A Deeper Look into 'Deep Learning of Aftershock Patterns Following Large Earthquakes': Illustrating First Principles in Neural Network Physical Interpretability

Arnaud Mignan(✉) and Marco Broccardo

Swiss Federal Institute of Technology, 8092 Zurich, Switzerland
arnaud.mignan@sed.ethz.ch

Abstract. In the last years, deep learning has solved seemingly intractable problems, boosting the hope to find (approximate) solutions to problems that now are considered unsolvable. Earthquake prediction - a recognized moonshot challenge - is obviously worthwhile exploring with deep learning. Although encouraging results have been obtained recently, deep neural networks (DNN) may sometimes create the illusion that patterns hidden in data are complex when this is not necessarily the case. We investigate the results of De Vries et al. [*Nature,* vol. 560, 2018] who defined a DNN of 6 hidden layers with 50 nodes each, and with an input layer of 12 stress features, to predict aftershock patterns in space. The performance of their DNN was assessed using ROC with AUC = 0.85 obtained. We first show that a simple artificial neural network (ANN) of 1 hidden layer yields a similar performance, suggesting that aftershock patterns are not necessarily highly abstract objects. Following first principle guidance, we then bypass the elastic stress change tensor computation, making profit of the tensorial nature of neural networks. AUC = 0.85 is again reached with an ANN, now with only two geometric and kinematic features. Not only seems deep learning to be "excessive" in the present case, the simpler ANN streamlines the process of aftershock forecasting, limits model bias, and provides better insights into aftershock physics and possible model improvement. Complexification is a controversial trend in all of Science and first principles should be applied wherever possible to gain physical interpretations of neural networks.

Keywords: Aftershock modelling · Pattern recognition · Applied deep learning

1 Introduction

Deep learning is rapidly rising as the go-to technique not only in data science [1, 2] but also for solving hard problems of Physics [3–5]. This is justified by the superior performance of deep learning in discovering complex patterns in very large datasets. One of the major advantages of deep neural networks (DNN) is that, generally, there is no need for feature extraction and engineering, as data can be used directly to train the network with great results. It comes as no surprise that DNNs are also becoming

I. Rojas et al. (Eds.): IWANN 2019, LNCS 11506, pp. 3–14, 2019.
https://doi.org/10.1007/978-3-030-20521-8_1

popular in statistical seismology [6] and give fresh hope for earthquake prediction [7], a challenge which has long been considered impossible [8]. The black-box nature of neural networks (NN) was considered an advantage in early attempts of predicting earthquakes from apparently complex patterns [9, 10]. Such a view is only valid in the context of pragmatic model applicability, not physical interpretability.

Recently, De Vries et al. [11], hereafter referred to as DeVries18, proposed a DNN to study the spatial distribution of aftershocks in the aftermath of a main seismic event. The goal of the authors was to design a stress-change based classifier for determining the spatial likelihood of aftershocks. Once the DNN was trained and tested (see Sect. 2.2 for details on the dataset and model), they provided a physical interpretation of the aftershock pattern. In particular, they analyzed three stress metrics and showed that each alone could lead to similar results as their DNN classifier. In this regard, the DeVries18 study is an improvement over the back-box approach [9, 10].

Designing a suitable DNN is a highly iterative process, based on hyper-parameterization tuning and, sometimes (even if not highly recommendable in this context), feature engineering. How do such choices affect, not only model performance, but physical interpretability? In view of the flexibility of deep learning, can we miss first-principles in the modelling process? We aim at answering these questions in the context of aftershock spatial pattern prediction, taking the DeVries18 study as baseline model.

2 Artificial Neural Networks in Statistical Seismology

2.1 Literature Survey

The Earthquake Prediction Challenge. '*The subject of Statistical Seismology aims to bridge the gap between physics-based models without statistics, and statistics-based models without physics*' [12] - This scientific domain can be divided roughly into two categories, with earthquakes as point sources (i.e. seismicity) in stochastic non-stationary processes or earthquakes as seismic waves radiating from finite sources.

