AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters®

RESEARCH LETTER
10.1029/2021GL097101

**Key Points:**

- This is the first attempt to apply a convolutional neural network (CNN) on shear wave splitting analysis
- The CNN is trained and verified using 86,903 human-labeled measurements and tested using synthetic data
- Application of the trained CNN on seismic data recorded in Alaska shows a high degree of consistency with human-labeled results

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

S. S. Gao,
sgao@mst.edu

**Citation:**

Zhang, Y., & Gao, S. S. (2022). Classification of teleseismic shear wave splitting measurements: A convolutional neural network approach. *Geophysical Research Letters*, *49*, e2021GL097101. https://doi.org/10.1029/2021GL097101

Received 17 NOV 2021
Accepted 3 JUN 2022

**Author Contributions:**

**Conceptualization:** Stephen S. Gao
**Data curation:** Yanwei Zhang
**Formal analysis:** Yanwei Zhang
**Funding acquisition:** Stephen S. Gao
**Investigation:** Yanwei Zhang, Stephen S. Gao
**Methodology:** Stephen S. Gao
**Project Administration:** Stephen S. Gao
**Resources:** Stephen S. Gao
**Software:** Yanwei Zhang
**Supervision:** Stephen S. Gao
**Validation:** Yanwei Zhang, Stephen S. Gao
**Visualization:** Yanwei Zhang
**Writing – original draft:** Yanwei Zhang
**Writing – review & editing:** Stephen S. Gao



# Classification of Teleseismic Shear Wave Splitting Measurements: A Convolutional Neural Network Approach

**Yanwei Zhang[1] and Stephen S. Gao[1]**

[1]Geology and Geophysics Program, Missouri University of Science and Technology, Rolla, MO, USA

**Abstract** Shear wave splitting (SWS) analysis is widely used to provide critical constraints on crustal and mantle structure and dynamic models. In order to obtain reliable splitting measurements, an essential step is to visually verify all the measurements to reject problematic measurements, a task that is increasingly time consuming due to the exponential increase in the amount of data. In this study, we utilized a convolutional neural network (CNN) based method to automatically select reliable SWS measurements. The CNN was trained by human-verified teleseismic SWS measurements and tested using synthetic SWS measurements. Application of the trained CNN to broadband seismic data recorded in south central Alaska reveals that CNN classifies 97.0% of human selected measurements as acceptable, and revealed ~30% additional measurements. To our knowledge, this is the first study to systematically explore the potential of a machine-learning based technique to assist with SWS analysis.

**Plain Language Summary** One of the routinely utilized seismological techniques to delineate the internal structure and dynamics of the Earth's crust and mantle is shear wave splitting (SWS) analysis. In order to obtain reliable measurements, it is necessary to perform manual verification of the numerous measurements, a task that is increasingly time consuming due to the rapid increase in the amount of seismic data recorded worldwide. In this study, by taking advantage of the recent revolutionary development of a machine-learning technique called the convolutional neural network (CNN), we systematically investigate the potential and performance of a CNN that is trained using human-labeled SWS measurements and tested using both synthetic and recorded data. The results demonstrate satisfactory performance of the CNN on both types of data. Although additional development of the CNN is needed to reach over-human performance, our tests suggest that if a human operator uses CNN-classified results for manual verification, an approximately 60% reduction of the effort will be achieved, and only about 3% of the measurements will be missed.

## 1. Introduction

It has long been recognized that the *P*-to-*s* converted phases from the core-mantle boundary such as SKS, PKS, and SKKS (hereafter referred to as XKS) split into orthogonally polarized fast and slow components in azimuthally anisotropic media (Ando et al., 1983; Long & Silver, 2009; Savage, 1999; Silver & Chan, 1991). The two splitting parameters, the polarization orientation of the fast component (fast orientation or $\phi$) and the time separation (splitting time or $\delta t$) between the two waves, reveal the orientation and splitting magnitude of the anisotropy, respectively. Over the past several decades, the shear wave splitting (SWS) analysis technique has been widely used to delineate azimuthal anisotropy in the upper mantle, where lattice preferred orientation of crystallographic axes of main constitute minerals such as olivine is the dominant cause of the observed anisotropy (Katayama & Karato, 2006; Savage, 1999; Silver, 1996; Zhang & Karato, 1995).

Several different methods have been utilized to measure the splitting parameters of the XKS phases, among which the transverse energy minimization method (Silver & Chan, 1991) is arguably the most reliable one for its stability for noisy data (Vecsey et al., 2008). In this method, a grid search procedure is applied to find the optimal pair of splitting parameters ($\phi$, $\delta t$) corresponding to the minimum XKS energy on the corrected transverse component. Numerous SWS studies suggest that in order to obtain reliable splitting measurements, an essential step is to visually verify all the measurements (e.g., Liu & Gao, 2013), as demonstrated in recent studies in North America (e.g., Liu et al., 2014; Yang et al., 2016, 2017; Yang et al., 2021). Due to the ever-increasing number of stations established around the world and the resultant exponential increase in the amount of data available for SWS analysis, this laborious task is increasingly time-consuming, and therefore alternate time-efficient yet reliable approaches are needed.

