

JOURNAL ARTICLE

Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning ^{FREE}

F Civilini ✉, R C Weber, Z Jiang, D Phillips, W David Pan

Geophysical Journal International, Volume 225, Issue 3, June 2021, Pages 2120–2134, <https://doi.org/10.1093/gji/ggab083>

Published: 15 March 2021 **Article history** ▼

SUMMARY

The costly power requirements of delivering seismic data back to Earth from planetary missions requires the development of algorithms for lander-side signal analysis for telemetry prioritization. This is difficult to explicitly program, especially if no prior seismic data are available from the planetary body. Deep learning computer vision has been used to generalize seismic signals on Earth for earthquake early warning problems but such techniques have not yet been expanded to planetary science. We demonstrate that Convolutional Neural Networks can be used to accurately catalogue planetary seismicity without local training data by building binary noise/signal classifiers from a single Earth seismic station and applying the models to moonquakes from the Apollo Passive Seismic Experiment (PSE) and the Lunar Seismic Profiling Experiment (LSPE). In order to promote generality and reduce the amount of training data, the algorithms use spectral images instead of time-series. Two- to five-layer convolution models are tested against a subset of 200 Grade-A events from the PSE and obtained station accuracy averages of 89–96 per cent. As the model was applied to an hour trace of data (30 min before and after the Grade-A event), additional detections besides the Grade-A event are unavoidable. In order to comprehensively address algorithm accuracy, additional seismic detections corresponding to valid signals such as other moonquakes or multiples within a particularly long event needed to be compared with those caused by algorithm error or instrument glitches. We developed an ‘extra-arrival accuracy’ metric to quantify how many of the additional

detections were due to valid seismic events and used it to select the three-layer model as the best fit. The three-layer model was applied to the entire LSPE record and matched the lunar day–night cycle driving thermal moonquake generation with fewer false detections than a recent study using Hidden Markov Models. We anticipate that these methods for lander-side signal detection can be easily expanded to non-seismological data and may provide even stronger results when supplemented with synthetic training data.

Keywords: Neural networks, fuzzy logic, Earthquake monitoring and test-ban treaty verification, Seismicity and tectonics

Subject: Seismology

Issue Section: Seismology

1 INTRODUCTION

Research in planetary seismology is fundamentally constrained by a lack of data. Until the recent detection of marsquakes during the Mars InSight mission (Giardini *et al.* 2020), the only other measurements of seismicity recorded outside of Earth was on the Moon during the Apollo missions between 1969 and 1977 (Lognonné 2005). These unique data sets have been periodically revisited using new seismological methods, including (i) ambient noise cross-correlations to determine shallow structure (Larose *et al.* 2005), (ii) array processing methods for core structure recovery (Weber *et al.* 2011) and (iii) Hidden Markov Models for moonquake detection and classification (Knapmeyer-Endrun & Hammer 2015; Dimech *et al.* 2017).

Seismic data from future missions can typically only be informed by theoretical signal-to-noise characteristics of the environment [e.g. Panning *et al.* (2018) for Europa; Mimoun *et al.* (2017) for Mars; and Lorenz *et al.* (2018) for Titan] or likely Earth-analogues [e.g. Zhan *et al.* (2014) with Antarctica as a Europa analogue]. Although objectives can be re-assessed after some initial data-collection upon lander arrival,

transfer of high-resolution data back to Earth is costly on lander power usage. The energy cost for a modern high-gain antenna is approximately 5 mJ/bit/AU (Lorenz 2015), and can be used to assess telemetry power requirements. The SEIS instrument package for the Mars InSight mission has a data allocation threshold of 38 Mbits/sol after compression (Lognonné *et al.* 2019), which takes approximately 0.133 MJ or 36.94 Watt-hours for telemetry back to Earth (0.7 AU). This represents about 1.5 per cent of the 2400 Watt-hours/sol output of the InSight lander (M. Panning, personal communication, 2019). Assuming a similar configuration to InSight, a mission to Europa (4–6 AU) would spend 211–316 Watt-hours or 8.8–13.2 per cent of the entire lander power output on telemetry data of the seismic instrument alone. This constraint is especially important as seismologists will likely be sharing the lander with science teams from other disciplines and with different objectives, some of which may be transferring much larger data.

This is a significant problem because typical earthquake detection algorithms used on Earth may not be suitable for planetary seismological studies. For example, a typical algorithm used for automatic earthquake detection on Earth is the short-term average and long-term average ratio (STA/LTA), which measures the differences of energy in traces of the time domain and marks a detection when this ratio surpasses a previously established threshold (Allen 1982). Although this algorithm requires low computational resources, it is entirely dependent on prior information of both the background noise and amplitude ranges of the expected seismic events, both of which will not be known in most planetary cases.

A much more accurate technique for seismic detection is template matching, which uses the correlation coefficient to detect seismic events in continuous data which match known templates (Gibbons & Ringdal 2006). Although this methodology can detect events with low signal-to-noise ratio, it is a time-domain method very dependent on the quality of the cross-correlation coefficient between the seismic signal and the template. Finally, the autocorrelation method detects earthquakes by autocorrelating segments of continuous data based on the principle that earthquakes are more similar to themselves

compared to noise (Brown *et al.* 2008). The high accuracy of this method is offset by a large computation requirement.

Machine learning is a general term used to describe a set of algorithms that extract information directly from data. Deep learning is a subset of machine learning based on the Universal Approximation Theorem, which states that any function can be approximated through a sum based on a linear operation on the inputs within a single non-linear function (Cybenko 1989). Machine learning algorithms have been applied to seismology classification problems over the past few decades with promising initial results. A binary classifier was built by Dysart & Pulli (1990) that could successfully distinguish between regional earthquakes and chemical explosions using spectral ratios. A similar classifier was developed to detect the differences between microearthquakes and quarry blasts by Musil & Plešinger (1996). Dai & MacBeth (1995) developed a neural network to distinguish *P* and *S* waves from noise in the time domain, but their algorithm was frequently susceptible to false arrivals. Due to small training sets and low computational power, these algorithms were primarily successful on very specialized cases and could not be extrapolated to other data sets.

Over the last several years, the development of GPU computing techniques, open-source high-level APIs such as *Keras* (Chollet *et al.* 2015) and *fastai* (Howard *et al.* 2018), and increased data collection from sensors such as smartphones have led to rapid advances in machine learning within the fields of computer vision, natural language processing and collaborative filtering. These new algorithms and workflows were quickly applied to seismology for a variety of tasks, including: earthquake detection (Meier *et al.* 2019), ground-motion prediction (Trugman & Shearer 2018) and tomography (Li & Harris 2018).

In this paper, we assess the accuracy of deep learning methods for planetary seismicity detection by using non-local training data. We build a binary classifier trained on spectrograms of Earth seismic events, test its accuracy on labelled seismic arrivals within the Apollo Passive Seismic (PSE) data set, and use it to assess the frequency of thermal moonquakes recorded within the Apollo 17 Lunar Seismic

Profiling Experiment (LSPE). The process of applying a previously built model to a new task or data set is known as ‘transfer learning’ (Caruana 1995). The arrivals are compared to human-picked signal reference times (Nakamura *et al.* 1981) and those obtained using Hidden Markov Models, an algorithm for temporal pattern recognition (Knapmeyer-Endrun & Hammer 2015; Dimech *et al.* 2017).

Moonquakes within the LSPE data, which are expected to be primarily thermal in origin (Duennebier & Sutton 1974), have not been systematically catalogued (Lognonné 2005; Dimech *et al.* 2017). A catalogue of thermal moonquakes can provide insight into the dynamics of the lunar regolith (e.g. Duennebier 1976), and detected non-thermal moonquakes can supplement PSE records and inform structure studies, which are crucial for formation and evolution models. Most importantly, a lightweight classifier that can discern between data and noise could be used on-board a future lander to prioritize data transfer despite not being trained with a local data set. This would have tremendous impact on the kinds of data, seismic or otherwise, that could be retrieved in future missions.

