Although data processing tasks still comprise a major part of the latest surge of ML applications in seismology, inverse problems are gaining rapidly (**Figure 4**). **Figure 4** shows an apparent downturn in the total number of publications in the two most recent years. The tally for 2022 is for only part of the year. Also, the figure includes publications on seismological ML model development, rather than ML applications. The decrease in the number of ML models published in seismology in 2021 appears to be real and can also be observed in the number of conference presentations. Some caveats are that conference presentations might be influenced by the pandemic and that while we have tried to be comprehensive in our coverage of ML publications in seismology, our database may be incomplete. Time will tell whether this trend persists.

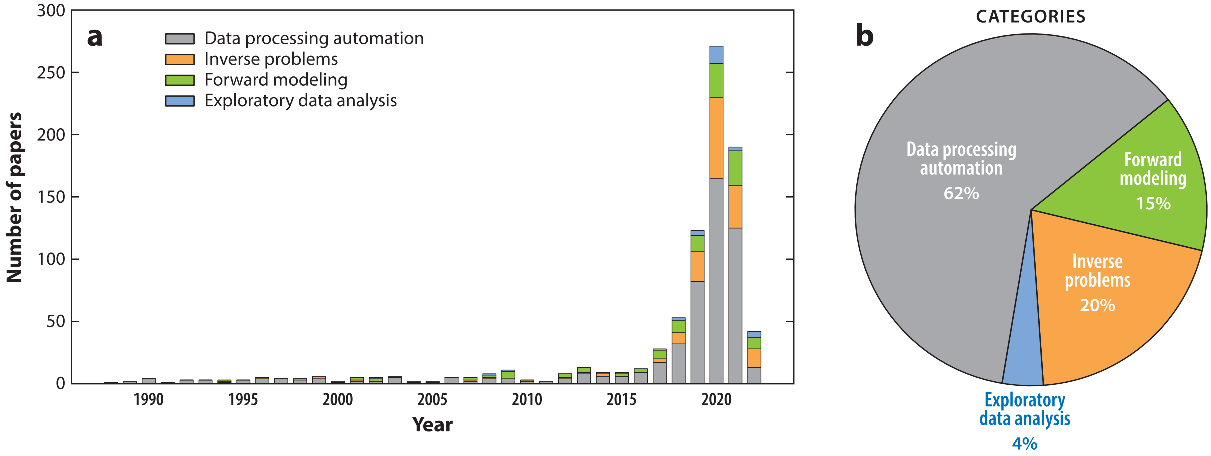

Figure 4

(*a*) Each bar shows the number of journal publications developing machine learning (ML) methods for a seismological task (including both earthquake and exploration seismology), published between January 1988 and May 2022. Bars are color coded based on category of seismological tasks. (*b*) A pie chart shows the share of seismological tasks for ML applications.

Supervised learning has dominated ML approaches in seismology to date; however, ML offers data-driven discovery, imaging, and interpretation of patterns in seismic data in a high-dimensional space as well. Seismologists are increasingly finding important applications for alternative approaches such as unsupervised learning and GANs, but exploratory analysis of high-dimensional seismic data has been only thinly investigated. We expect applications of unsupervised approaches in seismology will be a growth area (**Figure 5**).

The limitations on training data and generalization are the main challenges in solving inverse and forward problems using supervised DNNs. A common approach to mitigate this is to train a network on synthetic data and fine-tune it, or to perform transfer learning with field data.

Moving toward semisupervised approaches, where both labeled and unlabeled data are used for the training, is another likely growth area. PINNs, where the governing physical theories are imposed as constraints into data-driven ML models, are a promising direction for improved generalization. Real seismic data are often poorly sampled, noisy, incomplete, and unbalanced, all of which pose challenges for ML techniques. Combining data-driven ML methods with physical models could be transformative.

Neural networks are the main ML method used in seismological applications. Among the variety of neural network types, fully connected and CNNs are used the most. U-net and autoencoders form the most commonly used neural network architectures. Both of these have butterfly forms and are composed of an encoder that transfers the input data into a high-level but lower-dimensional representation and a decoder that generates the output (often with the same dimension as the input) from this low-dimensional representation. Encoder-decoder networks are clearly a highly suitable architecture for many seismological applications.

Event discrimination, detection, and phase picking are fairly mature applications. The emphasis now is shifting to earthquake characterization and exploratory data analyses. Even in well-studied applications, however, important issues are unresolved. For example, it is difficult to determine the relative performance, strengths, and weaknesses of each method due to a lack of standard benchmarking.