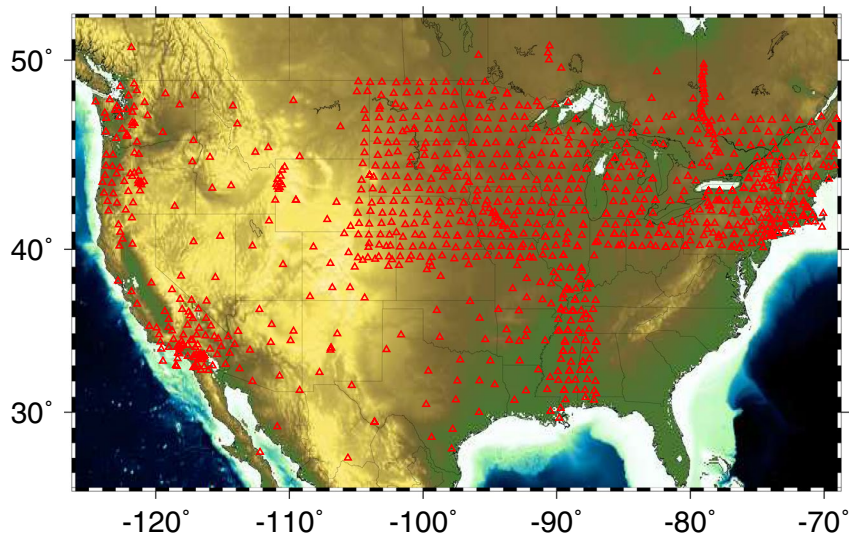


**Figure 1.** Distribution of seismic stations (red triangles; Liu et al., 2014; Yang et al., 2016, 2017) used for training the convolutional neural network.

In recent years, applications of machine learning (ML) based techniques on various scientific problems have dramatically increased, and over-human performance has been shown in diverse areas. Especially after AlphaGo showed unexpectable high performance (Silver et al., 2016), ML became widely known and attracted the attention of researchers from different fields. In geophysical research, ML has been applied in numerous studies and demonstrated considerable potential. Such applications include earthquake early warning (Li et al., 2018), earthquake detection and magnitude estimation (Lomax et al., 2019; Mousavi & Beroza, 2020; Perol et al., 2018), seismic phase picking (Dokht et al., 2019; McBrearty et al., 2019; Ross et al., 2018; Woollam et al., 2019; Zhu & Beroza, 2018), event classification (Linville et al., 2019; Titos et al., 2018), first-motion polarity determination (Ross et al., 2018), seismic denoising (Zhu et al., 2019), and earthquake prediction (Mignan & Broccardo, 2019; Rouet-Leduc et al., 2017). ML-based applications on structural seismological problems such as teleseismic tomography, receiver functions, and SWS analyses are relatively rare and are starting to takeoff (e.g., Bianco et al., 2019; Garcia et al., 2021).

In this study we design a CNN to classify the automatically determined SWS splitting parameters measured based on the set of procedures outlined in Liu and Gao (2013) into acceptable and unacceptable ones. Traditionally such kind of grouping was conducted by trained human operators and was the most time-consuming step in SWS analysis (e.g., Liu & Gao, 2013), although efforts have been made to make such processes fully automatic for both teleseismic (Link et al., 2022; Teanby et al., 2004) and local S-wave (Peng & Ben-Zion, 2004) splitting measurements with promising results. While many SWS studies group the measurements in to "good," "fair," "unacceptable," and "null," the vast majority of them use the first two categories (which are collectively called "acceptable" in this study) indistinguishably for interpretation. For the "nulls," which are characterized by a lack of XKS energy on the original transverse component while strong XKS energy appears on the radial component (e.g., Silver & Chan, 1991), we identify them in the pre-processing stage. The CNN is trained using published SWS measurements verified by human operators and is tested using splitting measurements obtained using synthetic data. The trained CNN is applied to data from 127 stations in south central Alaska and the results are compared with those reported in a recent study (Yang et al., 2021). To our knowledge, this is the first time when a ML-based technique is applied to SWS analysis.

## 2. Training Data Set and Preprocessing

Our CNN is trained and verified using 86,903 published human-labeled XKS SWS measurements recorded by 1,108 stations from Liu et al. (2014), Yang et al. (2016) and Yang et al. (2017) for the contiguous United States and adjacent areas (Figure 1). This data set was measured, manually verified, and ranked based on the procedures in Liu and Gao (2013), and contains 8,117 well-defined (ranks A and B) and 78,786 not well-defined (rank C

and rank N, which are for null measurements). The procedures start with an automatic SWS measuring step built based on the minimization of transverse energy method of Silver & Chan (1991). The automatically determined measurements are then auto-ranked based on the XKS signal to noise ratios on the original radial, original transverse, and the corrected transverse components (Liu et al., 2008). All the resulting measurements and their ranks are subsequently manually screened to validate the results based on the following criteria: goodness of the fitting between the fast and slow components, the linearity of the particle motion pattern, the robustness and uniqueness of the minimum value on the contour of remaining energy on the corrected transverse components, and the strength of the XKS arrival on the original radial and transverse components. If necessary, the XKS time window, the band-pass frequency, and the automatically determined ranking are manually adjusted to exclude non-XKS arrivals in the XKS window and to improve the signal to noise ratio.

The stations cover different tectonic areas of the North America (Figure 1) so that our training and verification data contain measurements from various geological backgrounds. Additionally, the data set contains stations with both azimuthally invariant and azimuthally varying individual measurements, indicating the presence of simple and complex anisotropy, respectively. The splitting measurement procedure takes the original radial and transverse components as the input data, grid-searches for the optimal splitting parameters, and generates corrected radial and transverse components. The optimal pair of splitting parameters produces a corrected transverse component that has the minimum XKS energy among all the candidate pairs of the splitting parameters. Our CNN takes all the four seismograms as input data and groups the optimal pair of parameters as either acceptable (ranks A and B) or unacceptable (ranks C and N). After testing with various combinations of the length and the onset of the time window for the four seismograms, a 50 s window centered at the theoretical arrival time of the XKS phase predicted based on the IASP91 Earth model is used in the study.

Acceptable measurements are labeled as array [1 0] and unacceptable measurements are labeled as [0 1]. Because the number of acceptable measurements is significantly greater than that of the unacceptable data in the training data set, to avoid overfitting, we balance the data set by setting different class weights to the two sets of data (nine for the acceptable ones and one for the unacceptable ones), which is a common practice in similar situations (Japkowicz & Stephen, 2002). After random shuffling, 80% of the measurements are used for training and 20% for validation.