Earthquake prediction remains the Grail of Seismology, as well as a moonshot challenge for all of society that is under the threat of large earthquakes. Earthquake predictability research has already gone through several cycles of enchantments and disillusions [13] with earthquake physics remaining derivative: It is still in a cataloguing phase of seismicity patterns, akin the naturalists collecting animals and plants in past centuries before modern biology emerged. It remains to be identified whether earthquake patterns are complex, in the holistic sense of the Critical Point Theory [14] which relates to Thermodynamics and Chaos, or complicated, in the reductionist sense of the recent Solid Seismicity Theory [15–17] which relates to Geometry. In this context, machine learning should help improving earthquake pattern recognition. While those algorithms are theoretically agnostic, they may give some insights into the theoretical directions to follow. For a review of pattern recognition algorithms used specifically for earthquake forecasting, see [18].

It should be mentioned that aftershock prediction (or forecasting) is different and much easier than mainshock prediction since aftershocks follow well known statistics in time (see review by [19]), space [17, 20, 21] and productivity [17, 22]. Operational aftershock forecasting is thus already possible [23].

Neural Network Models for Earthquake Pattern Prediction. The earliest attempts to apply ANNs to earthquake predictions dates back, to the best of our knowledge, to 1994 [24, 25]. Few studies followed in the next decade [9, 26, 27]. The next milestone was in 2007 with the comprehensive work of Panakkat and Adeli [28], one of the first to use deep learning (2 hidden layers) as well as radial basis function (RBF) networks, and the first to test recurrent neural networks (RNN). Neural network research for earthquake prediction took off from there, but mostly from a computer science perspective and from a limited number of teams [7, 29, 30].

All the aforementioned works used structured seismic data as input, i.e. earthquake catalogues in tabular format with location, occurrence time, and magnitude of events (except [24] who used seismic electric signals). For two different approaches to feature engineering on structured data, compare [27] for a financial market approach to e.g. [30] for a 'seismicity law' approach. The most common outputs are the predicted magnitude, occurrence time and location of large earthquakes. Overall, mixed results were obtained so far, with performance decreasing with increasing mainshock magnitude [7, 10]. The gain of using neural networks instead of simpler methods remains unclear since performance was rarely compared to a baseline model.

The most recent DNN model was proposed by De Vries et al. [11], which was different from previous studies in various ways. First, it did not try to predict mainshock characteristics but the spatial patterns of aftershocks (an early attempt at predicting aftershock spatial distribution had already been done by [31] but for one sequence only). DeVries18 used a global earthquake catalogue for aftershock binary classification (aftershocks present or not in geographic cells) but with features engineered from stress computed from mainshock rupture models. Their DNN model will be fully described in Sect. 2.2, and critically assessed and improved upon in Sect. 3 in terms of physical interpretability.

Let us mention that raw earthquake data is unstructured, in the form of seismic wave timeseries. To the best of our knowledge, neural networks have not yet been applied to seismic wave data to predict earthquakes. However Random Forests have recently been used to predict lab earthquakes [32]. We should also mention earthquake early warning (EEW), which consists in predicting the arrival time of S waves and surface waves based on P wave data (which is however not to be confused with earthquake prediction). This work was pioneered with NNs in 1996 [33] but NN applications in EEW still remain at an early stage. The best-known example is the MyShake project where a neural network classifies the amplitude and frequency content of smartphone movements to discriminate early earthquake shaking from human activities [34].

With the recent advances in applying CNNs [35, 36] and RNNs [37] to automatically pick seismic waves, one could easily imagine applying those techniques to predict a mainshock based on foreshock seismic waves [38], as already tested with Random Forests [32]. Moreover, those techniques tend to improve the quality of earthquake catalogues by increasing the number of events ten-folds [35]. which can in turn be used as features in earthquake prediction based on structured data. A recent meta-analysis indeed showed that an increase in the amount of micro-seismicity improves anomaly detection [39]. Hence, CNN-based earthquake catalogues would likely help improving model performances. A semi-supervised mixture model has also recently been proposed towards that goal [40]. Finally, autoencoder networks [41] and generative adversarial networks (GAN) [42] of seismic wave data could help reducing data dimensions, which could further facilitate feature engineering.

2.2 The DeVries18 Study

The DeVries18 study [11] defined a DNN for binary classification of spatial patterns of aftershocks. It was made of 6 hidden layers, each composed of 50 nodes with hyperbolic tangent activation and a 0.5 dropout rate, for a total of 13,451 weights and biases. Their input layer contained 12 nodes, representing features engineered from the elastic stress change tensor of the related mainshocks. Their workflow is illustrated in Fig. 1.