2 BACKGROUND

2.1 Geologic setting, seismicity and instrumentation

The Moon is thought to have formed towards the end of Earth’s accretion phase following an impact with a body of approximately 1/10th the mass of Earth (Canup 2004). Its structure is composed of a 34–43-km-thick crust which is thinner on the near side of the Moon, an underlying mantle which extends to 1450 km depth and a 250–430 km core differentiated into solid and liquid components (Lognonné *et al.* 2003; Wiczorek *et al.* 2012). The surface geology is distinguished by two main features: the highlands, composed of anorthositic the maria, fine-grained basalt rich in iron, magnesium and titanium (de Pater & Lissauer 2010).

Detailed investigations of the lunar subsurface were possible due to the installation of a four-instrument seismic network deployed during the Apollo missions (1969–1972) that kept transmitting data back to Earth until 1977, referred to as the Apollo Passive Seismic Experiment (PSE, Simmons *et al.* 1970). Each station was composed of a 3-axis long-period and a vertical-axis Short Period seismometer with centre frequencies of 0.45 and 8 Hz, respectively (Lognonné 2005; Nunn *et al.* 2020). Over 13 000 moonquakes have been detected in the long-period data, and more undetected signals are probably within the short-period records (Nakamura 2003; Lognonné 2005). It is important to note that the marked moonquake event times in the PSE catalogue are general times to direct human researchers on the location of the signal instead of precise wave arrivals (Nakamura *et al.* 1981).

Moonquakes are largely characterized by a strong body wave coda and a lack of dispersive surface waves induced by scattering properties of the near surface (Dainty & Toksöz 1981). There are four main types of seismic activity recorded on the Moon: (1) impacts from meteoroids or artificial sources (Oberst & Nakamura 1987), (2) deep quakes (between 700 and 1200 km) due to tidal stresses (Nakamura 2005), (3) high-frequency shallow quakes (<200 km) which have been associated by Nakamura *et al.* (1979) and later work with fault systems near deep impact basins or lobate scarps (Watters *et al.* 2019) and (4) rock fracturing due to thermal stresses induced by solar illumination (Duennebier & Sutton 1974).

Four seismometers recording at a sample-rate of 117.78 Hz (8-bit instruments, 5 volt dynamic range) were deployed in 1972 as part of the LSPE in a triangular configuration and left in listening mode until 1977 (Kovach & Watkins 1973; Bates *et al.* 1979; Heffels *et al.* 2017; Dimech *et al.* 2017; Nunn *et al.* 2020, Fig. 1). The purpose of the experiment was to determine regional shallow structure by recording seismic waves produced by explosive charges, the thrust of the lunar module ascent engine and the crash of the lunar module ascent stage (Kovach & Watkins 1973). Thermal moonquakes are hypothesized to make the bulk of the LSPE signals due to their high frequency, but those records have not been fully explored (Dimech *et al.* 2017).

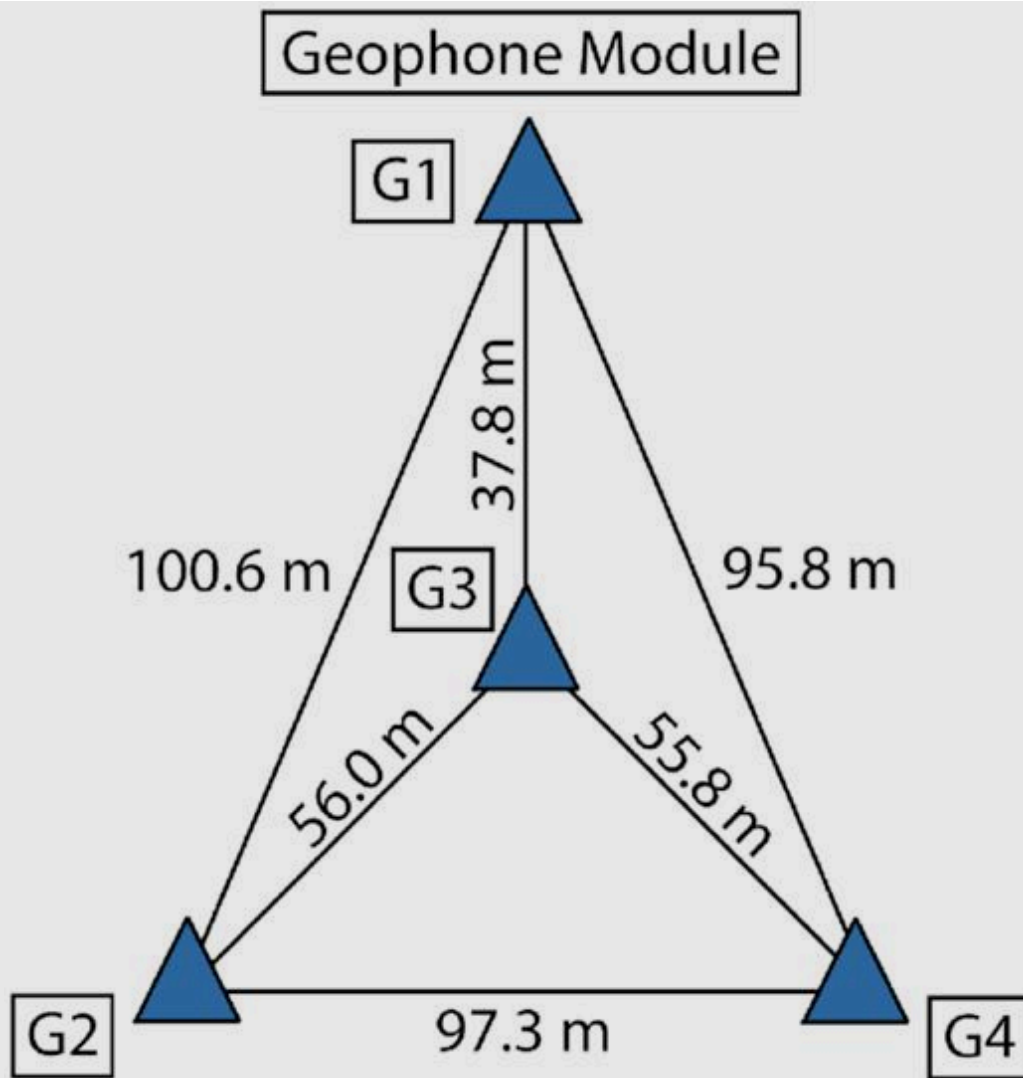


Figure 1. Schematic of geophones used in the LSPE. The geophones are displayed as blue triangles with labels G1 through G4, respectively. Distances between each geophone were obtained from Heffels *et al.* (2017).

Seismic data from the LSPE has low signal-to-noise ratio, a limited dynamic range and instrument glitches. Due to these technical hurdles and the emergent characteristic of certain moonquake arrivals, traditional methods for seismic detection do not work well and researchers typically carry out manual detections. Fig. 2 displays STA/LTA for an hour trace of Geophone 1 on 1976-08-17. Subplots (a) and (c) display STA/LTA results and values using a trigger on (red lines) value of 1.5, a trigger off (blue lines) value of 0.5 and windows of 10 and 20 s. Subplots (b) and (d) display results using the same triggers but windows of 20 and 40 s. Results were computed using the `classic_sta_lta` utility within ObsPy (Bereuther *et al.* 2010). The poor quality of the data means that time-series methods such as STA/LTA

will either only trigger for the largest events or contain many false detections.

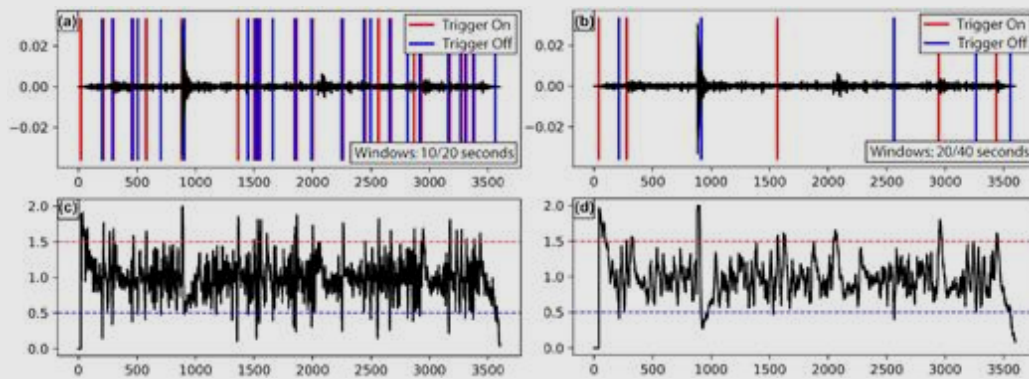


Figure 2. STA/LTA detections (a) and values (c) for 1 hr of LSPE seismic data using a trigger on of 1.5 (red lines), a trigger off of 0.5 (blue lines) and windows of 10 and 20 s. STA/LTA detections (b) and values (d) for the same trace using the same thresholds but windows of 20 and 40 s.