Labeled data sets are required for building supervised models or testing unsupervised models. Even for seismological tasks (e.g., earthquake detection and phase picking) where ample labeled

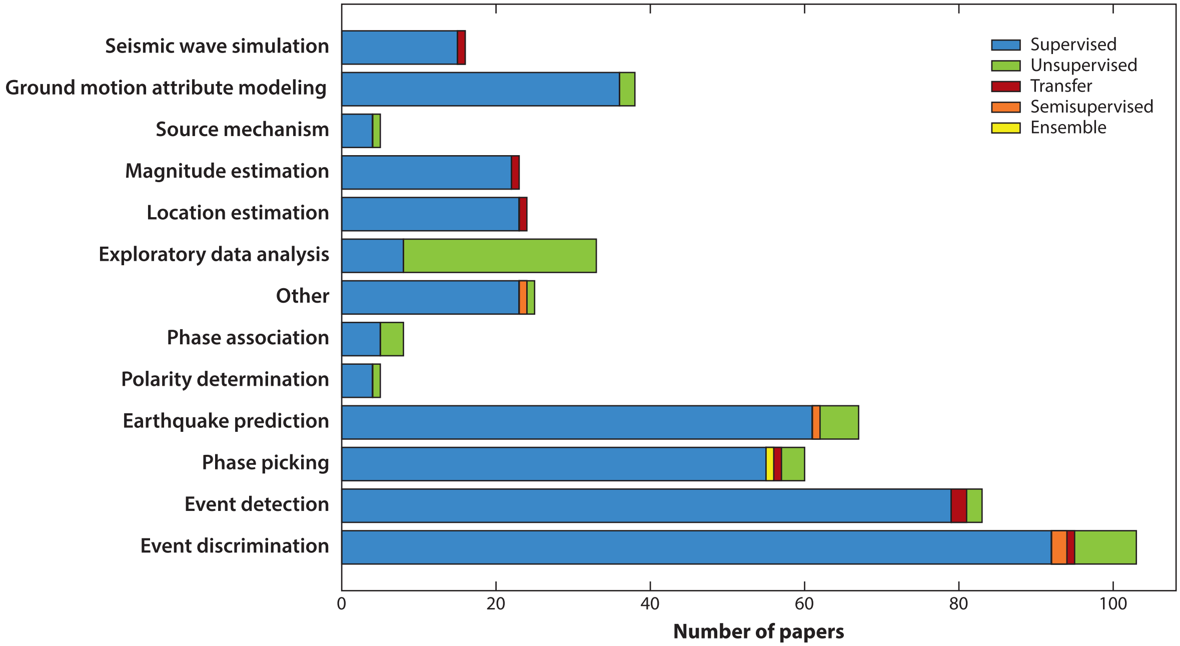

Figure 5

The ratio of various machine learning approaches used for each seismological task.

data exist, the reliability of those labels is variable and uncertain. Two analysts will differ in their measurement of the arrival time of a seismic phase in an earthquake signal, which introduces bias. A challenging task in building training data sets is quality control of the labels. There are only a few seismic data sets that can serve as benchmarks (e.g., Mousavi et al. 2019a, Michelini et al. 2021); however, these data sets are suitable for only some of the tasks we have outlined. Standard benchmark data sets can serve as ground truth and accelerate progress in application of ML methods. Efficient simulation methods for fast generation of synthetic data at a large scale could also play an important role.

Deep learning and more complex ML techniques have successfully improved the performance of some tasks; however, this does not guarantee their suitability for other problems and data types. Mignan & Broccardo (2020) and Albert & Linville (2020) showed that simpler ML and traditional methods can match the performance of deep-learning models for aftershock forecasting and infrasound classification, respectively. Simpler and more transparent methods that can be tied to the physical properties of the waveforms, yet provide a similar performance, are preferable to less interpretable approaches.

SUMMARY POINTS

1. Deep-learning detectors/pickers are already widely used in practice and have shown great performance resulting in more complete earthquake catalogs with unprecedented spatiotemporal resolution in diverse local environments.
2. More complete earthquake catalogs could lead to an improvement in both statistical and physics-based forecasting, and in artificial intelligence–based approaches.

3. Although data processing tasks comprise a major part of studies in the latest surge of ML applications in seismology, inverse problems are gaining rapidly.

FUTURE ISSUES

1. We expect future applications of unsupervised approaches for exploratory analysis of high-dimensional seismic data will be a growth area.
2. Semisupervised and physics-based neural networks are promising directions to improve model generalization in low training data situations.
3. Systematic benchmarking through open source frameworks and benchmark data sets can play a key role in evaluating the models and progress in the field.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank the Editorial Committee for the invitation and the reviewer for the insightful comments. We also thank Jannes Münchmeyer for providing a figure presenting an example of TEAM model performance for earthquake early warning. G.C.B. was supported by the Air Force Research Laboratory under contract FA9453-19-C-0073.