## 3. Structure and Training of the CNN

The CNN is built on Keras which is a high-level neural networks application programming interface (Gulli & Pal, 2017). Following Perol et al. (2018), we designed a CNN with eight 1-D convolutional layers followed by a full-connect layer. For each splitting measurement, the input for CNN includes four seismic waveforms each with 1,000 nodes (50 s seismic waveforms with a sampling rate of 20 Hz) in length. Rectified Linear Unit is applied as the activation function between each layer (Nair & Hinton, 2010). The output is the probability of the acceptable and unacceptable measurements (Table S1 and Figure S1 in Supporting Information S1). The probability is given by Softmax which is a popular activation function for classification problems (Goodfellow et al., 2016). The equation can be shown as:

$$p(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{2} e^{x_j}} \tag{1}$$

where $j = 1, 2$ and $i = 1, 2$ represent the 2 nodes of the final layer, and $p(x)_1$ and $p(x)_2$ represent the probability of acceptable and unacceptable measurements, respectively. If the probability of the acceptable measurements is greater than a threshold, this measurement is considered as an acceptable one. Because this is a bi-class classification problem and the training data set is balanced, the threshold probability of acceptable measurements used in this study is 0.5.

To numerically reveal the difference between the CNN-predicted and human-labeled results, the cross-entropy loss is applied as the cost function (Goodfellow et al., 2016). The equation can be shown as:

$$L = -\sum_{i=1}^{n} p_i \log(q_i) \ (n = 2) \tag{2}$$

where $n$ is 2 representing the two types of measurements (acceptable and unacceptable), $p$ is the probability of the CNN predicted result and $q$ is the human-labeled result. The weights of the CNN are updated to minimize the cost function Adamalgorithm (Kingma & Ba, 2014) with a learning rate of 0.001 during each training iteration. In each iteration, 100 measurements are randomly selected to train the CNN and each measurement in the training data set is used for 64 times. The training history of accuracy and loss value (Figure S2 in Supporting Information S1) show a high level of similarity between the trends of the curves for the training and validation data sets, suggesting a low probability of overfitting. In addition, both curves become nearly flat at the highest epoch numbers, which suggests a low possibility of underfitting.

## 4. Testing With Synthetic SWS Measurements

Two sets of synthetic SWS measurements are generated based on Kong et al. (2015) to test the performance of the CNN. Firstly, the radial component of a pre-splitting XKS wave is defined as

$$R(t) = A_0 \sin(2\pi f t) e^{-\alpha t} \tag{3}$$

where $A_0 = 5000$ is the amplitude of the pre-splitting XKS wave, $f = 0.125$ Hz is the frequency, and $\alpha$ is the decaying factor which randomly changes from 0.1 to 0.5. After penetrating the anisotropic layer, which has a fixed $\phi$ of 0° and a randomly assigned $\delta t$ ranging from 0.5 to 2.0 s, the shear wave splits into the fast and slow components

$$S_f(t) = R(t)\cos(\theta) \tag{4}$$

and

$$S_s(t) = -R(t - \delta t)\sin(\theta) \tag{5}$$

where $S_f$ and $S_s$ are the fast and slow waves, and $\theta$ is the angle between $\phi$ and the back azimuth (BAZ) of the event. The epicentral distances of the events are randomly assigned in the range of 90°–120°, and the focal depths vary from 20 to 50 km. Finally, $S_f$ and $S_s$ are projected to the north-south and east-west components and random noise is added to generate synthetic seismograms for SWS measurements. The signal-to-noise ratio (SNR) is defined as $\max(R(t))/\max(N(t))$, where $N(t)$ is the trace of the random noise.

It can be demonstrated mathematically that the energy on the original transverse component and the reliability of the resulting splitting parameters are dependent on $\theta$, which is the angular difference between the BAZ and the fast orientation (Silver & Chan, 1991). In the modulo-90° domain, close-to-null measurements, which are characterized by hardly observable XKS energy on the original transverse component, are dominant if $\theta$ is less than 15° or greater than 75°. The relationship between $\theta$ and the reliability of the measurements, as well as the influence of $\theta$ on CNN's ability to correctly separate the acceptable measurements from the unacceptable ones, can be quantified using synthetic data.

For this purpose, we produce 72 groups of synthetic seismograms with SNR ranging from 4 to 10. Each group has 1,000 measurements. The BAZ (which equals to $\theta$ because the fast orientation is set as 0° in the model) of the $n$th group is ($n$-1)*5°. The other parameters, including the SNR, used for generating the synthetic seismograms in Equations 3–5 are the same among the groups. The same data processing procedures used to generate the uniform SWS database for North America (Liu et al., 2014; Yang et al., 2016, 2017) are applied to the synthetic waveforms to automatically determine the splitting measurements, which are then grouped by the trained CNN into acceptable and unacceptable categories. The results suggest that the number of CNN-accepted measurements reduces rapidly when the BAZ approaches 0° and 90° in the modulo−90° domain (Figure 2a), due to the weak XKS energy (relative to that of the noise) on the original transverse component. Such measurements are either ranked as C or N by human operators, depending on the noise level and the preference of the operators. When the BAZ is ≥15° from the fast or slow orientations in the modulo−90° domain, clear XKS energy is present on the original transverse component, and consequently the trained CNN successfully identified almost all the measurements with a rate of success >99% (Figure 2a).