2.2 Convolutional neural networks (CNN)

In its most general form, neural networks are a set of connections uniting inputs and operations (LeCun *et al.* 2015). A deep learning Fully Connected Neural Network (FCNN) can solve problems by transforming a set of input values using a weighted linear transform within an optimization function, and then iterating on the weights by minimizing an error function through a process called stochastic gradient descent (Robbins & Monro 1951; Bottou *et al.* 2018). The purpose of the non-linearity, or activation function, applied to linear transform is to reduce the sensitivity of the output to changes in the weights (Nielsen 2015). A grouping of these linear-nonlinear operations is typically called a hidden layer (Cybenko 1989).

Neural net architecture refers to the number of layers and activation functions (non-linearities) that are used to model the training set. Despite the validity of the Universal Approximation Theorem (Cybenko 1989), it has long been known that a single hidden layer neural network may not be able to learn the desired function in practice, as approximating a complicated function using a single non-linearity may require an unfeasibly large layer that in turn would overfit and not generalize correctly (LeCun 1989; Goodfellow *et al.* 2016).

Accuracy assessment for machine learning problems using varied neural network depth and layer size is an active area of research. Several landmark studies have established general relationships between neural net architecture and the ability to resolve an output function. Sum-product networks, which use computational nodes composed of products or weighted sums of inputs (analogous to traditional artificial neural nets), were used to demonstrate that shallow and deep networks need to grow exponentially and linearly, respectively, to represent families of polynomial functions (Poon & Domingos 2011; Delalleau & Bengio 2011). Montúfar *et al.* (2014) expanded on the work done by Delalleau & Bengio (2011) on sum-product networks for rectifier and maxout networks (Glorot *et al.* 2011; Goodfellow *et al.* 2013) and found the number of resolvable linear regions increases exponentially with increasing numbers of hidden layers.

Although fully connected neural networks can be applied to computer vision problems by setting up each image pixel as an element in an input vector, this becomes computationally intensive for large images (Karpathy 2018). The CNN architecture (LeCun 1989) instead inputs an image as a 3-D array and is able to extract features using a set of convolution kernels which are connected to a subset of pixels or receptive field.

The convolution $S(i, j)$ of a receptive field $F(i, j)$ and a 2-D kernel $K(m, n)$

$$\begin{aligned} S(i, j) &= (F * K)(i, j) = (K * F)(i, j) \\ &= \sum_m \sum_n F(i - m, j - n) K(m, n), \end{aligned}$$

is:

where i and j are the row
(1)

and column indices of the receptive field pixels and m and n are the row and column indices of the kernel (Goodfellow *et al.* 2016). The kernel then ‘slides’ and is convoluted to a receptive field beside the original. This process is repeated until the kernel has been applied to the entire image to create an activation map. In terms of the images used in this study, this process finds the locations in the spectrogram in which there are contrasts of low and high energy.

Eq. (1) describes 2-D matrices, but typically both images and kernels are 3-D matrices with depth reflecting RGB colour values. In practice, a network will learn a kernel that activates across visual features like edges or groups of colours (Krizhevsky *et al.* 2012; Karpathy 2018). In deeper networks trained on photos of people, for example, kernels could reflect more complex features such as eyes or hands (Zeiler & Fergus 2014). The 2-D activation maps produced by each kernel are then stacked where the depth of the layer corresponds to the number of kernels. The number of kernels used in the algorithms are primarily driven by the complexity of the distinguishing features of the classification and the available GPU memory (Krizhevsky *et al.* 2012; Zeiler & Fergus 2014).

The output convolution layer is then passed through a non-linear function and ‘pooled’ prior to stochastic gradient descent. The Rectified Linear Unit or ReLU [$f(x) = \max(0, x)$], is the non-linearity typically applied to image classification problems due to its speed compared to other activation functions (Nair & Hinton 2010; Krizhevsky *et al.* 2012).

Pooling refers to a data-reduction procedure which condenses the information in the convolution layer to a smaller size through

$$L = \frac{L - F_p}{S_p} + 1,$$

averaging or the max() operation (Jarrett *et al.* 2009):

(2)

where F_p and S_p are the sizes of the pooling function receptive field and its stride. The validity of the Universal Approximation Theorem, as described at the beginning of this section for FCNNs, also applies to CNNs, and multiple convolution and non-linearities are needed to resolve complex features (e.g. Krizhevsky *et al.* 2012). As the number of layers in the neural network increases, the spatial size of activation maps become smaller in size but the depth increases.

Stochastic Gradient Descent is the process by which model parameters are updated in the direction of the steepest negative gradient relative to an error function (Robbins & Monro 1951; Bottou *et al.* 2018). The loss function commonly used in image-classification problems is cross-entropy or log loss, which measures the performance of a model where the prediction is a value between 0 and 1 (Bottou *et al.* 2018). For two

classes, this is measured by: $-y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$, where y is (3)

the true label probability and \hat{y} is the probability of the predicted values given by the classifier, which takes the observed data as the input.

2.3 Applications of deep learning to earthquake detection

Over the last few years, deep learning has increasingly been applied to earthquake detection and classification. Ross *et al.* (2018b) used a 4.5 million three-component seismogram training and validation set from Southern California Seismic Network data set to train a CNN capable of detecting *P*- and *S*-wave phases with 99 per cent accuracy. Their final model, which contained four convolution layers (with 21, 15, 11 and 9 filters) and two fully connected layers, used each component of the time-series similar to how a typical image will have three individual RGB channels. This was a natural extension of prior work by Ross *et al.* (2018a), which had used CNNs to detect *P*-wave arrivals and first-motion polarities with an accuracy of 95 per cent (3 convolution layers of 21, 15 and 11 filters and 2 fully connected layers). A comprehensive study testing various deep learning architectures found that CNNs were more accurate than fully connected or recurrent neural networks in determining earthquake first-arrivals (Meier *et al.* 2019).

3 METHODS

Our workflow was typical of CNN problems with an added transfer learning application: (i) publicly available seismic data from a single station were downloaded for a variety of earthquake magnitudes; (ii) 20 images of 20-s windows were taken in both noise and earthquake segments of each seismic trace, converted to spectrograms and used to populate a binary-classification training set; (iii) a classification model was built using two-layer to five-layer CNN architecture; (iv) each of the models were applied to data from the PSE and compared with known

hand-picked events and (v) the best-fitting model was applied to LSPE data and compared with a prior study using Hidden Markov Models.

3.1 Training sets

Creating a labelled training set is the first step in building a deep learning classifier. The training set for this manuscript was earthquake and noise data from a single Earth seismic station. We used the *IRIS* utility *PyWeed* to retrieve data from seismic station Piñon Flats Observatory in California (PFO, 33.612°N –116.459°E) between 2005 and 2019 for all earthquakes magnitude 3 and above located within one degree of the site (approximately 110 km). We avoided small events for this training set in order to train a clear distinction between noise and event segments of a spectrogram. As will be discussed later, a spectrogram colour maximum needed to be fixed for all events to enable a clear distinction between pre-event and event energy. Smaller events would require a very low overall spectrogram colour maximum, which would have the by-product of increasing the colour intensity of the noise. Only data from the vertical component was used for this training set as it has the clearest *P*-wave arrival.

The site was chosen due to its location between the San Andreas and San Jacinto fault zones and its continuous operation. The set of diverse events had varied spectrogram characteristics which promoted algorithm generalization. We used 100 samples-per-second traces starting from 100 s before and 20 s after the *P*-wave arrival to avoid any possible overlaps due to incorrectly marked *P*-wave arrival times. Each trace was then visually inspected to check for earthquake doublets or incorrect *P*-wave arrival times. A total of 716 events were used: 582 in the training set and 134 in the validation set.