DeVries18 defined aftershocks as all events - as catalogued in the International Seismological Center (ISC) global catalogue [43] - which occurred between one second and one year after a mainshock, and within 100 km horizontally of the mainshock rupture (50 km vertically). They repeated the operation for 199 mainshocks which occurred between 1968 and 2012. Then they gridded aftershocks in 5 km-wide XYZ cells, each cell labelled 1 if it contained *at least* one aftershock, 0 otherwise. They retrieved mainshock rupture data (geometry, mechanism, slip) from the Finite-Source Rupture Model Database (SRCMOD) [44]. They computed the elastic stress tensor $\Delta\boldsymbol{\sigma}$ [45] at the centroid of each cell and finally defined their DNN input layer based on the 12 following features (normalized by 10^{-6}): the absolute values of the six independent components of $\Delta\boldsymbol{\sigma}$, which are $|\Delta\sigma_{xx}|$, $|\Delta\sigma_{xy}|$, $|\Delta\sigma_{xz}|$, $|\Delta\sigma_{yy}|$, $|\Delta\sigma_{yz}|$, $|\Delta\sigma_{zz}|$, and their opposites $-|\Delta\sigma_{xx}|$, $-|\Delta\sigma_{xy}|$, $-|\Delta\sigma_{xz}|$, $-|\Delta\sigma_{yy}|$, $-|\Delta\sigma_{yz}|$, $-|\Delta\sigma_{zz}|$.

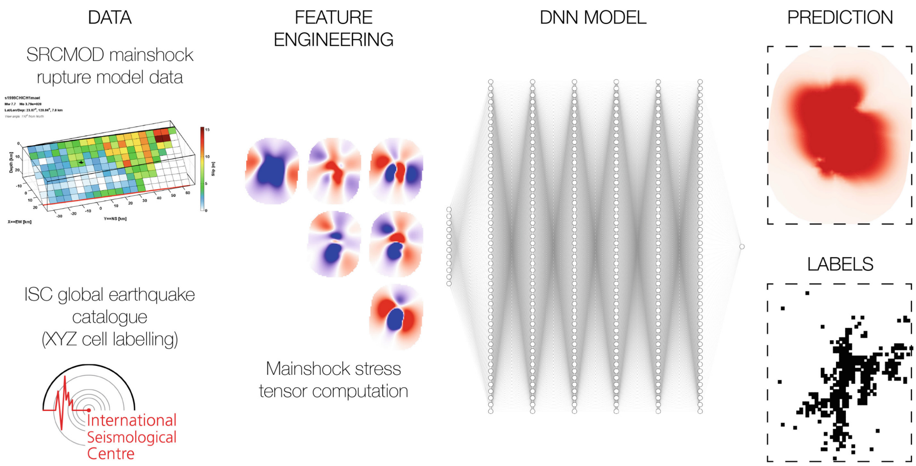

Fig. 1. The DeVries18 workflow to predict aftershock patterns with deep learning (illustrated with SRCMOD eventTag 1999CHICHI01MAxx).

Their model was trained on c. 75% of the data, with each XYZ cell considered one sample. They tested the performance of their DNN on the remaining 25% of the data, by accessing model accuracy via the Receiver Operating Characteristic (ROC) curve, calculating the Area Under the Curve (AUC). Based on their model input and topology, DeVries18 obtained AUC = 0.849, which appears impressive compared to AUC = 0.583 (near-random performance) obtained for the classical Coulomb failure criterion [46], which represents the main earthquake-triggering model of the current paradigm. It remains unclear why such low performance was obtained for Coulomb stress. It is possible that cherry-picking or overfitting had been previously achieved or it is maybe DeVries18, investigating Coulomb stress on a global dataset, that made assumptions which are too generic (e.g. by not investigating possible changes in the regional stress field [17]). This debate is worth mentioning but it is however outside the scope of the present paper.

Previous NN studies on earthquake prediction (Sect. 2.1) did not relate their findings to any physical process (black-box approach) although they assumed, explicitly or implicitly, that highly non-linear patterns were due to complex processes which can only be investigated holistically (i.e. mainshock as a critical point). To the best of our knowledge, only DeVries18 used their DNN results to seek for interpretable and meaningful physical patterns.