LITERATURE CITED

Adeli H, Panakkat A. 2009. A probabilistic neural network for earthquake magnitude prediction. *Neural Netw.* 22:1018–24

Aden-Antoniów F, Frank WB, Seydoux L. 2022. An adaptable random forest model for the declustering of earthquake catalogs. *J. Geophys. Res. Solid Earth* 127:e2021JB023254

Albert S, Linville L. 2020. Benchmarking current and emerging approaches to infrasound signal classification. *Seismol. Res. Lett.* 91:921–29

Barkaoui S, Lognonné P, Kawamura T, Stutzmann É, Seydoux L, et al. 2021. Anatomy of continuous Mars SEIS and pressure data from unsupervised learning. *Bull. Seismol. Soc. Am.* 111:2964–81

Beroza GC, Segou M, Mousavi SM. 2021. Machine learning and earthquake forecasting—next steps. *Nat. Commun.* 12:4761

Beyreuther M, Carniel R, Wassermann J. 2008. Continuous hidden Markov models: application to automatic earthquake detection and classification at Las Canãdas caldera, Tenerife. *J. Volcanol. Geotherm. Res.* 176:513–18

Beyreuther M, Wassermann J. 2008. Continuous earthquake detection and classification using discrete Hidden Markov Models. *Geophys. J. Int.* 175:1055–66

bin Waheed U, Haghighat E, Alkhalifah T, Song C, Hao Q. 2021. PINNeik: Eikonal solution using physics-informed neural networks. *Comput. Geosci.* 55:104833

Bose M, Wenzel F, Erdik M. 2008. PreSEIS: a neural network-based approach to earthquake early warning for finite faults. *Bull. Seismol. Soc. Am.* 1:366–82

Braeuer B, Bauer K. 2015. A new interpretation of seismic tomography in the southern Dead Sea basin using neural network clustering techniques. *Geophys. Res. Lett.* 42:9772–80

Brunsvik B, Morra G, Cambiotti G, Chiaraluce L, Di Stefano R, et al. 2021. Three-dimensional paganica fault morphology obtained from hypocenter clustering (L'Aquila 2009 seismic sequence, Central Italy). *Tectonophysics* 804:228756

Chamarczuk M, Nishitsuji Y, Malinowski M, Draganov D. 2020. Unsupervised learning used in automatic detection and classification of ambient-noise recordings from a large-N array. *Seismol. Res. Lett.* 91:370–89

Chin T-L, Chen K-Y, Chen D-Y, Lin D-E. 2020. Intelligent real-time earthquake detection by recurrent neural networks. *IEEE Trans. Geosci. Remote Sens.* 58(8):5440–49

Chin T-L, Chen K-Y, Chen D-Y, Wang T-H. 2021. An attention-based hypocenter estimator for earthquake localization. *IEEE Trans. Geosci. Remote Sens.* 60:5905510

Corradini M, McBrearty IW, Trugman DT, Satriano C, Johnson PA, Bernard P. 2022. Investigating the influence of earthquake source complexity on back-projection images using convolutional neural networks. *Geophys. J. Int.* 229:1824–39

Dai H, MacBeth C. 1995. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophys. J. Int.* 120:758–74

Datta A, Wu DJ, Zhu W, Cai M, Ellsworth WL. 2022. DeepShake: shaking intensity prediction using deep spatiotemporal RNNs for earthquake early warning. *Seismol. Soc. Am.* 93:1636–49

Derras B, Bard PY, Cotton F. 2014. Towards fully data driven ground-motion prediction models for Europe. *Bull. Earthq. Eng.* 12:495–516

Dickey J, Borghetti B, Junek W, Martin R. 2020. Beyond correlation: a path-invariant measure for seismogram similarity. *Seismol. Res. Lett.* 91:356–69

Dowla FU, Taylor SR, Anderson RW. 1988. Seismic discrimination with artificial neural networks: preliminary results with regional spectral data. *Bull. Seismol. Soc. Am.* 80:1346–73

Esposito AM, Giudicepietro F, Scarpetta S, D'Auria L, Marinaro M, Martini M. 2006. Automatic discrimination among landslide, explosion-quake, and microtremor seismic signals at Stromboli volcano using neural networks. *Bull. Seismol. Soc. Am.* 96:1230–40

Feng B, Fox GC. 2020. TSEQPREDICTOR: spatiotemporal extreme earthquakes forecasting for Southern California. arXiv:2012.14336v1 [physics.geo-ph]