Fig. 3 displays the time-series (a, b, c) and spectrograms (d, e, f) of seismic events on Earth and the Moon. Subplots (b) and (e) correspond to an impact moonquake recorded during the PSE at Apollo station 12 and subplots (c) and (f) are from an LSPE thermal moonquake at Apollo station 17. The background noise level on the Moon appears high due to low signal amplitude and the poor dynamic range of the Apollo seismic instruments in comparison to the modern seismic instruments which recorded the training set. As a result, a classifier built on Earth data

which has a much higher signal-to-noise ratio would produce a large number of erroneous detections when applied to lunar noise. Spectrograms also establish a fixed image dimension in which to view data, unlike time-series which require *a priori* knowledge of the maximum amplitude of the quake trace. In other words, if the continuous data is visualized as a series of time-series images, the number of pixels within the plot would vary depending on whether the seismic signal was small or large compared to the fixed reference. Small seismic signals containing fewer pixels would not only require more data to be generalized by the algorithms compared to large signals, but also may be incorrectly classified as noise if the image window is large. This makes spectral data optimal for future real-time analysis on a lander, as this scaling factor does not need to be accounted for in the methodology. Therefore, we decided to build the classifier using spectrograms instead of time-series. Spectrograms were computed using the SciPy (Virtanen *et al.* 2020) function `signal.spectrogram` and the colour scheme was limited to 250 000 counts² Hz⁻¹ so that the colour intensity of the noise field would not vary depending on the size of the earthquake. This procedure is the equivalent to clipping data outside of plot bounds in a time-series image.

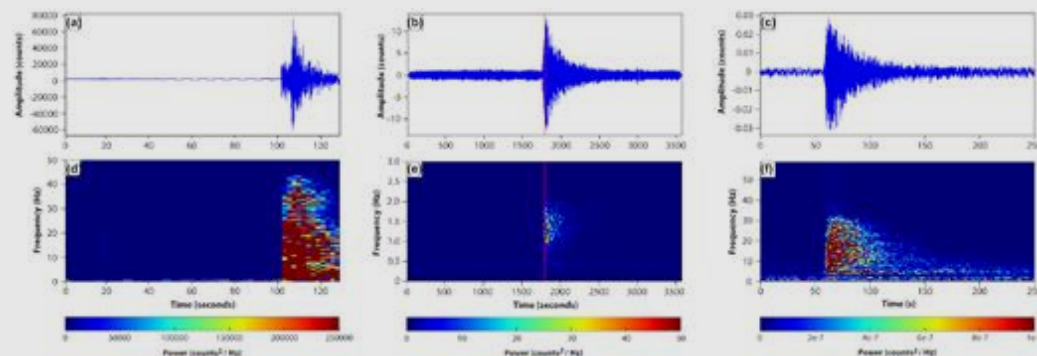


Figure 3. Seismic traces of (a) a magnitude 3.7 earthquake recorded at the Piñon Flats Observatory station in California, (b) an impact moonquake from PSE Station 12 with the event arrival time from Nakamura *et al.* (1981) marked as a vertical line and (c) a thermal moonquake recorded at Geophone 1 during the LSPE. Spectrograms of each event are displayed in (d), (e) and (f).

The Piñon Flats Observatory contains both a broadband and a short period instrument. We used spectrograms from the broad-band instrument so we could record a mix of long-period and short-period

energy to better generalize the spectrogram image. As will be detailed in the discussion, we found that the broad-band data was generalized enough to be used for both the long-period PSE and short-period LSPE instruments. We assessed using events from both instruments within the training set, but decided against it as the changes in the maximum colour scheme of the spectrograms would lead to two distinct representation of images.

Data augmentation is a technique in computer vision that can be used to improve the performance of models by creating new images in the training set which are slightly modified versions of the original images. In computer vision problems using photos, this is typically done by cropping, zooming or rotating images. However, applying data augmentation in this manner will decrease model accuracy for our data set, as cropped or rotated spectrograms will lose their identifying characteristics. Instead, we create additional spectrograms by sliding the data window across the noise and earthquake segments with a small overlap. In addition to increasing the amount of data in the training set, this teaches the model how to recognize an earthquake frame where only a small amount of the window is the earthquake signal, which is crucial for on-time earthquake detection. We chose 20 sliding windows across the noise and earthquake segments with 1 s overlap starting at 0 s for the noise and at 81 s for the earthquake (19 s prior to the onset of the *P* wave at 101 s). The total number of images in the training and validation sets after data augmentation was 22 560 and 5240, respectively. We chose to use Earth data to initially validate the models instead of immediately using moonquakes due to the complexity of the moonquake data, which would require the creation of an additional hand-picked data set as the detections listed in the catalogue are general signal reference times which are not accurate enough for training (Nakamura *et al.* 1981).

3.2 CNN architecture

CNN architecture is typically determined empirically. In other words, most studies try several types of architectures and present whichever performs the best on the validation set (e.g. Ross *et al.* 2018b). In our case, we will be applying the model which performs best on the PSE to

the LSPE instead of simply choosing the one with the highest validation accuracy on the Earth-seismic data set. Although increasing layers and parameters enables the modelling of more complicated functions and features, it can lead to models sampling noise in the training data, also known as overfitting (Srivastava *et al.* 2014). As we are trying to design a lightweight and flexible classifier that could be included on a lander, we tested a simple two-layer model with one dropout layer (with a value of 50 per cent), one fully connected layer, and incrementally added convolution layers. Dropout is a feature where neurons have a probability of being deactivated (in our case 50 per cent), and is commonly used to reduce overfitting (Hinton *et al.* 2012).

A single desktop machine using 25 CPU cores and a batch size of 32 for a NVIDIA P5000 GPU was used for this study. As typical for machine learning supervised methodologies, most of the processing power is used in the training phase. We will discuss the efficiency of the algorithms for real-time applications on the lander in the discussion. The models were built using the *fastai* (Howard *et al.* 2018) Python library and run through the data 50 times (also known as epochs). Loss with learning rate was analyzed using the *fastai lr_find* routine (Fig. S1) and used to choose a learning rate of $1e-4$. 50 epochs were deemed sufficient as training did not further improve accuracy and the validation loss did not greatly surpass the training loss for any of the models (Fig. S2). Smaller learning rates closer to the minimum loss were not chosen due to the possibility of overfitting the training set. The input spectrograms (224×224 pixels) have RGB values and thus a depth of three. The first convolution layer contains 16 filters and their number is doubled for each subsequent layer. We chose a receptive field of size 3, a zero padding of two, and a stride of one for each convolution layer and a receptive field size of 2 and a stride of two for the pooling layer (eq. 2). After the convolution layers, dropout is applied and a fully connected layer is used to sort the outputs into probabilities of the two classes. A schematic for the architecture of the two-layer model is displayed in Fig. 4.

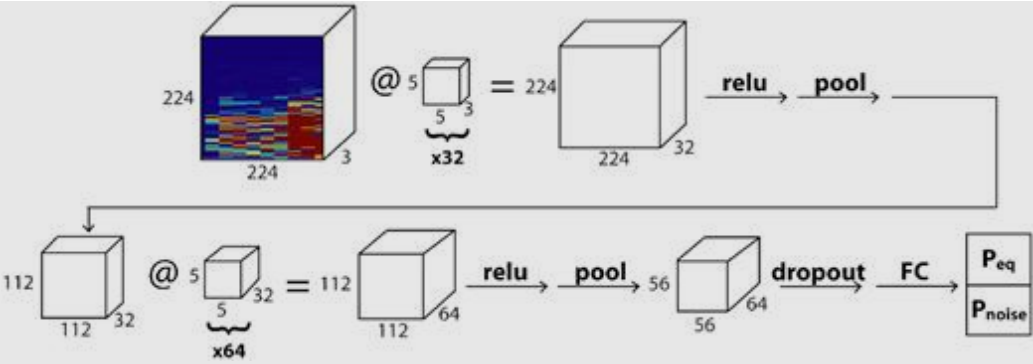


Figure 4. Architecture for the 2-layer CNN used in this study. An input spectrogram of size 224×224 with depth 3 (RGB) is convoluted with 32 kernels of dimensions 5×5×3 then has relu and pooling operations applied. The resulting activation maps are then passed through this process again, dropout is applied and a fully connected layer (FC) is used to sort the volume into a two-element vector of quake or noise probabilities (P_{eq} and P_{noise}).