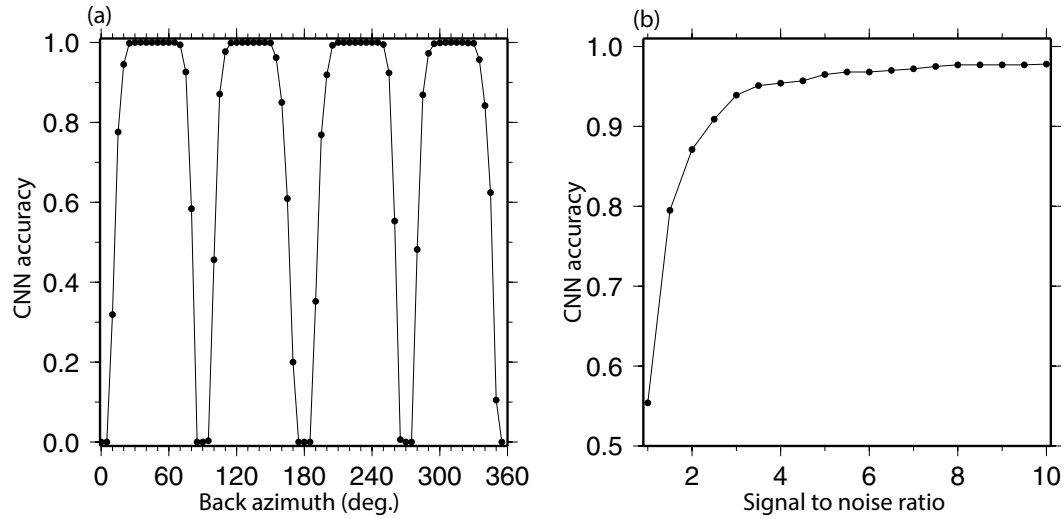


**Figure 2.** Performance tests of the convolutional neural network (CNN) using synthetic data set. (a) 72 groups of synthetic shear wave splitting (SWS) measurements with different back azimuths. (b) 20 groups of synthetic SWS measurements with different signal-to-noise ratios.

To test the performance of the CNN on data with different SNR values, 20 groups of synthetic SWS measurements are generated, each with 1,000 measurements. The SNR of each group varies from 1 to 10 with an interval of 0.5. Based on the results of the first test (Figure 2a), the BAZ of each of the events used for the test is at least 15° from the fast or slow orientation. The results show that the success rate of CNN is over 90% for SNR ≥2.5 and over 97% when SNR ≥6.5 (Figure 2b).

## 5. Application to SWS Measurements in South Central Alaska

We next apply the trained CNN to broadband seismic data in south central Alaska recorded by 127 stations with variable recording period from 1988 to October 2019. The procedures to request (from the Data Management Center of the Incorporated Research Institutions for Seismology) and preprocess the XKS data follow those described in Liu and Gao (2013) and are identical to those used by Yang et al. (2021). In total 19,960 pairs of splitting parameters are obtained at 127 Stations (Figure 3a). The SNR based auto-ranking procedure of Liu et al. (2008) ranked 6,314 measurements from 127 stations as potentially acceptable (Figure 3b) and 13,646 measurements as unacceptable. Note that the Liu et al. (2008) approach was intentionally designed for excluding only the measurements that are impossible to be acceptable, for the purpose of reducing the amount of effort for the subsequent step of visual screening and at the same time minimizing the risk of missing potentially acceptable measurements.

The trained CNN is then applied to classify the 6,314 pairs of splitting parameters (Figure 3b). When a threshold value of 0.5 for the probability to be acceptable is used, a total of 2668 pairs from 124 stations are determined as acceptable (Figure 3c). We apply three set of conditions to refine the CNN-selected results. These conditions are necessary to exclude false positives caused by limitations in the current CNN, as described in Section 6.2 below. First, because the synthetic tests show that a measurement cannot be reliably classified by CNN (and human operators, as discussed below) when the difference between the BAZ and the fast orientation is less than ∼15° (Figure 2a), we excluded such near-null measurements, and the remaining data set contains 2387 pairs of measurements from 122 stations (Figure 3d). Second, measurements with large standard deviations (≥15° for $\phi$ or ≥1.5 s for $\delta t$) are excluded, resulting in a total of 1,751 pairs of measurements from 118 stations (Figure 3e). Third, because measurements with large splitting times are rarely found in SWS studies in Alaska and elsewhere and are frequently associated with erroneously determined splitting parameters, we remove the 221 pairs (or 12.6%) of measurements with splitting times ≥2.0 s, leaving 1,530 pairs of measurements from 115 stations in the final data set (Figure 3f).