3 Applying First Principles to Neural Network Interpretability

3.1 Was High Abstraction Required to Predict Aftershock Patterns?

Deep learning has dramatically improved the state-of-the-art in computer vision, among other fields [1], and it is often accepted that defining a larger and deeper DNN does not hurt model performance. However, a deeper NN can be interpreted as a model of higher abstraction. In computer vision, for instance, a first layer may represent simple shapes, a second layer parts of a face (such as eye, nose, ear), and a third layer, different faces. When aftershock patterns are predicted by a 6-hidden-layer DNN [11], it captivates the collective imagination as to the degree of abstraction that seismicity patterns carry. This can explain a certain media euphoria about artificial intelligence predicting earthquakes [47–50]. This is unfortunately misleading. As proven below, a shallow neural network can predict aftershock patterns with a similar performance as the DNN of DeVries18.

We should first emphasize that DeVries18 did not flatten aftershock maps as in computer vision made with fully-connected feedforward networks, but used each XYZ cell as sample. This turns the study case from unstructured to structured data, with features defined in tabular form (X, Y, Z, stress features). Standard machine learning algorithms are performant on such data. This especially means that we only deal with a handful of features instead of the hundreds or thousands that usually call for deep learning.

We simplified the topology of the DeVries18 model and found, for example, that both a DNN with topology 12-8-8-1 (with dropout rate reduced to 0.2) or just an ANN with topology 12-30-1 yield a similar AUC of 0.85 (Fig. 2). The newly predicted aftershock patterns remained similar to the ones of DeVries18, with spatial lobes related to the spatial distribution of the stress features (similar as in Fig. 1). This demonstrates that deep learning is so far unnecessary for aftershock pattern prediction. We will show how simple aftershock patterns are in the next subsection.

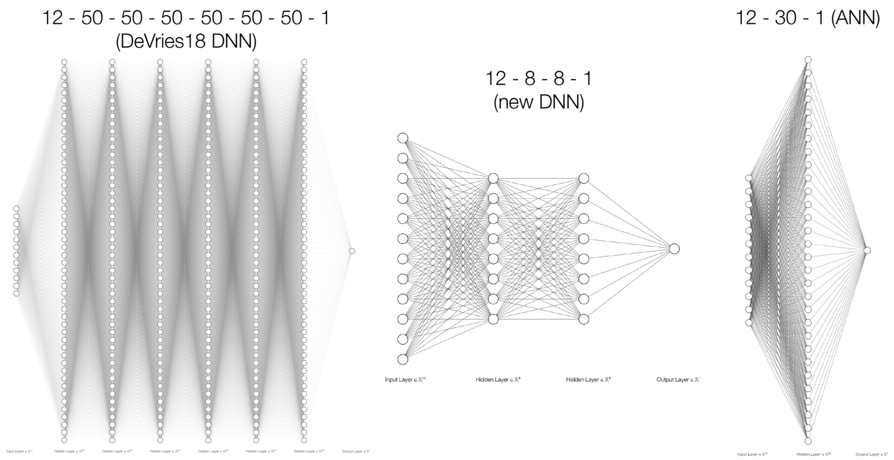

Fig. 2. Different topologies with similar AUC performances and similar predicted aftershock patterns. Network topology plots generated with alexlenail.me/NN-SVG/.

3.2 Were Stress Metrics the Most Pertinent Physical Parameters?

Taking Advantage of the Tensorial Definition of Neural Networks. NNs consist of chains of tensor operations, and following also the Universal Approximation Theorem, we can intuitively deduct that any stress tensor could be mimicked by a certain NN topology and weight combination.

Assuming the linearized theory of elasticity, the stress-change tensor can be generally written as

$$\Delta\boldsymbol{\sigma} = \boldsymbol{C}\boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{C}$ is the 4th order elasticity tensor (which, in the case of isotropic elasticity, has two independent constants, i.e. the Lamé parameters) and where $\boldsymbol{\varepsilon}$ is the linear strain tensor defined as the symmetric part of the displacement gradient

$$\boldsymbol{\varepsilon}(\boldsymbol{r},\boldsymbol{d}) = \frac{1}{2}\left(\nabla\boldsymbol{u}(\boldsymbol{r},\boldsymbol{d}) + \nabla\boldsymbol{u}^{\boldsymbol{T}}(\boldsymbol{r},\boldsymbol{d})\right) \tag{2}$$

where $\boldsymbol{u}(\boldsymbol{r}, \boldsymbol{d})$ is the displacement field at a distance $\boldsymbol{r}$ from the rupture, and $\boldsymbol{d}$ is the finite rupture displacement. Note the possible parallel between Eq. (1) and a neuron linear function $z = \boldsymbol{w}^{\mathbf{T}}x + \mathrm{b}$ where the activation $g(z)$ (tanh, relu, or else) could also relate to triggering of an earthquake above a threshold $\Delta\sigma$.