We found that the three-layer, four-layer and five-layer models all had accuracy of over 99.9 per cent on the validation set (459 778; 1 179 394 and 4 407 554 trainable parameters). We chose to name our ensemble of models ‘MoonNet’. An additional layer of dropout after the second convolution layer was added for the four- and five-layer models to reduce overfitting. The two-layer model had slightly fewer parameters than the three-layer model (455 298) and decreased accuracy (99.0 per cent). The layers of each architecture are displayed in Tables S1–S4 in the [supplementary materials](#) and the results are summarized in Table 1. It is important to remember that the validation set is composed of spectrogram images of earthquakes not in the training set, not moonquakes. However, model validation using earthquakes is still a necessary metric to assess the spectrogram generalization. Model validation using moonquakes will be examined using the PSE catalogue in Section 3.3.

Table 1.

Training loss, validation loss, accuracy and number of parameters for each model.

Model	Training loss	Validation loss	Accuracy	Parameters
MoonNet2L	0.012183	0.038161	99.0 per cent	455 298

MoonNet3L	0.000003	0.000002	>99.9 per cent	459 778
MoonNet4L	0.001005	0.000166	>99.9 per cent	1 179 394
MoonNet5L	0.001810	0.001221	>99.9 per cent	4 407 554

3.3 Computing detections from CNN

Once the classifier is built, it can be used to assess new data as long as they are in the format of the training and validation sets (i.e. a 20-s window $224\times224\times3$ spectrogram). The most current PSE moonquake catalogue contains over 13 000 events (Nakamura *et al.* 1981). The events are graded based on their signal-to-noise ratio, initial impulse coherence, the shape of the envelope function, and the appearance of the event on one or more of the long-period channels: 263 are considered Grade-A, 1018 Grade-B, 7123 Grade-C and the remaining 4653 are considered poor events (Bulow *et al.* 2005). In the current evaluation of Nakamura *et al.* (1981), the grade of the events is not based on an aggregate of all of the instruments. In other words, an event could be classified as Grade-A due to a high-amplitude response in one instrument, but might also contain a much smaller response on a different instrument that would otherwise be considered a much lower grade. Therefore, we chose to use the station which recorded the greatest amplitude.

We downloaded the Apollo PSE miniSEED continuous waveform data from the GEOSCOPE Observatory (Geoscope 1982) through the IPGP datacenter fpt (ftpcdd.ipgp.fr). We extracted hour-long records corresponding to 30 min on either side of each Grade-A event. Unfortunately, data gaps are present in the records so certain events were not available. As some channels had instrument errors and did not display a proper Grade-A quality event, only the channel with the maximum trace value was used. We then visually inspected all remaining one-hour traces to check that the highest-amplitude response selected did not contain an instrument glitch close to the start of the event. Overall, we used 200 out of 263 of the Grade-A events: 38 for Apollo 12, 77 for Apollo 14, 27 for Apollo 15 and 58 for Apollo 16. 18 traces of the 200 events were from the vertical component while the

rest were from the horizontal components. The catalogue of final Grade-A events used is provided in the [supplementary materials](#).

It is unclear what pre-processing was done to the PSE data prior to distribution as it is not documented. The data is mostly centred around zero and some filtering has been done as it is not distributed into even counts. As the data does not contain gaps, it is likely that it was interpolated prior to distribution. However, these details are not crucial as our process to create the spectrogram cleans the data and the classifier will only use the general shape of the spectrogram distribution.

A median despiking routine was applied to remove brief, high-amplitude signals produced by glitches or data dropout (Phillips 2017). The classifier assessed 20-s spectrogram windows across each trace. A few alterations to the data were required to increase the similarity between spectrograms of moonquakes and the earthquakes in the training set. Moonquakes are much weaker in power than earthquakes, so a colour-scheme with a maximum of $50 \text{ counts}^2 \text{ Hz}^{-1}$ was chosen so that the noise-signal contrast is similar to the Earth training set (Fig. 3d). In addition, a 1.0 Hz highpass filter was applied to the LSPE data to remove low-frequency background noise, which is prevalent on the Moon but not present in the Earth spectrograms. This filter was not necessary for the PSE instruments.

The result of the applied classifier is a vector of probabilities describing the likelihood of the two categories at each second of the record. We classify these results as 'detection'. An algorithm was then developed to sort this probability vector into moonquake 'arrivals', that is the time in which the first seismic wave reaches the seismometer, and ignore occasional false detections. The algorithm uses a sliding window of length w seconds where it tracks the number of noise and quake classifications. If it finds a number of quake detections equal to e , then an arrival flag is placed. An input confidence c is the minimum ratio of correct to incorrect predictions required (i.e., threshold of the maximum number of incorrect frames in w). An integer variable d is the number of ignored frames after a first detection used to prevent multiple detections for the same waveform. Several variations of the parameters were tested for this data set. As will be further discussed in

the results section, we decided to use $w = 30$, $e = 30$, $c = 0.7$ and $d = 20$ for both the PSE and LSPE data.

After the accuracy of the models was determined using the PSE data, the algorithms were applied to the four LSPE geophones. The starting LSPE data used for this study were hourly interpolated and unfiltered SAC files produced by Phillips (2017), which accounted for time overlap errors, sync errors, and data gaps. The same median despiker used in the PSE data was applied to each trace. Due to differences in hardware between the two data sets, the highpass filter was changed to 10 Hz and the spectrogram colour maximum was lowered to $1\text{e-}6 \text{ counts}^2 \text{ Hz}^{-1}$ to account for the higher sample-rate. Original LSPE data can be retrieved from Data ARchives and Transmission System (DARTS) database at url <https://darts.isas.jaxa.jp/planet/seismology/apollo/PSE.html> (Muir et al., 2000)

4 RESULTS

4.1 Testing architecture

Table 1 displays the training loss, validation loss, accuracy and number of parameters for each of the 2- to 5-layer models. Each model had an accuracy of 99 per cent or higher with the validation set. The 2- and 3-layer models have a single dropout layer, while the 4- and 5-layer models have a dropout layer after the second convolution layer and before the fully-connected layer (Tables S1–S4). The second dropout layer in the 4- and 5-layer models was necessary to prevent data overfitting. The updated 4-layer model still overfit the data between epochs 15–30 but minimized with further training (Fig. S2C). This is not surprising given that the number of parameters greatly increases by adding the fourth and fifth layers (Table 1). The amount of overfit on these models is very low, and does not automatically disqualify the models from being tested with the PSE data.

One of the main appeals of using a supervised algorithm for lander-side detection is that the primary computational (and therefore energy) costs occur during the training phase on Earth. Quantifying the energy usage during the testing phase greatly depends on the eventual

hardware configuration of the lander, but general estimates can be obtained using utilities. We used the *energy-usage* Python package (Friedler *et al.* 2019) to compute the energy usage for the two main processing steps that would occur on a lander: the spectrogram generation and spectrogram assessment. The spectrogram generation, which the real-time application on the lander would be at a rate of 1 per second, was measured to be 13.97 Watts. The spectrogram comparison takes approximately 20.28 Watts using a batch size of 16. The architecture of the code was not set up for real-time analysis, so it is difficult to estimate the power consumption for assessing a single spectrogram. Determining an ideal hardware setup to maximize power efficiency for a lander is a complex problem that requires thorough experimentation. Although the real-time testing rate (one spectrogram created and tested per second) is slow enough that a GPU is not necessary, it may be more efficient to operate on a buffer of several minutes to an hour to take advantage of CPU or GPU parallelization.

4.2 Assessing PSE arrivals

The 2- to 5-layer architectures were compared against hand-picked signal reference times on Grade-A events of PSE data (Nakamura *et al.* 1981). A sample PSE event for S16 is displayed in Fig. 5. It shows the time-series with the catalogue moonquake arrival (A, red line) and the arrival determined by MoonNet3L (B, green line). The spectrogram of the trace (c) demonstrates that the algorithm is able to place the arrival right at the beginning of the increased energy of the spectrogram. Statistically assessing the performance of the classifier is a non-trivial task for this data set due to differences between each seismometer including: different geographical locations on the lunar surface, coupling differences, and technical glitches. This is further complicated by the time difference between the catalogued signal reference times (Nakamura *et al.* 1981) and actual wave arrival times. Although this time difference is usually within a minute, certain waveform arrivals occur several minutes after the signal reference time (e.g. Fig. S3).