Florez MA, Caporale M, Buabthong P, Ross ZE, Asimaki D, Meier M-A. 2022. Data-driven synthesis of broadband earthquake ground motions using artificial intelligence. *Bull. Seismol. Soc. Am.* 112(4):1979–96

Gatti F, Clouteau D. 2020. Towards blending physics-based numerical simulations and seismic databases using generative adversarial network. *Comput. Methods Appl. Mech. Eng.* 372:113421

Geller RJ, Jackson DD, Kagan YY, Mulargia F. 1997. Earthquakes cannot be predicted. *Science* 275:1616–17

Gong J, Fan W, Parnell-Turner R. 2022. Microseismicity indicates atypical small-scale plate rotation at the Quebrada transform fault system, East Pacific Rise. *Geophys. Res. Lett.* 49(3):e2021GL097000

Gong L-W, Zhang H, Chen S, Chen L-J. 2021. Three-dimensional modeling of the Xichang crust in Sichuan, China by machine learning. *Appl. Sci.* 12(6):2955

Hammer C, Beyreuther M, Ohrnberger M. 2012. A seismic-event spotting system for volcano fast-response systems. *Bull. Seismol. Soc. Am.* 102:948–60

Hao M, Pascal A. 2022. QuakeLabeler: a fast seismic data set creation and annotation toolbox for AI applications. *Seismol. Res. Lett.* 93:997–1010

Hara S, Fukahata Y, Iio Y. 2019. P-wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets Space* 71:127

Hernandez PD, Ramírez JA, Soto MA. 2022. Deep-learning-based earthquake detection for fiber-optic distributed acoustic sensing. *J. Lightwave Technol.* 40:2639–50

Hincks T, Aspinall W, Cooke R, Gernon T. 2018. Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science* 359:1251–55

Holtzman BK, Paté A, Paisley J, Waldhauser F, Repetto D. 2018. Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Sci. Adv.* 4:eaao2929

Hsu T-Y, Huang C-W. 2021. Onsite early prediction of PGA using CNN with multi-scale and multi-domain P-waves as input. *Front. Earth Sci.* 9:626908

Hu J, Jin C, Zhang H, Hu L, Wang Z. 2022. Support vector regression for developing ground-motion models for Arias intensity, cumulative absolute velocity, and significant duration for the Kanto region, Japan. *Seismol. Res. Lett.* 93:1619–35

Jenkins WF, Gerstoft P, Bianco MJ, Bromirski PD. 2022. Unsupervised deep clustering of seismic data: monitoring the Ross Ice Shelf, Antarctica. *J. Geophys. Res. Solid Earth* 126:e2021JB021716

Jiang C, Zhang P, White MCA, Pickle R, Miller MS. 2022. A detailed earthquake catalog for Banda arc–Australian plate collision zone using machine-learning phase picker and an automated workflow. *Seismic Rec.* 2:1–10

Jozinović D, Lomax A, Štajduhar I, Michelini A. 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophys. J. Int.* 222(2):1379–89

Käufl P, Valentine AP, Trampert J. 2016. Probabilistic point source inversion of strong-motion data in 3-D media using pattern recognition: a case study for the 2008 M_w 5.4 Chino Hills earthquake. *Geophys. Res. Lett.* 43:8492–98

Kerh T, Ting SB. 2005. Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system. *Eng. Appl. Artif. Intel.* 18:857–66

Khosravikia F, Clayton P. 2021. Machine learning in ground motion prediction. *Comput. Geosci.* 148:104700

Kim D, Lekić V, Ménard B, Baron D, Taghizadeh-Popp M. 2020. Sequencing seismograms: a panoptic view of scattering in the core-mantle boundary region. *Science* 368:1223–28

Kim G, Ku B, Ko H. 2020. Multifeature fusion-based earthquake event classification using transfer learning. *IEEE Geosci. Remote Sens. Lett.* 18(6):974–78

Köhler A, Myklebust EB, Mæland S. 2022. Enhancing seismic calving event identification in Svalbard through empirical matched field processing and machine learning. *Geophys. J. Int.* 230:1305–17

Köhler A, Ohrnberger M, Scherbaum F. 2009. Unsupervised feature selection and general pattern discovery using Self-Organizing Maps for gaining insights into the nature of seismic wavefields. *Comput. Geosci.* 35:1757–67

Kong Q, Inbal A, Allen RM, Lv Q, Puder A. 2019. Machine learning aspects of the MyShake global smartphone seismic network. *Seismol. Res. Lett.* 90:546–52