Following first principles, one shall define the NN input layer from displacement data directly, avoiding any model assumption. Having a neural network doing (eventually) a mapping from deformation (mainshock geometry and kinematics) back to deformation (simplified to aftershocks occurring or not) avoids making any assumption on stress (elasticity versus poroelasticity theory, plasticity, *etc.*), material properties (Lamé parameters), and other unknowns. In particular, this avoids having large uncertainties potentially affecting the quality of the classifier, or in other words, this avoids theoretical model bias. Recall that deformation is measurable and should be used as input layer while stress is derivative, representing subjective feature engineering.

New Neural Network with Displacement-Based Input Layer. The two variables in Eq. (2) are the distance to the rupture r and the rupture displacement d. We here define r as the minimum distance between cell XYZ and the mainshock rupture plane, and d as the mean slip on the rupture obtained from the SRCMOD database [44] (observe that both r and d are scalars in this version and orthogonal features). r represents a geometric feature and d a kinematic feature. Such a simple parameterization of aftershock patterns is compatible with the observation that aftershocks occur closest to the mainshock rupture with their likelihood decreasing as a power-law with increasing distance [17, 20, 21]. Considering the mean slip provides a physical link to deformation and a way to scale aftershock patterns to the size of the mainshock [17].

Those two features are found sufficient to get a similar model performance as the DeVries18 model, with both a simple DNN 2-6-6-1 or with an ANN 2-30-1 yielding AUC = 0.84−0.86 (Fig. 3; note that we here changed tanh activation to rectified linear unit; a dropout rate of 0.2 was again used for the DNN). The predicted aftershock patterns are simpler and blurrier than the ones produced by DeVries18 but generalize well (Figs. 3 and 4).

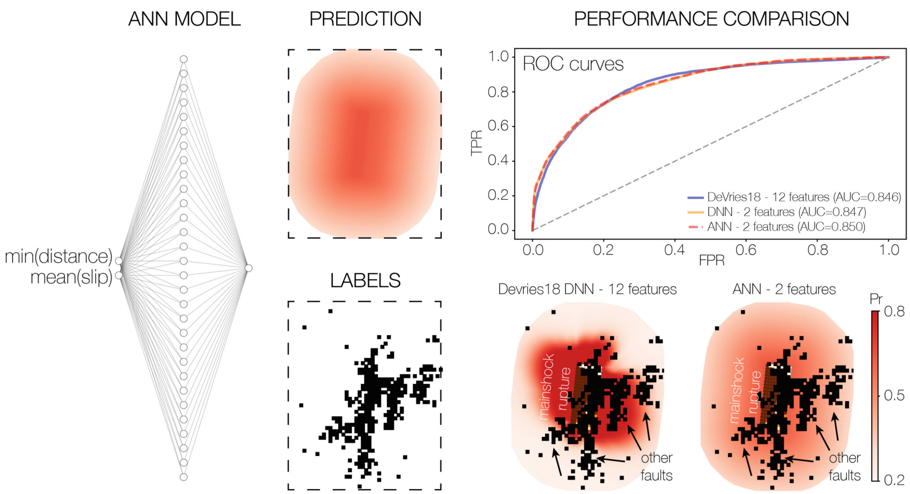

Fig. 3. 2-features ANN model and comparison with the DeVries18 deep learning performance. Aftershock maps for SRCMOD eventTag 1999CHICHI01MAxx; ROC curves for the full test set.

Fig. 4. Other examples from the test set of aftershock pattern predictions made by the 12-features DeVries18 DNN and the new 2-features ANN model (SRCMOD eventTags s1995KOBEJA01YOSH, s2005KASHMI01SHAO, and s2010DARFIE01ATZO).

