Dear *Nature* editors,

We are happy that people have downloaded our data and been able to run the code.

That said, these specific comments are not worthy of publication in *Nature*. They do not move the field forward; in fact, they are either incorrect or miss the point of the science entirely while simultaneously managing to sound extraordinarily condescending.

Overall, the comments can be categorized into three parts: (1) the idea of 'data leakage' potentially inflating results, (2) the fact that a random forest approach performs similarly to a neural network, and (3) the fact that we are learning a simple signal. We address each of these in turn.

(1) The concern about 'data leakage' inflating results makes no sense in the scientific context. As explained in our paper, we partition the training/testing data sets randomly, based on distinct mainshock, and aftershocks are selected based on a simple, fixed time-window approach. The commenters correctly point out that this approach means that in a few training/testing examples, Mainshock B is included in the aftershock sequence of Mainshock A. Taken totally out of context, this seems like it could inflate results. But this is not at all true when you consider the specific scientific approach.

    For example, say that Mainshock A is assigned to the training data set, and Mainshock B is assigned to the testing data set, but Mainshock B is included as one of Mainshock A's aftershocks. The network would be partially trained on the aftershock sequence of Mainshock A (using the stress changes caused by Mainshock A as inputs). Because Mainshock B is included in the aftershock sequence of Mainshock A, the network could then be tested on some of the same aftershocks, but using the stress changes caused by Mainshock B as inputs instead. The network is mapping modeled stress changes to aftershocks, and this mapping will be entirely different for the example in the training data set and the example in the testing data set, although they overlap geographically. There's no information in the training data set that would help the network perform well on the testing data set – instead, the network is being asked in the testing data set to explain the same aftershocks that it has seen in the training data set, but with a different mainshock. If anything, this would hurt performance on the testing data set.

    As a result of this 'data leakage,' the authors of these comments claim that we have inflated the performance of the neural network. As noted above, we randomly partitioned the data into a training and testing data set and set aside the testing data set early on. This is a standard approach. In the final evaluation, maximum shear stress change, the von-Mises yield criterion, and the neural network all performed similarly well (AUC scores of 0.85) on our testing data set (see Figure 2 of the paper).

    The comparable performance of the neural network, maximum shear stress change, and the von-Mises yield criterion was one of the central results of the paper. The neural network learned a forecast that is highly correlated with, and therefore performs similarly to, these physical quantities. This was profoundly interesting because it suggested that these kinds of approaches may be useful for identifying specific physical quantities or phenomena that could potentially play a role in earthquake triggering. Maximum shear stress change and the von-Mises yield criterion have, to date, not been widely used in the earthquake triggering literature.

The authors of these comments write that when they repartitioned the training/testing data sets, they obtain AUC scores of 0.77 for the neural network. They also obtain AUC scores of 0.77 for maximum shear stress change and the von-Mises yield criterion. The absolute AUC values certainly differ in absolute terms because our data set was randomly selected, but the neural network is no more of an improvement evaluated on our testing data set than it is on theirs. Puzzlingly though, the authors of these comments state that there are concerned because, based on their partitioning of the data, "In terms of predictive performance, the machine learning methods are not an improvement over traditional techniques of the maximum change in shear stress or the von-Mises yield criterion."

This is the entire point of the paper. The neural network seems to identify maximum shear stress change and the von-Mises yield criterion as useful quantities in distinguishing grid cells. (These quantities are not "traditional techniques" as these commenters state; they have been little-used to date.)

(2) The authors of these comments say that the paper "gives the misleading impression that only deep learning is capable of learning the aftershocks." In the paper, we used a neural network as a tool to gain insight into aftershock patterns; we do not suggest that another machine learning method would not be useful.

Neural networks and random forests often perform similarly well for shallow or non-perceptual machine learning tasks. This is no surprise. The insightful result in the paper was that a neural network learned a location forecast that is highly correlated with simple physics-based stress quantities. The fact that another machine learning method may also be able to provide these insights is missing the point. It's like saying "We wrote the same paragraph with a pencil instead of a pen." Science has not been advanced.

(3) The fact that the neural network learns a simple pattern is the entire point of the paper. The neural network learned a pattern that is highly correlated with remarkably simple, but little-used quantities – maximum shear stress and the second invariant of the deviatoric stress tensor. As noted above, we make a big deal about this in the paper, because it is the whole point.

These comments were made without any scientific context. We are earthquake scientists and our goal was to use a machine learning approach to gain some insight into aftershock location patterns. We accomplished this goal. The authors of these comments do not – we will be disappointed if *Nature* publishes them.

Phoebe DeVries and Brendan Meade