Kriegerowski M, Petersen GM, Vasyura-Bathke H, Ohrnberger M. 2019. A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. *Seismol. Res. Lett.* 90:510–16

Ku B, Kim G, Ahn J-K, Lee J, Ko H. 2020. Attention-based convolutional neural network for earthquake event classification. *IEEE Geosci. Remote Sens. Lett.* 18(12):2057–61

Kuang W, Yuan C, Zhang J. 2021. Real-time determination of earthquake focal mechanism via deep learning. *Nat. Commun.* 12:1432

Kuyuk HS, Ohno S. 2018. Real-time classification of earthquake using deep learning. *Procedia Comput. Sci.* 140:298–305

Lapins S, Goitom B, Kendall J-M, Werner MJ, Cashman KV, Hammond JOS. 2021. A little data goes a long way: automating seismic phase arrival picking at Nabro volcano with transfer learning. *J. Geophys. Res. Solid Earth* 126(7):e2021JB021910

Lara F, Lara-Cueva R, Larco JC, Carrera EV, León R. 2021. A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano. *J. Volcanol. Geotherm. Res.* 490:107142

Larson J, Kramar D, Leonard K. 2020. A geostatistical analysis of seismicity in Oklahoma using regression trees and neural networks. *Phys. Geogr.* 42(4):334–50

Lee SC, Han SW. 2002. Neural-network-based models for generating artificial earthquakes and response spectra. *Comput. Struct.* 80:1627–38

Li BQ, Smith JD, Ross ZE. 2021. Basal nucleation and the prevalence of ascending swarms in Long Valley caldera. *Sci. Adv.* 7(35):eabi8368

Liao W-Y, Lee E-J, Mu D, Chen P, Rau R-J. 2021. ARRU phase picker: attention recurrent-residual U-net for picking seismic P- and S-phase arrivals. *Seismol. Res. Lett.* 92(4):2410–28

Licciardi A, Bletery Q, Rouet-Leduc B, Ampuero J-P, Juhel K. 2022. Instantaneous tracking of earthquake growth with elastogravity signals. *Nature* 606:319–24

Lin J-T, Melgar D, Thomas A, Searcy J. 2021. Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning. *J. Geophys. Res. Solid Earth* 126(10):e2021JB022703

Linville L, Pankow K, Draelos T. 2019. Deep learning models augment analyst decisions for event discrimination. *Geophys. Res. Lett.* 46:3643–51

Lomax A, Michelini A, Jozinović D. 2019. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismol. Res. Lett.* 90:517–29

Lomax A, Satriano C, Vassallo M. 2012. Automatic picker developments and optimization: FilterPicker—a robust, broadband picker for real-time seismic monitoring and earthquake early warning. *Seismol. Res. Lett.* 83:531–40

Madureira G, Ruano AE, Ruano MG. 2013. On-line operation of an intelligent seismic detector. *Soft Comput. Appl.* 85:531–42

Masotti M, Falsaperla S, Langer H, Spampinato S, Campanini R. 2006. Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy. *Geophys. Res. Lett.* 33:L20304

McBrearty IW, Delorey AA, Johnson PA. 2022. Pairwise association of seismic arrivals with convolutional neural networks. *Seismol. Res. Lett.* 90:503–9

McBrearty IW, Gomberg J, Delorey AA, Johnson PA. 2019. Earthquake arrival association with backprojection and graph theory. *Bull. Seismol. Soc. Am.* 109(6):2510–31

McKean SH, Priest JA, Dettmer J, Eaton DW. 2019. Quantifying fracture networks inferred from microseismic point clouds by a Gaussian mixture model with physical constraints. *Geophys. Res. Lett.* 46:11008–17

McLellan M, Audet P. 2020. Uncovering the physical controls of deep subduction zone slow slip using supervised classification of subducting plate features. *Geophys. J. Int.* 223:94–110

Michelini A, Cianetti S, Gaviano S, Giunchi C, Jozinovic D, Lauciani V. 2021. INSTANCE—the Italian seismic dataset for machine learning. *Earth Syst. Sci. Data Discuss.* 13(12):5509–44

Mignan A. 2014. The debate on the prognostic value of earthquake foreshocks: a meta-analysis. *Sci. Rep.* 4:4099

Mignan A, Broccardo M. 2020. Neural network applications in earthquake prediction (1994–2019): meta-analytic and statistical insights on their limitations. *Seismol. Res. Lett.* 91(4):2330–42

Moseley B, Markham A, Nissen-Meyer T. 2020a. Solving the wave equation with physics-informed deep learning. arXiv:2006.11894 [physics.comp-ph]