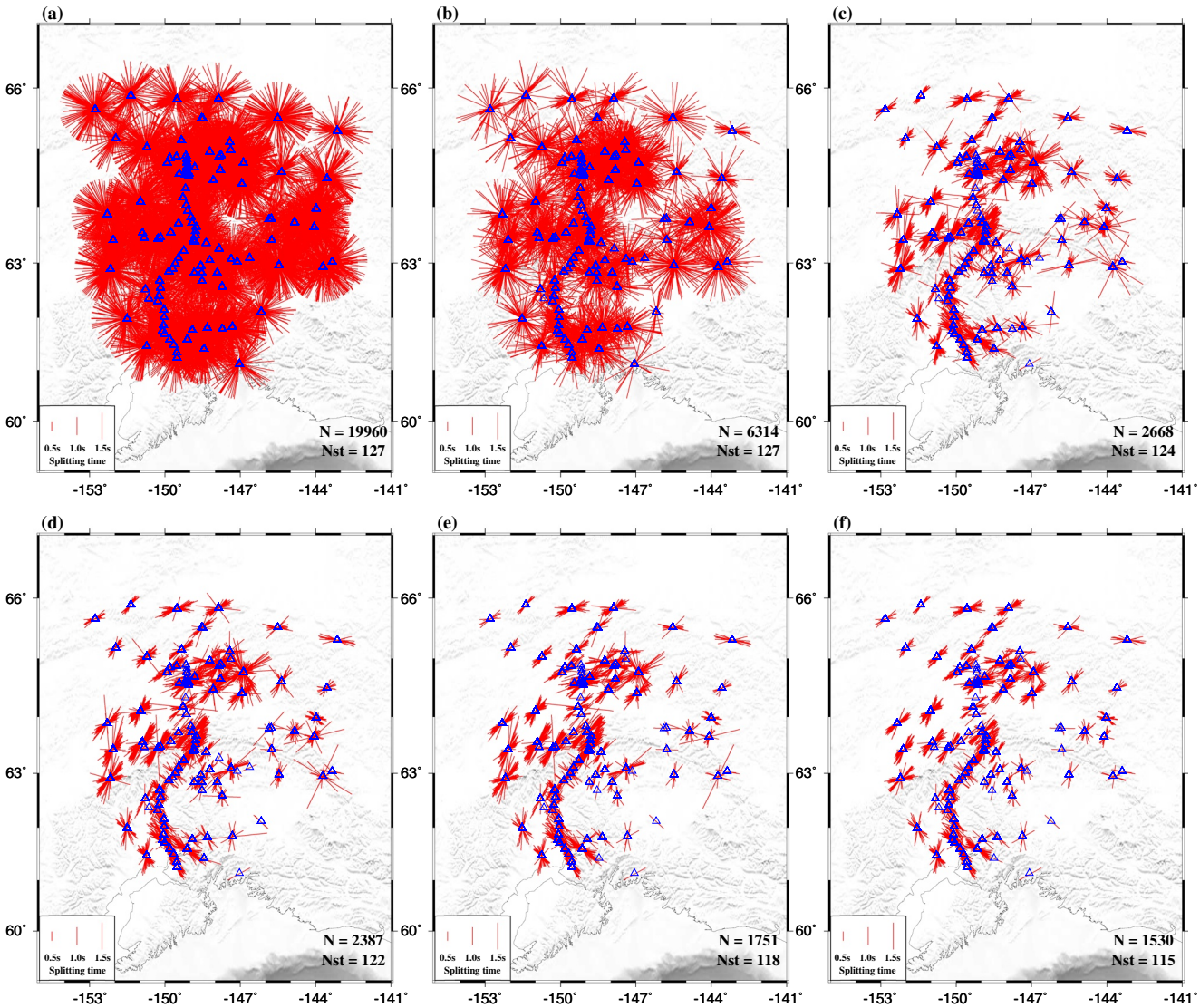


**Figure 3.** SWS measurements (red bars) in south central Alaska plotted at the recording stations (blue triangles). The orientation of the bars represents the fast orientation, and the length is proportional to the splitting time. (a) All the measurements recorded by stations in the study area. (b) Results of auto-ranking based on the approach of Liu et al. (2008) which was designed as a pre-screening step to reduce human workload in the subsequent manual screening step. (c) Results of convolutional neural network with a threshold of 0.5. (d) Same as (c) by after removing measurements for which the angular difference between the back azimuth and the fast or slow wave polarization orientations is smaller than 15°. (e) Same as (d) but after removing measurements with standard deviation of $\phi > 15°$ or standard deviation of $\delta t > 1.5$ s. (f) Same as (e) but after removing measurements with $\delta t > 2.0$ s. The number of measurements (N) and the number of stations (Nst) are shown in the lower right corner of each plot.

Understandably, as more conditions are applied to the CNN classified measurements, the number of remaining measurements reduces, while the consistency among the measurements at the stations increases (Figures 3c–3f). Note that due to the existence of complex anisotropy especially in the central and southern portions of the study area (Yang et al., 2021), variability among the measurements at the same stations is present at some of the stations. Such a variability does not necessarily indicate inaccurately determined results but are mostly the result of azimuthal and piercing-point variations of the splitting parameters. The former is usually an indicator of complex anisotropy (Rümpker & Silver, 1998; Silver & Savage, 1994), and the latter is the result of a station being located at the boundary of two or more regions with different characteristics of anisotropy (Alsina & Snieder, 1995; Jia et al., 2021).