Interpreting the DeVries18 Stress-Based Input Layer. In order to interpret their DNN, DeVries18 tested various stress metrics [51] and concluded that the sum A of absolute values of independent components of $\Delta\boldsymbol{\sigma}$, the von Mises yield criterion $\sqrt{3\Delta J_2}$, and the maximum change in shear stress $\Delta\tau$, respectively

$$\begin{cases} A = |\Delta\sigma_{xx}| + |\Delta\sigma_{yy}| + |\Delta\sigma_{zz}| + |\Delta\sigma_{xy}| + |\Delta\sigma_{xz}| + |\Delta\sigma_{yz}| \\ \sqrt{3\Delta J_2} = \sqrt{\Delta I_1^2(\Delta\boldsymbol{\sigma}') - 3\Delta I_2(\Delta\boldsymbol{\sigma}')} \\ \Delta\tau = |\Delta\boldsymbol{\sigma}_1 - \Delta\boldsymbol{\sigma}_3|/2 \end{cases} \tag{3}$$

(where $\Delta\boldsymbol{\sigma}' = \Delta\boldsymbol{\sigma} - (\Delta\boldsymbol{\sigma} : I)/3 \cdot I$ is the deviatoric stress change tensor with I the identity matrix; ΔI_1 and ΔI_2 are the 1st and 2nd invariants) yield similar AUC scores as their DNN prediction (i.e. AUC = 0.85). In fact, since they had already obtained a similar result in 2017 [51], their best stress metrics should have later been used as baseline model. The DeVries18 conclusion should then have been that deep learning - in fact - did not help improving scores obtained from simple stress indices. Instead, their study was misinterpreted by Nature News as *'Artificial intelligence nails predictions of earthquake aftershocks'* [47].

How can we explain the observation that both displacement (r, d) and stress metrics lead to similar performances? We see that in both the DeVries18 feature engineering and the three metrics of Eq. (3), only absolute values of the stress components are considered, which means that all dipolar information is lost. What remains at first order is the distance r from mainshock rupture (subject to some rotations at second order) and a spatial scaling, which can be calibrated by mainshock rupture displacement d (see e.g. [17] for an analytical expression of static stress changes along r for different mainshock magnitudes). The top row of Fig. 4 indeed shows that aftershock patterns do not seem to be correlated to the stress lobe geometry but simply with distance from the mainshock rupture (bottom row).

Interestingly, the AUC seems to plateau at c. 0.85 for the full test set for any mainshock-based feature and all the tested hyper-parameterizations. This suggests that valuable information is currently missing from the models. Comparing stress-based and displacement-based pattern maps shows that some aftershock clusters occur on well-defined lineaments which are likely representative of other fault segments (Fig. 3). Deviations from seismicity spatial laws have already been related to the presence of faults at proximity of the mainshock rupture [16, 17, 21]. We suggest that model performance could further improve, once features based on fault network data are added.

4 Conclusions

The present work is not a purely intellectual exercise on Occam's Razor, nor an incremental work on neural network topology simplification and feature engineering. It has important implications in operational aftershock forecasting, physical research, and communication of AI results in general:

(1) By entirely bypassing stress computation, aftershock forecasting could be streamlined with the proposed geometric and kinematic features possibly retrievable in real-time. It also avoids making physical assumptions that may lead to some model bias (compare Fig. 3 to Fig. 1). For this, we took advantage of the tensorial nature of a neural network. We also used an ANN instead of a DNN in view of the very small number of features.

(2) By using first principles, we were able to show that aftershock patterns are simple after all, in agreement with the literature on the statistical properties of the spatial distribution of aftershock [17, 20, 21]. While this does not tell us what the physics is behind the mapping of mainshock deformation to aftershock deformation, it already clarifies the importance of distance to rupture and rupture slip as main physical measurable parameters.