Moseley B, Nissen-Meyer T, Markham A. 2020b. Deep learning for fast simulation of seismic waves in complex media. *Solid Earth* 11:1527–49

Mousavi SM, Beroza GC. 2020a. Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Trans. Geosci. Remote Sens.* 58(11):8211–24

Mousavi SM, Beroza GC. 2020b. A machine-learning approach for earthquake magnitude estimation. *Geophys. Res. Lett.* 47(1):e2019GL085976

Mousavi SM, Ellsworth W, Zhu W, Beroza GC. 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* 11:3952

Mousavi SM, Horton SP, Langston CA, Samei B. 2016a. Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophys. J. Int.* 207:29–46

Mousavi SM, Langston CA, Horton SP. 2016b. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics* 81:V341–55

Mousavi SM, Sheng Y, Zhu W, Beroza GC. 2019a. STanford EArthquake Dataset (STEAD): a global data set of seismic signals for AI. *IEEE Access* 7:179464–76

Mousavi SM, Zhu W, Ellsworth W, Beroza G. 2019b. Unsupervised clustering of seismic signals using deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* 11:1693–97

Mousavi SM, Zhu W, Sheng Y, Beroza GC. 2019c. CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.* 9:10267

Münchmeyer J, Bindi D, Leser U, Tilmann F. 2021a. Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophys. J. Int.* 226(2):1086–104

Münchmeyer J, Bindi D, Leser U, Tilmann F. 2021b. The transformer earthquake alerting model: a new versatile approach to earthquake early warning. *Geophys. J. Int.* 225:646–56

Münchmeyer J, Leser U, Tilmann F. 2022a. A probabilistic view on rupture predictability: All earthquakes evolve similarly. *Geophys. Res. Lett.* 49:e2022GL098344

Münchmeyer J, Woollam J, Rietbrock A, Tilmann F, Lange D, et al. 2022b. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Adv. Geophys.* 61:151–77

Nakano M, Sugiyama D, Hori T, Kuwatani T, Tsuboi S. 2019. Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismol. Res. Lett.* 90:530–38

Otake R, Kurima J, Goto H, Sawada S. 2021. Deep learning model for spatial interpolation of real-time seismic intensity. *Seismol. Soc. Am.* 91:3433–43

Panakkat A, Adeli H. 2007. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *Int. J. Neural Syst.* 17:13–33

Paolucci R, Gatti F, Infantino M, Smerzini C, Özcebe AG, Stupazzini M. 2018. Broadband ground motions from 3D physics-based numerical simulations using artificial neural networks. *Bull. Seismol. Soc. Am.* 108:1272–86

Park Y, Mousavi SM, Zhu W, Ellsworth WL, Beroza GC. 2020. Machine-learning-based analysis of the Guy-Greenbrier, Arkansas earthquakes: a tale of two sequences. *Geophys. Res. Lett.* 47(6):e2020GL087032

Pawley S, Schultz R, Playter T, Corlett H, Shipman T, et al. 2018. The geological susceptibility of induced earthquakes in the Duvernay play. *Geophys. Res. Lett.* 45:1786–93

Peng P, He Z, Wang L, Jiang Y. 2020. Microseismic records classification using capsule network with limited training samples in underground mining. *Sci. Rep.* 10:13925

Perol T, Gharbi M, Denolle M. 2018. Convolutional neural network for earthquake detection and location. *Sci. Adv.* 4:e1700578

Picozzi M, Iaccarino AG. 2021. Forecasting the preparatory phase of induced earthquakes by recurrent neural network. *Seismol. Res. Lett.* 3:17–36

Pulli JJ, Dysart PS. 1990. An experiment in the use of trained neural networks for regional seismic event classification. *Geophys. Res. Lett.* 17:977–80

Retailleau L, Saurel J-M, Laporte M, Lavayssière A, Ferrazzini V, et al. 2022. Automatic detection for a comprehensive view of Mayotte seismicity. *C. R. Géosci.* 354(S2):153–70

Ristea N-C, Radoi A. 2021. Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves. *IEEE Geosci. Remote Sens. Lett.* 19:7502305

Rodriguez AB, Benitez C, Zuccarello L, Angelis SD, Ibanez JM. 2020. Bayesian monitoring of seismo-volcanic dynamics. *IEEE Trans. Geosci. Remote Sens.* 58:892–902

Ross ZE, Ben-Zion Y. 2014. Automatic picking of direct P, S seismic phases and fault zone head waves. *Seismol. Res. Lett.* 199:368–81

Ross ZE, Meier M-A, Hauksson E. 2018a. P wave arrival picking and first-motion polarity determination with deep learning. *J. Geophys. Res. Solid Earth* 123:5120–29