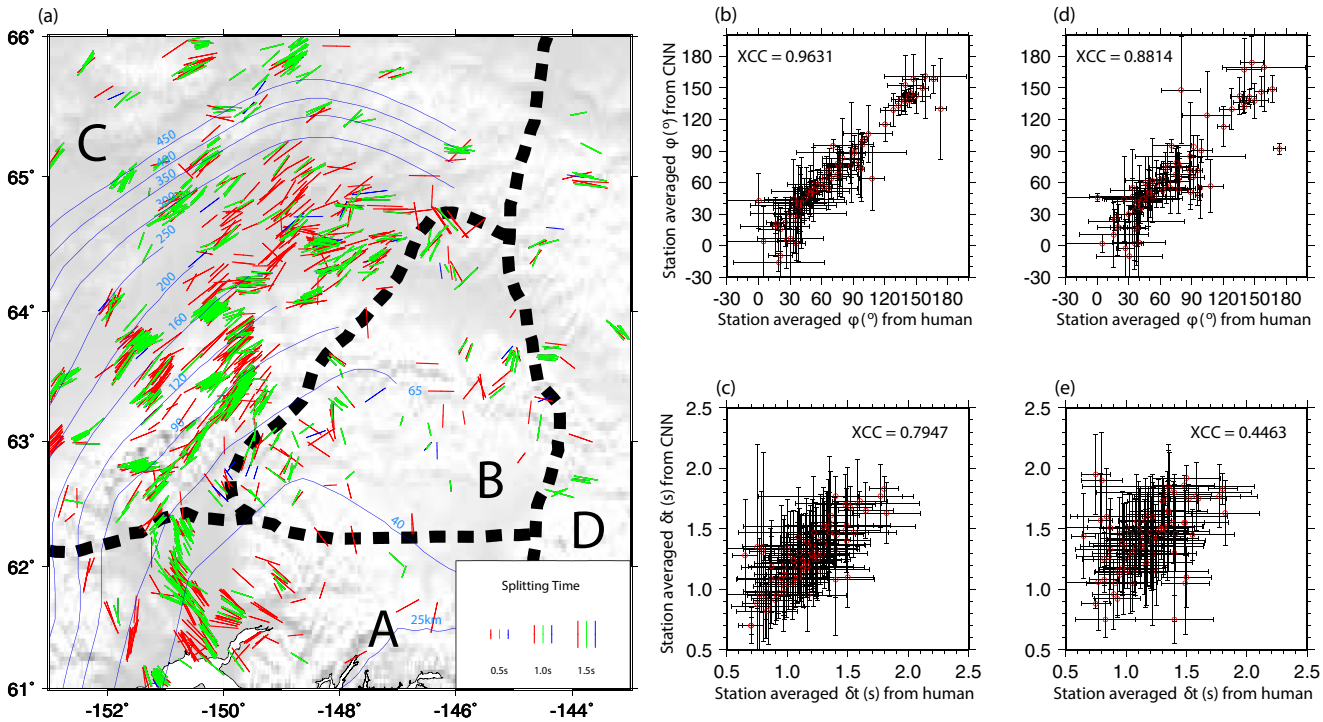


**Figure 4.** (a) Comparison of human-determined (blue bars; Yang et al., 2021) and convolutional neural network (CNN)-selected (red bars) shear wave splitting measurements in south central Alaska. Green bars are measurements accepted by both CNN and human operators. All the measurements are plotted above the XKS ray-piercing points at 200 km deep which is the most likely depth of the anisotropic layer (Yang et al., 2021). The contour lines show the depth of the subducted Pacific slab, and the thick dashed lines separate four regions (a–d) with different patterns of splitting measurements (Yang et al., 2021). The CNN results are the same as those shown in Figure 3f. (b) Cross-plot of human and CNN determined station-averaged $\phi$ measurements. The black bars are the standard deviation. (c) Same as (b) but for station-averaged $\delta t$ measurements. (d and e) are respectively the same as (b and c) but the CNN station averages were computed using results that were determined by CNN as acceptable but rejected by the human operators. XCC: Cross-correlation coefficient.

## 6. Discussion

### 6.1. Comparison With Human-Determined Measurements

After manually verifying the automatically ranked measurements (Figure 3b), Yang et al. (2021) obtained 971 measurements from 106 stations, among which 952 (98.1%) are within the 2668 measurements classified as acceptable by CNN before the application of the three conditions (Figure 3c). This suggests that if a human operator uses the CNN accepted results (Figure 3c) rather than results from the SNR based ranking system (Figure 3b) as the starting point for manual verification, an approximately 60% reduction in the number of measurements to be verified will be achieved, and only less than 2% of the measurements will be missed.

To objectively compare the human (Yang et al., 2021) and CNN determined final results (Figure 3f), the same set of three conditions applied to the CNN-accepted data are applied to the 971 human-determined measurements, and the remaining human-determined data set contains 865 measurements from 102 stations (Figure 4a), among which 816 from 100 stations are in the final data set that CNN determined (Figure 3f). In other words, CNN missed merely 49 (5.7%) of the human determined measurements. Most of the missed measurements are in Area B (Figure 4a) where the interaction of two flow systems with nearly orthogonal directions leads to weak anisotropy with small splitting times (Yang et al., 2021).

In spite of the fact that the station averaged splitting parameters from the CNN and human determined results show a high similarity with a cross-correlation coefficient (XCC) of 0.9631 for $\phi$ and 0.7947 for $\delta t$ (Figures 4b and 4c), the number of measurements in the final CNN-determined data set (Figure 3f) is about twice as many as that in the human-determined data set. This difference might be caused by the fact that the human operators applied a stricter set of standards when verifying the measurements. To test this possibility, we computed station averages using only the CNN determined measurements that were not selected by the human operators.

The results (Figures 4d and 4e) show a reduced similarity between the two sets of data, with reduced XCC values of 0.8814 for $\phi$ and 0.4463 for $\delta t$. The most obvious explanation for this reduced similarity is that some measurements with a marginal quality were classified as acceptable by CNN but were rejected by the human operators.

### 6.2. Comparison With a Fully Automated Non-Machine Learning SWS Measurement Approach

Several non-machine learning methods have been proposed to measure SWS parameters in a completely automated manner (e.g., Link et al., 2022; Teanby et al., 2004). Here, we choose the latest one, SplitRacerAUTO (Link et al., 2022), to compare with the CNN-based approach proposed in this study. The MATLAB-based SplitRacerAUTO can automatically select the XKS time window and categorize the splitting measurements. The same data set in Alaska is applied to test the performance of SplitRacerAUTO against our ML-based approach. To be consistent with the parameters used in our preprocessing step, we used the frequency range of 0.04–0.5 Hz for the band-pass filter in SplitRacerAUTO, and kept all the other parameters the same as the default values. The results show that this method accepts 950 measurements from 110 stations, and 467 of them (48.1%) are contained in results of Yang et al. (2021). After applying the same constraints that produced Figure 3f, there are 586 measurements remained and 404 (46.7%) of them are in Yang et al. (2021). In other words, the automated procedure missed 53.3% of the human-determined measurements, while our CNN-based approach merely missed 5.7% of them. The results of the automatically determined splitting parameters and results after various constraints are plotted in Figure S4 of Supporting Information S1 using the same style as Figure 3 for easy comparison. Similarly, comparisons of human and SplitRacerAUTO determined measurements similar to Figure 4 are plotted in Figure S5 of Supporting Information S1. Comparing Figures 4 and S5 in Supporting Information S1, it is clear that the CNN-based approach resulted in a significantly greater number of measurements than SplitRacerAUTO, especially in areas with weaker anisotropy such as Area B in Figures 4 and S5 in Supporting Information S1.