(3) Deep learning branding helps to captivate the popular imagination. This, in itself, is fine for model applications and to revive potentially moribund topics. However, it has dangerous consequences in physical interpretability of neural networks, as demonstrated above. Whenever possible, first principles should be applied even if this means using shallower rather than deeper learning. This can only be decided on a case-to-case basis.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
2. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)
3. Carleo, G., Troyer, M.: Solving the quantum many-body problem with artificial neural networks. Science **355**, 602–606 (2017)
4. Han, J., Jentzen, A., Weinan, E.: Solving high-dimensional partial differential equations using deep learning. PNAS **115**(34), 8505–8510 (2018)
5. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. Phys. Rev. Lett. **120**, 024102 (2018)
6. Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J., Gerstoft, P.: Machine learning in seismology: turning data into insights. Seismol. Res. Lett. **90**(1), 3–14 (2019)
7. Panakkat, A., Adeli, H.: Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. Comput.-Aided Civ. Infrastruct. Eng. **24**, 280–292 (2009)
8. Geller, R.J., Jackson, D.D., Kagan, Y.Y., Mulargia, F.: Earthquakes cannot be predicted. Science **275**(5306), 1616–1617 (1997)
9. Brodi, B.: A neural-network model for earthquake occurrence. J. Geodyn. **32**, 289–310 (2001)
10. Moustra, M., Avraamides, M., Christodoulou, C.: Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. Expert Syst. Appl. **38**, 15032–15039 (2011)
11. DeVries, P.M.R., Viégas, F., Wattenberg, M., Meade, B.J.: Deep learning of aftershock patterns following large earthquakes. Nature **560**, 632–634 (2018)
12. Vere-Jones, D., Ben-Zion, Y., Zuniga, R.: Statistical seismology. Pure Appl. Geophys. **162**, 1023–1026 (2005)
13. Mignan, A.: Retrospective on the Accelerating Seismic Release (ASR) hypothesis: controversy and new horizons. Tectonophysics **505**, 1–16 (2011)
14. Sornette, D.: Critical Phenomena in Natural Sciences, Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer, New York (2009). https://doi.org/10.1007/3-540-33182-4
15. Mignan, A.: Seismicity precursors to large earthquakes unified in a stress accumulation framework. Geophys. Res. Lett. **39**, L21308 (2012)
16. Mignan, A.: Static behaviour of induced seismicity. Nonlin. Process. Geophys. **23**, 107–113 (2016)
17. Mignan, A.: Utsu aftershock productivity law explained from geometric operations on the permanent static stress field of mainshocks. Nonlin. Process. Geophys. **25**, 241–250 (2018)
18. Tiampo, K.F., Shcherbakov, R.: Seismicity-based earthquake forecasting techniques: ten years of progress. Tectonophysics **522–523**, 89–121 (2012)
19. Mignan, A.: Modeling aftershocks as a stretched exponential relaxation. Geophys. Res. Lett. **42**, 9726–9732 (2015)
20. Richards-Dinger, K., Stein, R.S., Toda, S.: Decay of aftershock density with distance does not indicate triggering by dynamic stress. Nature **467**, 583–586 (2010)
21. Hainzl, S., Brietzke, G.B., Zöller, G.: Quantitative earthquake forecasts resulting from static stress triggering. J. Geophys. Res. **115**, B11311 (2010)
22. Båth, M.: Lateral inhomogeneities of the upper mantle. Tectonophysics **2**(6), 483–514 (1965)