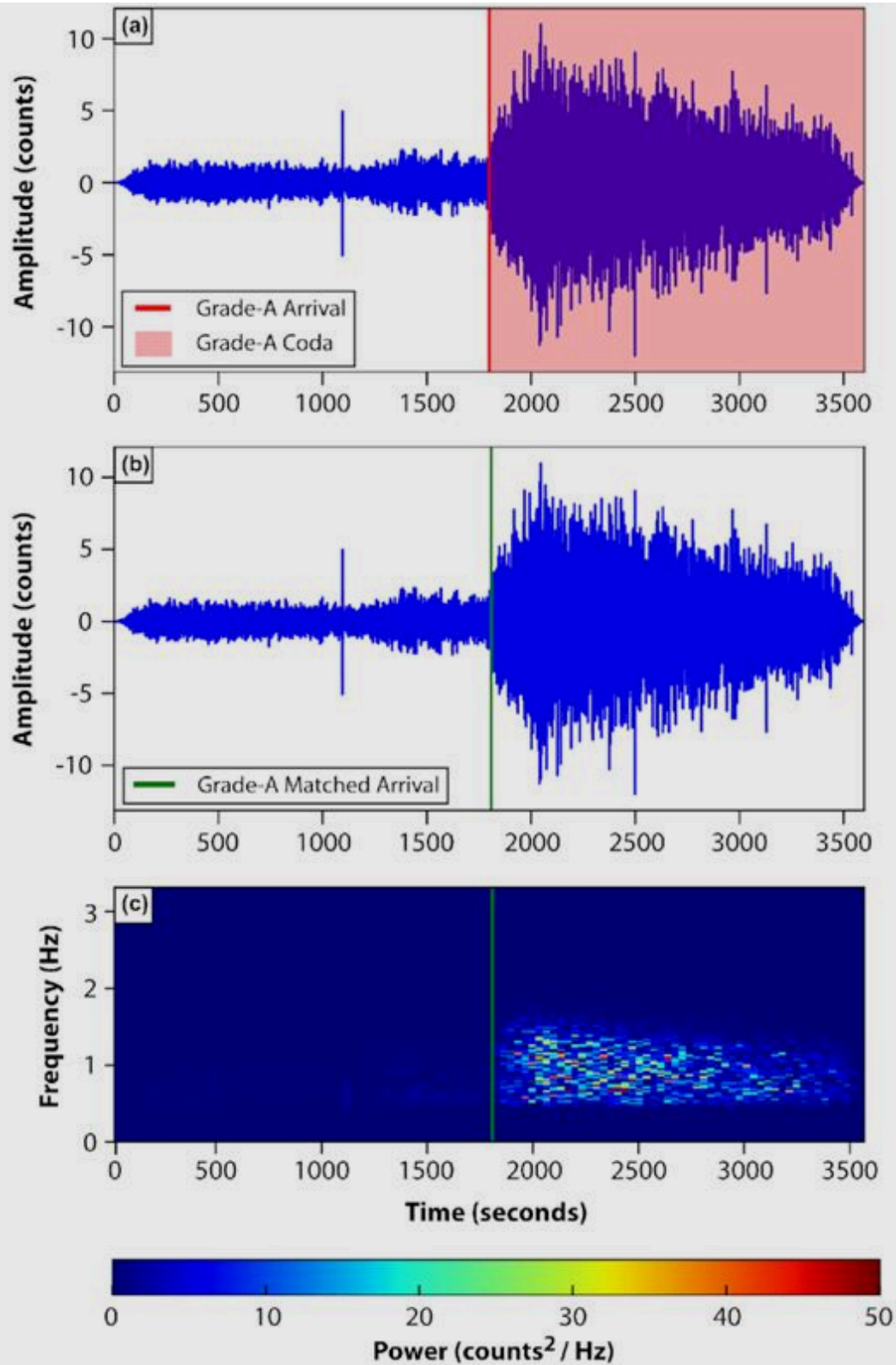


Figure 5. Example of deep learning detection for a sample Grade-A PSE event (S16 channel MH2, impact event on 1975-08-10). (a) Time-series of a Grade-A PSE event with catalogue arrival (line) and coda (shaded). (b) Time-series of the same event with deep learning arrival by MoonNet3L, which was classified to be within the leniency window for the event (line). (c) Spectrogram of the PSE Grade-A event.

It is unclear what signal occurs at approximately 1200, as it is not in the catalogue and it does not look like a typical instrument glitch.

We assessed our results by comparing the deep learning arrivals to the signal reference times the catalogued Grade-A and non-Grade-A events. First, we checked whether or not the Grade-A event was detected within a time window of several minutes. We assessed the validity of the remaining arrivals by checking whether or not they corresponded to arrivals within the Grade-A window, the Grade-A coda, or the window or coda of a non-Grade-A event, and summarized the results using two metrics: (1) a 'Grade-A percentage' and (2) and 'extra-arrival accuracy'.

We define the Grade-A percentage as the number of obtained Grade-A arrivals divided by the total number of Grade-A events. This is a complicated metric to measure in practice because the time difference between the signal reference time in the catalogue and the arrival of the seismic wave in the data is not consistent from event to event. For example, the marked signal reference time in Fig. S3 does not have a noticeable high amplitude in the time-series (Fig. S3a), nor does it have high energy at any frequency band (Fig. S3c) until several minutes later. To resolve this problem, we created a parameter called a 'leniency window', in which arrivals within a certain time window from the marked signal reference time are considered 'on time' and count for the Grade-A percentage accuracy. Fig. S3 used a 300-s leniency window, so the deep learning arrivals (Fig. S3b) were classified as 'Grade-A coda' instead of corresponding to the main arrival.

Additional arrivals within the same seismic event are often unavoidable due to the low signal-to-noise ratio of the lunar data and the long data trace used for comparison (30 minutes before and after the Grade-A event). This is not necessarily a negative for our algorithm as our objective is to isolate data segments for telemetry instead of building a precise lander-side catalogue. Additionally, non-Grade-A events are sometimes present in the data and found by the algorithm. However, these additional arrivals can be cumbersome when assessing the overall accuracy of the models as they need to be distinguished between valid seismic signals and model errors or instrument glitches. This is made more complicated due to a lack of accurate times in the

PSE catalogue, which requires the use of a leniency window for both the Grade-A event and any additional events found within the trace.

Due to this, we developed an ‘extra-arrival accuracy’ metric to assess the validity of extra arrivals within each trace to further quantify the accuracy of each model. We define the extra-arrival accuracy as the sum of the arrivals corresponding within the Grade-A window, Grade-A coda, and window/coda of a non-Grade-A event divided by the total number of extra arrivals. In other words, it is the ratio of arrivals with a known seismic origin and those that are presumed to be errors of the model or instrument. A high extra-arrival accuracy means that additional arrivals are primarily within either the same Grade-A event or other non-Grade-A events marked in the PSE catalogue (model is accurate), while a low extra-arrival accuracy means that many arrivals are placed in the noise of the seismic trace or due to instrument glitches (model is inaccurate).

Fig. 6 displays Grade-A and extra-arrival accuracy for Station 12 for MoonNet2L through 5L and 1–10 min windows. Similar plots for Stations 14, 15 and 16 are displayed in Figs S4–S6 in the [supplementary materials](#). As may be expected, a larger leniency window increases the accuracy of both the Grade-A and the extra-arrival accuracy at the expense of a less meaningful result. The objective is to use a window that is small, but large enough to account for difficult to observe arrivals in data such as those in Fig. S3. We chose a 300-s leniency window (despite the event in Fig. S3 being incorrectly classified) as larger windows typically do not increase accuracy and smaller windows have drastically lower accuracy (Fig. 6). This characteristic is not unique to this station and is also observed in Stations 14, 15 and 16 (Figs S4–S6).