### 6.3. Limitations of the Current CNN and Suggested Next Steps

In spite of the satisfactory performance of our CNN on both synthetic and real data, a major drawback of the current CNN is that it does not have the capability to adjust the data processing parameters including the beginning and end times of the XKS window and the bandpass filtering frequencies. For a small portion of the measurements, such adjustments are required in order to obtain reliable results. For instance, if the epicentral distance is smaller than 90°, part of the S wave energy can be included in the XKS window, leading to unreliable results. Therefore, it is necessary to design and train a CNN that can automatically recognize the necessity and make such adjustments. One of the approaches is to design a separate pre-processing CNN for picking the arrival time of the XKS arrival, similar to those designed for picking the onset time of P or S waves from local events (Zhu & Beroza, 2018). Additional work is needed to find the optimal ending time of the XKS window, and to detect the dominant frequency range of the noise and perform band pass filtering to enhance the SNR when strong noise is present. Alternatively, the optimal XKS window can be determined during the pre-processing stage using non-CNN based approaches such as the time-frequency spectrum technique recently proposed by Link et al. (2022).

## 7. Conclusions

In this study, we have established a CNN to automatically classify teleseismic SWS measurements. The CNN is trained by published human-labeled datasets and tested using synthetic SWS measurements to evaluate its performance against different levels of noise and its dependence on the difference between the fast orientation and the back-azimuth of the events. When the SNR is greater than 6.5, more than 97% of the non-null synthetic measurements can be correctly accepted by the CNN. Application of the CNN to data from south central Alaska shows that it can classify almost all human-accepted measurements (98.1%) as acceptable when a threshold probability of 0.5 is used. The study suggests a high potential for CNN-based methods to significantly improve the efficiency of measuring SWS parameters.

## Data Availability Statement

The Python codes to train and test the CNN have been uploaded to GitHub under the address https://github.com/YW-Zhang94/CNN-SWS.git, and the seismic data used in the seismic data used in the study (in SAC-Seismic Analysis Codes format with a total size of about 2.1 GB) can be found under https://figshare.com/articles/dataset/CNN_SWS_data/19904833. All the waveform data used in the study are openly accessible from the IRIS Data Management Center (https://ds.iris.edu/ds/nodes/dmc/, last accessed March 2019), under the main network codes of AK (https://doi.org/10.7914/SN/AK), AT (https://doi.org/10.7914/SN/AT), DW (https://doi.org/10.7914/SN/DW), IM (International Miscellaneous Stations), IU (https://doi.org/10.7914/SN/IU), TA (https://doi.org/10.7914/SN/TA), XE (https://doi.org/10.7914/SN/XE_1999), XR (https://doi.org/10.7914/SN/XR_2004), XV (https://doi.org/10.7914/SN/XV_2014), YE (https://doi.org/10.7914/SN/YE_2011), and YV (https://doi.org/10.7914/SN/YV_2006).

**Acknowledgments**
The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms and related metadata used in this study. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048. We thank Frederik Link, an anonymous reviewer, and Editor Daoyuan Sun for constructive reviews that significantly improved the manuscript. The study was partially supported by the U.S. National Science Foundation under awards 1830644 and 1919789 to S.G.

## References

Alsina, D., & Snieder, R. (1995). Small-scale sublithospheric continental mantle deformation: Constraints from SKS splitting observations. *Geophysical Journal International*, *123*(2), 431–448. https://doi.org/10.1111/j.1365-246X.1995.tb06864.x

Ando, M., Ishikawa, Y., & Yamazaki, F. (1983). Shear wave polarization anisotropy in the upper mantle beneath Honshu, Japan. *Journal of Geophysical Research*, *88*(B7), 5850–5864. https://doi.org/10.1029/JB088iB07p05850

Bianco, M. J., Gerstoft, P., Olsen, K. B., & Lin, F. C. (2019). High-resolution seismic tomography of Long Beach, CA using machine learning. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-50381-z

Dokht, R. M., Kao, H., Visser, R., & Smith, B. (2019). Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters*, *90*(2A), 481–490. https://doi.org/10.1785/0220180308

Garcia, J. A., Waszek, L., Tauzin, B., & Schmerr, N. (2021). Automatic identification of mantle seismic phases using a Convolutional Neural Network. *Geophysical Research Letters*, *48*(18), e2020GL091658. https://doi.org/10.1029/2020GL091658

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine learning basics. *Deep learning*, *1*(7), 98–164.

Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449. https://doi.org/10.3233/IDA-2002-6504

Jia, Y., Liu, K. H., Kong, F., Liu, L., & Gao, S. S. (2021). A systematic investigation of piercing-point-dependent seismic azimuthal anisotropy. *Geophysical Journal International*, *227*(3), 1496–1511. https://doi.org/10.1093/gji/ggab285

Katayama, I., & Karato, S. I. (2006). Effect of temperature on the B-to C-type olivine fabric transition and implication for flow pattern in subduction zones. *Physics of the Earth and Planetary Interiors*, *157*(1–2), 33–45. https://doi.org/10.1016/j.pepi.2006.03.005

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kong, F., Gao, S. S., & Liu, K. H. (2015). A systematic comparison of the transverse energy minimization and splitting intensity techniques for measuring shear-wave splitting parameters. *Bulletin of the Seismological Society of America*, *105*(1), 230–239. https://doi.org/10.1785/0120140108

Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773–4779. https://doi.org/10.1029/2018GL077870