Ross ZE, Meier M-A, Hauksson E, Heaton TH. 2018b. Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am.* 108:2894–901

Ross ZE, Yue Y, Meier M-A, Hauksson E, Heaton TH. 2019. PhaseLink: a deep learning approach to seismic phase association. *J. Geophys. Res. Solid Earth* 124:856–69

Rouet-Leduc B, Hulbert C, Johnson PA. 2019. Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nat. Geosci.* 12:75–79

Rouet-Leduc B, Hulbert C, Lubbers N, Barros K, Humphreys J, Johnson PA. 2017. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* 44:9276–82

Rouet-Leduc B, Hulbert C, McBrearty IW, Johnson PA. 2020. Probing slow earthquakes with deep learning. *Geophys. Res. Lett.* 47(4):e2019GL085870

Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning representations by back-propagating errors. *Nature* 323:533–36

Saad OM, Chen Y. 2021. CapsPhase: capsule neural network for seismic phase classification and picking. *IEEE Trans. Geosci. Remote Sens.* 60:5904311

Seydoux L, Balestriero R, Poli P, De Hoop M, Campillo M, Baraniuk R. 2020. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nat. Commun.* 11:3972

Shi P, Grigoli F, Lanza F, Beroza GC, Scarabello L, Wiemer S. 2022. MALMI: an automated earthquake detection and location workflow based on machine learning and waveform migration. *Seismol. Res. Lett.* 93(5):2467–83

Shiloh L, Eyal A, Giryes R. 2019. Efficient processing of distributed acoustic sensing data using a deep learning approach. *J. Lightwave Technol.* 37:4755–62

Smith JD, Azizzadenesheli K, Ross ZE. 2020. Eikonet: solving the Eikonal equation with deep neural networks. *IEEE Trans. Geosci. Remote Sens.* 59(12):10685–96

Smith JD, Ross ZE, Azizzadenesheli K, Muir JB. 2022. HypoSVI: hypocentre inversion with Stein variational inference and physics informed neural networks. *Geophys. J. Int.* 228:698–710

Snover D, Johnson CW, Bianco MJ, Gerstoft P. 2021. Deep clustering to identify sources of urban seismic noise in Long Beach, California. *Seismol. Soc. Am.* 92:1011–22

Song C, Alkhalifah T, bin Waheed U. 2020. Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks. *Geophys. J. Int.* 225:846–59

Soto H, Schurr B. 2020. DeepPhasePick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks. *Geophys. J. Int.* 227(2):1268–94

Steinberg A, Vasyura-Bathke H, Gaebler P, Ohrnberger M, Ceranna L. 2021. Estimation of seismic moment tensors using variational inference machine learning. *J. Geophys. Res. Solid Earth* 126:e2021JB022685

Tan YJ, Waldhauser F, Ellsworth WL, Zhang M, Zhu W, et al. 2021. Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 Central Italy sequence. *Seismic Rec.* 1:11–19

Thomas AM, Inbal A, Searcy J, Shelly DR, Bürgmann R. 2021. Identification of low-frequency earthquakes on the San Andreas fault with deep learning. *Geophys. Res. Lett.* 48(13):e2021GL093157

Tian X, Zhang W, Zhang X, Zhang J, Zhang Q, et al. 2020. Comparison of single-trace and multiple-trace polarity determination for surface microseismic data using deep learning. *Seismol. Res. Lett.* 91:1794–803

Tibi R, Linville L, Young C, Brogan R. 2019. Classification of local seismic events in the Utah region: a comparison of amplitude ratio methods with a spectrogram-based machine learning approach. *Bull. Seismol. Soc. Am.* 109:2532–44

Titos M, Bueno A, Garcia L, Benitez C, Segura JC. 2020. Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geosci. Remote Sens. Lett.* 17:869–73

Trugman DT, Shearer PM. 2018. Strong correlation between stress drop and peak ground acceleration for recent M 1–4 earthquakes in the San Francisco Bay area. *Bull. Seismol. Soc. Am.* 108:929–45

Turhan Taner M, Lu L, Baysal E. 1988. Unified method for 2-D and 3-D refraction statics with first break picking by supervised learning. *SEG Tech. Program Expand. Abstr.* 1988:172–74

Uchide T. 2020. Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophys. J. Int.* 223:1658–71

Van Den Ende M, Ampuero J-P. 2020. Automated seismic source characterisation using deep graph neural networks. *Geophys. Res. Lett.* 47(17):e2020GL088690