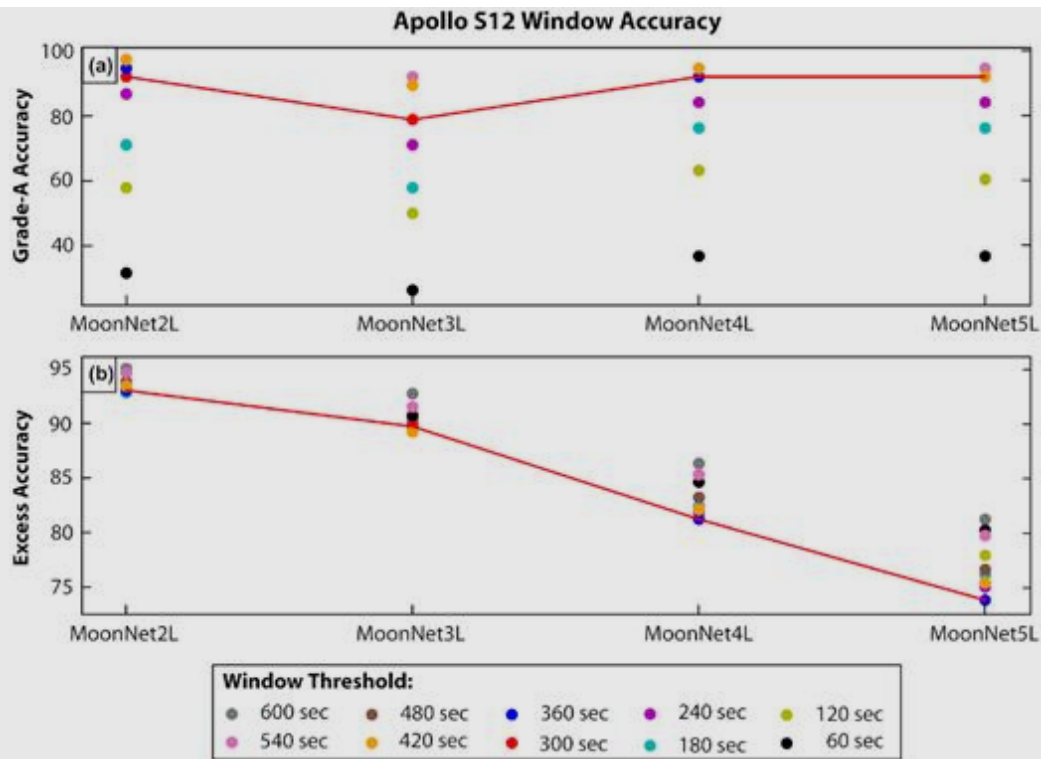


Figure 6. Accuracy of PSE arrivals for S12. (a) Grade-A event and (b) extra-arrival accuracy (whether or not extra arrivals are due to seismic signals or noise) for MoonNet 2L through 5L for 1–10-min leniency windows. The 300-s leniency window is highlighted using a red line in both subplots.

Bar graphs showing the distribution of the 300-s window arrivals for the 2–5 layer models are displayed in Fig. 7, Figs S7–S9 and are summarized in Table 2. In Fig. 7(a), Grade-A arrivals are displayed in bar format as Grade-A matched (blue) or not matched (orange). In subplot (b), the extra arrivals are divided into Grade-A 300-s window (blue), Grade-A coda (orange), 300-s window of a non-Grade-A event (green), non-Grade-A event coda (red) or unassigned (purple). Arrivals marked as unassigned do not match seismic signals in the PSE catalogue and most likely correspond to arrivals incorrectly placed in the noise of the trace.

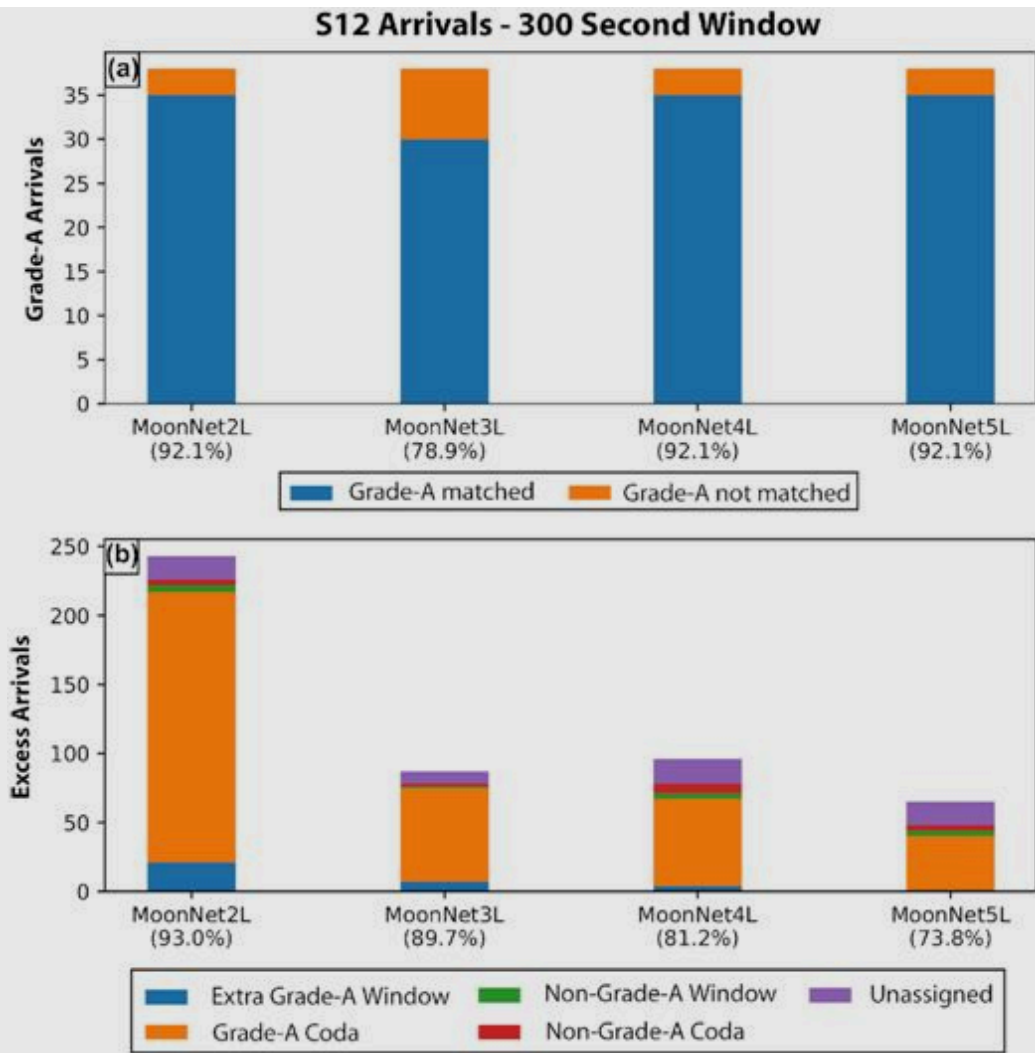


Figure 7. Bar graphs showing the distribution of Grade-A and extra-arrival accuracy for PSE Station 12. (a) Grade-A arrivals are classified as Grade-A matched (blue) or not matched (orange). (b) Extra arrivals are classified as arrivals in the Grade-A 300-s window (blue), Grade-A coda (orange), 300-s window of a non-Grade-A event (green), non-Grade-A event coda (red) or none of the above (purple).

Table 2.

Grade-A and extra-arrival accuracy (marked as -A and -Ex. respectively) for each model using a 300-s leniency window. The column Avg. represents the average accuracy for each station while the rows Avg.-A and Avg.-Ex display the average accuracy values for each model. The values in the extra-arrival accuracy correspond to extra Grade-A window + Grade-A coda + non-Grade-A window + non-Grade-A coda / total additional arrivals.