Link, F., Reiss, M. C., & Rumpker, G. (2022). An automatized XKS-splitting procedure for large data sets: Extension package for SplitRacer and application to the USArray. *Computers & Geosciences*, *158*, 104961. https://doi.org/10.1016/j.cageo.2021.104961

Linville, L., Pankow, K., & Draelos, T. (2019). Deep learning models augment analyst decisions for event discrimination. *Geophysical Research Letters*, *46*(7), 3643–3651. https://doi.org/10.1029/2018GL081119

Liu, K. H., Elsheikh, A., Lemnifi, A., Purevsuren, U., Ray, M., Refayee, H., et al. (2014). A uniform database of teleseismic shear wave splitting measurements for the western and central United States. *Geochemistry, Geophysics, Geosystems*, *15*(5), 2075–2085. https://doi.org/10.1002/2014GC005267

Liu, K. H., & Gao, S. S. (2013). Making reliable shear-wave splitting measurements. *Bulletin of the Seismological Society of America*, *103*(5), 2680–2693. https://doi.org/10.1785/0120120355

Liu, K. H., Gao, S. S., Gao, Y., & Wu, J. (2008). Shear wave splitting and mantle flow associated with the deflected Pacific slab beneath northeast Asia. *Journal of Geophysical Research*, *113*, B01305. https://doi.org/10.1029/2007JB005178

Lomax, A., Michelini, A., & Jozinović, D. (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismological Research Letters*, *90*(2A), 517–529. https://doi.org/10.1785/0220180311

Long, M. D., & Silver, P. G. (2009). Shear wave splitting and mantle anisotropy: Measurements, interpretations, and new directions. *Surveys in Geophysics*, *30*(4), 407–461. https://doi.org/10.1007/s10712-009-9075-1

McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, *90*(2A), 503–509. https://doi.org/10.1785/0220180326

Mignan, A., & Broccardo, M. (2019). One neuron versus deep learning in aftershock prediction. *Nature*, *574*(7776), E1–E3. https://doi.org/10.1038/s41586-018-0438-y

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019GL085976

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. *Proceedings of the 27th international conference on international conference on machine learning*, 807–814.

Peng, Z., & Ben-Zion, Y. (2004). Systematic analysis of crustal anisotropy along the Karadere–Duzce branch of the North Anatolian fault. *Geophysical Journal International*, *159*, 253–274. https://doi.org/10.1111/j.1365-246X.2004.02379.x

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578. https://doi.org/10.1126/sciadv.1700578

Ross, Z. E., Meier, M. A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*, 5120–5129. https://doi.org/10.1029/2017JB015251

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282. https://doi.org/10.1002/2017GL074677

Rümpker, G., & Silver, P. G. (1998). Apparent shear-wave splitting parameters in the presence of vertically varying anisotropy. *Geophysical Journal International*, *135*(3), 790–800. https://doi.org/10.1046/j.1365-246X.1998.00660.x

Savage, M. K. (1999). Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Reviews of Geophysics*, *37*(1), 65–106. https://doi.org/10.1029/98RG02075

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484. https://doi.org/10.1038/nature16961

Silver, P. G. (1996). Seismic anisotropy beneath the continents: Probing the depths of geology. *Annual Review of Earth and Planetary Sciences*, *24*(1), 385–432. https://doi.org/10.1038/nature16961

Silver, P. G., & Chan, W. W. (1991). Shear wave splitting and subcontinental mantle deformation. *Journal of Geophysical Research*, *96*(B10), 16429–16454. https://doi.org/10.1029/91JB00899

Silver, P. G., & Savage, M. K. (1994). The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers. *Geophysical Journal International*, *119*(3), 949–963. https://doi.org/10.1111/j.1365-246X.1994.tb04027.x

Teanby, N. A., Kendall, J. M., & van der Baan, M. (2004). Automation of shear-wave splitting measurements using cluster analysis. *Bulletin of the Seismological Society of America*, *94*(2), 453–463. https://doi.org/10.1785/0120030123

Titos, M., Bueno, A., Garcia, L., & Benitez, C. (2018). A deep neural networks approach to automatic recognition systems for volcano-seismic events. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(5), 1533–1544. https://doi.org/10.1109/JSTARS.2018.2803198

Vecsey, L., Plomerová, J., & Babuška, V. (2008). Shear-wave splitting measurements-Problems and solutions. *Tectonophysics*, *462*(1–4), 178–196. https://doi.org/10.1016/j.tecto.2008.01.021

Woollam, J., Rietbrock, A., Bueno, A., & De Angelis, S. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, *90*(2A), 491–502. https://doi.org/10.1785/0220180312

Yang, B. B., Liu, K. H., Dahm, H. H., & Gao, S. S. (2016). A uniform database of teleseismic shear-wave splitting measurements for the western and central United States: December 2014 update. *Seismological Research Letters*, *87*(2A), 295–300. https://doi.org/10.1785/0220150213

Yang, B. B., Liu, Y., Dahm, H., Liu, K. H., & Gao, S. S. (2017). Seismic azimuthal anisotropy beneath the eastern United States and its geodynamic implications. *Geophysical Research Letters*, *44*(6), 2670–2678. https://doi.org/10.1002/2016GL071227

Yang, Y., Gao, S. S., Liu, K. H., Kong, F., & Fu, X. (2021). Mantle flow in the vicinity of the eastern edge of the Pacific-Yakutat slab: Constraints from shear wave splitting analyses. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB022354. https://doi.org/10.1029/2021JB022354

Zhang, S., & Karato, S. I. (1995). Lattice preferred orientation of olivine in simple shear deformation and the flow geometry of the upper mantle of the Earth. *Nature*, *375*, 774–777. https://doi.org/10.1038/375774a0

Zhu, W., & Beroza, G. C. (2018). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476–9488. https://doi.org/10.1109/TGRS.2019.2926772