23. Gerstenberger, M.C., Wiemer, S., Jones, L.M., Reasenberg, P.A.: Real-time forecasts of tomorrow's earthquakes in California. Nature **435**, 328–331 (2005)
24. Lakkos, S., Hadjiprocopis, A., Compley, R., Smith, P.: A neural network scheme for earthquake prediction based on the seismic electric signals. In: Proceedings of the IEEE Conference on Neural Networks and Signal Processing, pp. 681–689. IEEE, Ermioni (1994)
25. Alves, E.I.: Notice on the predictability of earthquake occurrences. Memórias e Notícias **117**, 51–61 (1994)
26. Liu, Y., Wang, Y., Li, Y., Zhang, B., Wu, G.: Earthquake prediction by RBF neural network ensemble. In: Yin, F.-L., Wang, J., Guo, C. (eds.) ISNN 2004. LNCS, vol. 3174, pp. 962–969. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28648-6_153
27. Alves, E.I.: Earthquake forecasting using neural networks: results and future work. Nonlin. Dyn. **44**, 341–349 (2006)
28. Panakkat, A., Adeli, H.: Neural network models for earthquake magnitude prediction using multiple seismicity indicators. Int. J. Neural Syst. **17**(1), 13–33 (2007)
29. Martínez-Álvarez, F., Reyes, J., Morales-Esteban, A., Rubio-Escudero, C.: Determining the best set of seismicity indicators to predict earthquakes. Two case studies Chile and the Iberian Peninsula. Knowl.-Based Syst. **50**, 198–210 (2013)
30. Asencio-Cortés, G., Martínez-Álvarez, F., Morales-Esteban, A., Reyes, J.: A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. Knowl.-Based Syst. **101**, 15–30 (2016)
31. Madahizadeh, R., Allamehzadeh, M.: Prediction of aftershocks distribution using artificial neural networks and its application on the May 12, 2008 Sichuan earthquake. JSEE **11**(3), 111–120 (2009)
32. Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J., Johnson, P.A.: Machine learning predicts laboratory earthquakes. Geophys. Res. Lett. **44**, 9276–9282 (2017)
33. Leach, R., Dowla, F.: Earthquake early warning system using real-time signal processing. In: Proceedings of the 1996 IEEE Signal Processing Society Workshop, pp. 463–472. IEEE, Kyoto (1996)
34. Kong, Q., Allen, R.M., Schreier, L., Kwon, Y.-W.: MyShake: a smartphone seismic network for earthquake early warning and beyond. Sci. Adv. **2**, e1501055 (2016)
35. Perol, T., Gharbi, M., Denolle, M.: Convolutional neural network for earthquake detection and location. Sci. Adv. **4**, e1700578 (2018)
36. Ross, Z.E., Meier, M.-A., Hauksson, E.: P wave arrival picking and first-motion polarity determination with deep learning. J. Geophys. Res. Solid Earth **123**, 5120–5129 (2018)
37. Ross, Z.E., Yue, Y., Meier, M.-A., Hauksson, E.: Phaselink: a deep learning approach to seismic phase association. J. Geophys. Res. Solid Earth (2019). https://doi.org/10.1029/2018jb016674
38. Bouchon, M., Karabulut, H., Aktar, M., Özalaybey, S., Schmittbuhl, J., Bouin, M.P.: Extended nucleation of the 1999 Mw 7.6 Izmit earthquake. Science **331**(6019), 877–880 (2011)
39. Mignan, A.: The debate on the prognostic value of earthquake foreshocks: a meta-analysis. Sci. Rep. **4**, 4099 (2014)
40. Mignan, A.: Asymmetric Laplace mixture modelling of incomplete power-law distributions: application to 'seismicity vision'. In: Arai, K., Kapoor, S. (eds.) CVC 2019. AISC, vol. 944, pp. 30–43. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17798-0_4
41. Valentine, A.P., Trampert, J.: Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data. Geophys. J. Int. **189**, 1183–1202 (2012)

42. Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., Andrews, J.: Machine learning seismic wave discrimination: application to earthquake early warning. Geophys. Res. Lett. **45**, 4773–4779 (2018)
43. International Seismological Center. http://www.isc.ac.uk/. Accessed 29 Jan 2019
44. Finite-Source Rupture Model Database. http://equake-rc.info/SRCMOD/. Accessed 29 Jan 2019
45. Okada, Y.: Surface deformation due to shear and tensile faults in a half-space. Bull. Seismol. Soc. Am. **75**(4), 1135–1154 (1985)
46. King, G.C.P.: Fault interaction, earthquake stress changes, and the evolution of seismicity. Treatise Geophys. **4**, 225–255 (2007)
47. Nature News: Artificial intelligence nails predictions of earthquake aftershocks. https://www.nature.com/articles/d41586-018-06091-z. Accessed 29 Jan 2019
48. The New York Times: A.I. is Helping Scientists Predict When and Where the Next Big Earthquake Will Be. https://www.nytimes.com/2018/10/26/technology/earthquake-predictions-artificial-intelligence.html. Accessed 29 Jan 2019
49. Futurism: Google's AI can help predict where earthquake aftershocks are most likely. https://futurism.com/the-byte/aftershocks-earthquake-prediction. Accessed 29 Jan 2019
50. The Verge: Google and Harvard team up to use deep learning to predict earthquake aftershocks. https://www.theverge.com/2018/8/30/17799356/ai-predict-earthquake-aftershocks-google-harvard. Accessed 29 Jan 2019
51. Meade, B.J., DeVries, P.M.R., Faller, J., Viegas, F., Wattenberg, M.: What is better than coulomb failure stress? A ranking of scalar static stress triggering mechanisms from 10^5 mainshock-aftershock pairs. Geophys. Res. Lett. **44**, 11409–11416 (2017)