Walter JI, Ogwari P, Thiel A, Ferrer F, Woelfel I. 2021. easyQuake: putting machine learning to work for your regional seismic network or local earthquake study. *Seismol. Res. Lett.* 92:555–63

Wang C-Y, Huang T-C, Wu Y-M. 2022. Using LSTM neural networks for onsite earthquake early warning. *Seismol. Soc. Am.* 93:814–26

Wang J, Teng T-C. 1997. Identification and picking of S phase using an artificial neural network. *Bull. Seismol. Soc. Am.* 87:1140–49

Wang J, Teng T-L. 1995. Artificial neural network-based seismic detector. *Bull. Seismol. Soc. Am.* 85:308–19

Wang K, Ellsworth WL, Beroza GC, Williams G, Zhang M, et al. 2019. Seismology with dark data: image-based processing of analog records using machine learning for the Rangely earthquake control experiment. *Seismol. Res. Lett.* 90:553–62

Wang Q, Guo Y, Yu L, Li P. 2017. Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. *IEEE Trans. Emerg. Top. Comput.* 8:148–58

Wang T, Trugman D, Lin Y. 2021. SeismoGen: seismic waveform synthesis using GAN with application to seismic data augmentation. *J. Geophys. Res. Solid Earth* 126(4):e2020JB020077

Wang Z, Zentner I, Zio E. 2020. Accounting for uncertainties of magnitude- and site-related parameters on neural network-computed ground-motion prediction equations. *Bull. Seismol. Soc. Am.* 110:629–46

Wei Z, Zhao L. 2022. P-wave velocity structure of the lower crust and uppermost mantle beneath the Sichuan–Yunnan (China) region. *Seismol. Res. Lett.* 93(4):2161–75

Wiszniowski J. 2019. Estimation of a ground motion model for induced events by Fahlman's Cascade Correlation Neural Network. *Seismol. Res. Lett.* 131:23–31

Withers KB, Moschetti MP, Thompson EM. 2020. A machine learning approach to developing ground motion models from simulated ground motions. *Geophys. Res. Lett.* 47(6):e2019GL086690

Woollam J, Münchmeyer J, Tilmann F, Rietbrock A, Lange D, et al. 2022. SeisBench—a toolbox for machine learning in seismology. *Seismol. Res. Lett.* 93:1695–709

Wu Y, Wei J, Pan J, Chen P. 2019. Research on microseismic source locations based on deep reinforcement learning. *IEEE Access* 7:39962–73

Xiao Z, Wang J, Liu C, Li J, Zhao L, Yao Z. 2021. Siamese earthquake transformer: a pair-input deep-learning model for earthquake detection and phase picking on a seismic array. *J. Geophys. Res. Solid Earth* 126(5):e2020JB021444

Yang S, Hu J, Zhang H, Liu G. 2021. Simultaneous earthquake detection on multiple stations via a convolutional neural network. *Seismol. Res. Lett.* 92(1):246–60

Yano K, Shiina T, Kurata S, Kato A, Komaki F, et al. 2021. Graph-partitioning based convolutional neural network for earthquake detection using a seismic array. *J. Geophys. Res. Solid Earth* 126(5):e2020JB020269

Yoon CE, O'Reilly O, Bergen KJ, Beroza GC. 2015. Earthquake detection through computationally efficient similarity search. *Sci. Adv.* 1(11):e1501057

Yu Z, Wang W. 2022. FastLink: a machine learning and GPU-based fast phase association method and its application to Yangbi M_S 6.4 aftershock sequences. *Geophys. J. Int.* 230:673–83

Zhang H, Innanen KA, Eaton DW. 2021. Inversion for shear-tensile focal mechanisms using an unsupervised physics-guided neural network. *Seismol. Res. Lett.* 92(4):2282–94

Zhang X, Zhang J, Yuan C, Liu S, Chen Z, Li W. 2020. Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Sci. Rep.* 10:1941

Zhu W, Beroza GC. 2019. PhaseNet: a deep-neural-network-based seismic arrival time picking method. *Geophys. J. Int.* 216:261–73

Zhu W, McBrearty IW, Mousavi SM, Ellsworth WL, Beroza GC. 2022a. Earthquake phase association using a Bayesian Gaussian mixture model. *J. Geophys. Res. Solid Earth* 127(5):e2021JB023249

Zhu W, Mousavi SM, Beroza GC. 2020. Seismic signal augmentation to improve generalization of deep neural networks. *Adv. Geophys.* 61:151–77

Zhu W, Tai KS, Mousavi SM, Bailis P, Beroza GC. 2022b. An end-to-end earthquake detection method for joint phase picking and association using deep learning. *J. Geophys. Res. Solid Earth* 127(3):e2021JB023283