MoonNet2L	MoonNet3L	MoonNet4L	MoonNet5L	Avg.
-----------	-----------	-----------	-----------	------

S12-A	35 /38 (92.1 per cent)	30/38 (78.9 per cent)	35/38 (92.1 per cent)	35/38 (92.1 per cent)	89 per cent
S12-Ex.	21+196+5+4 /243 (93.0 per cent)	7+68+1+2 /87 (89.7 per cent)	4+63+4+7 /96 (81.2 per cent)	1+39+4+4 /65 (73.8 per cent)	88 per cent
S14-A	76/77 (98.7 per cent)	76/77 (98.7 per cent)	76/77 (98.7 per cent)	76/77 (98.7 per cent)	99 per cent
S14-Ex.	69+300+18+15 /551 (73.0 per cent)	41+114+7+2 /271 (60.5 per cent)	51+117+11+8 /384 (48.7 per cent)	36+15+10+6 /243 (27.6 per cent)	57 per cent
S15-A	27/27 (100.0 per cent)	25/27 (92.6 per cent)	27/27 (100.0 per cent)	27/27 (100.0 per cent)	98 per cent
S15-Ex.	11+96+2+0 /132 (82.6 per cent)	6+56+1+0 /74 (85.1 per cent)	7+26+1+0 /54 (63.0 per cent)	5+20+1+0 /43 (60.5 per cent)	77 per cent
S16-A	56/58 (96.6 per cent)	53/58 (91.4 per cent)	56/58 (96.6 per cent)	57/58 (98.3 per cent)	96 per cent
S16-Ex.	37+222+3+5 /333 (80.2 per cent)	8+53+0+3 /91 (70.3 per cent)	27+155+4+4 /262 (72.5 per cent)	14+25+4+4 /116 (40.5 per cent)	71 per cent
Avg.-A	97 per cent	92 per cent	97 per cent	98 per cent	
Avg.-Ex.	80 per cent	71 per cent	61 per cent	40 per cent	

The deep learning models were able to match the Grade-A events at 78.9–92.1 per cent for S12, 98.7 per cent for S14, 92.6–100 per cent for S15 and 91.4–98.3 per cent for S16 (Table 2). The accuracy of the additional arrivals depended greatly on the model and station used, with MoonNet5L and S14 having the worst average accuracy at 40 and 57 per cent, respectively. We found that MoonNet2L determined approximately twice or more arrivals as the other models, which were primarily located in the coda of the Grade-A event (Fig. 7).

4.3 Application to the LSPE

Each of deep learning models was able to obtain an average accuracy greater than 90 per cent on the grade-A PSE subset (Table 2). Due to this, we used the extra-arrival accuracy to determine which model to apply to the LSPE data. The model with the highest extra-arrival accuracy was MoonNet2L (80 per cent), but this result is in part due to the large number of arrivals. Therefore, we chose to apply the model with the second highest average extra-arrival accuracy, MoonNet3L (71 per cent) to the LSPE data set.

The four geophones were operational for approximately 256 d (Phillips 2017). Hourly segments of seismic data were run on the desktop machine described in Section 3.2. One segment could be assessed by the model in about a minute, so the entire data set for one geophone takes approximately 4–5 d of processing time. The most time-consuming segment of processing is the CPU-driven creation of spectrograms (approximately 50 s), which can be easily improved by using additional CPU cores. An example result of the MoonNet3L model applied to an hour trace of LSPE is displayed in Fig 8 alongside the Hidden Markov Models detections from Dimech *et al.* (2017). We can observe through visual inspection that the results from the MoonNet3L model produce more accurate arrivals than the Dimech *et al.* (2017) with fewer false or duplicate detections.

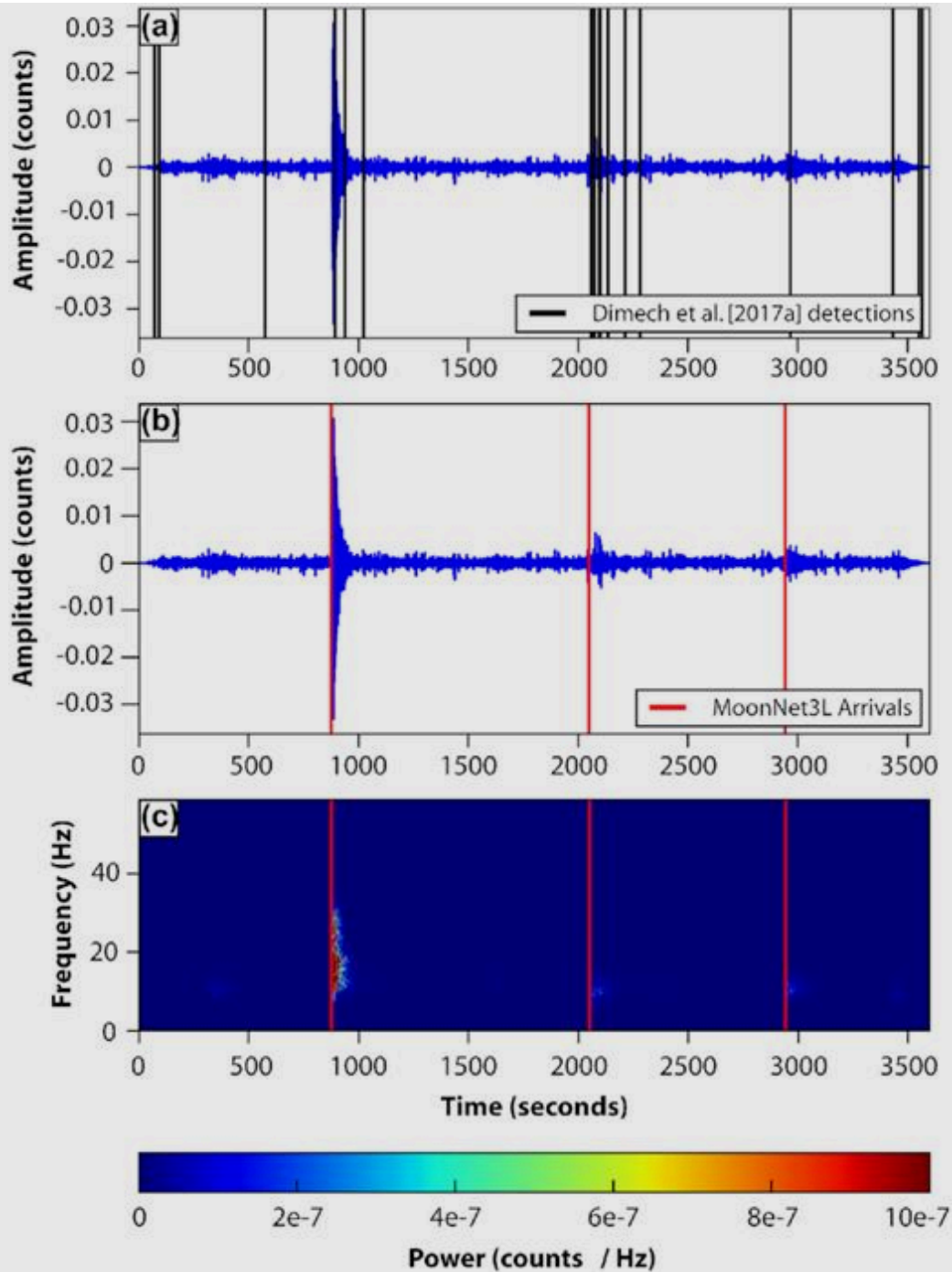


Figure 8. LSPE comparison between MoonNet3L and the Dimech *et al.* (2017) study for an hour trace of Geophone 1 on 1976-08-17. Time-series with detections from Dimech *et al.* (2017) (black lines) and MoonNet3L arrivals are displayed in (a) and (b), respectively. [C] Spectrogram with the MoonNet3L arrivals.

The arrivals in each hour segment were then summed for each day (Fig. 9). We found that the daily moonquakes detected by MoonNet3L number between a few to over 80 and vary over time with a periodicity of approximately a month, consistent with the day/night cycle that drives the generation of thermal moonquakes (Fig. 9, Duennebier &

Sutton 1974). The periodicity is visually similar to that of the Dimech *et al.* (2017) study, but as can be expected from the results shown in Fig. 8, the latter almost has an order of magnitude greater number of detections (Fig. 9b). The full catalogue for each of the geophones is provided as a CSV file in the [supplementary materials](#). To our knowledge, this is the first catalogue of lunar seismicity obtained with machine learning techniques.

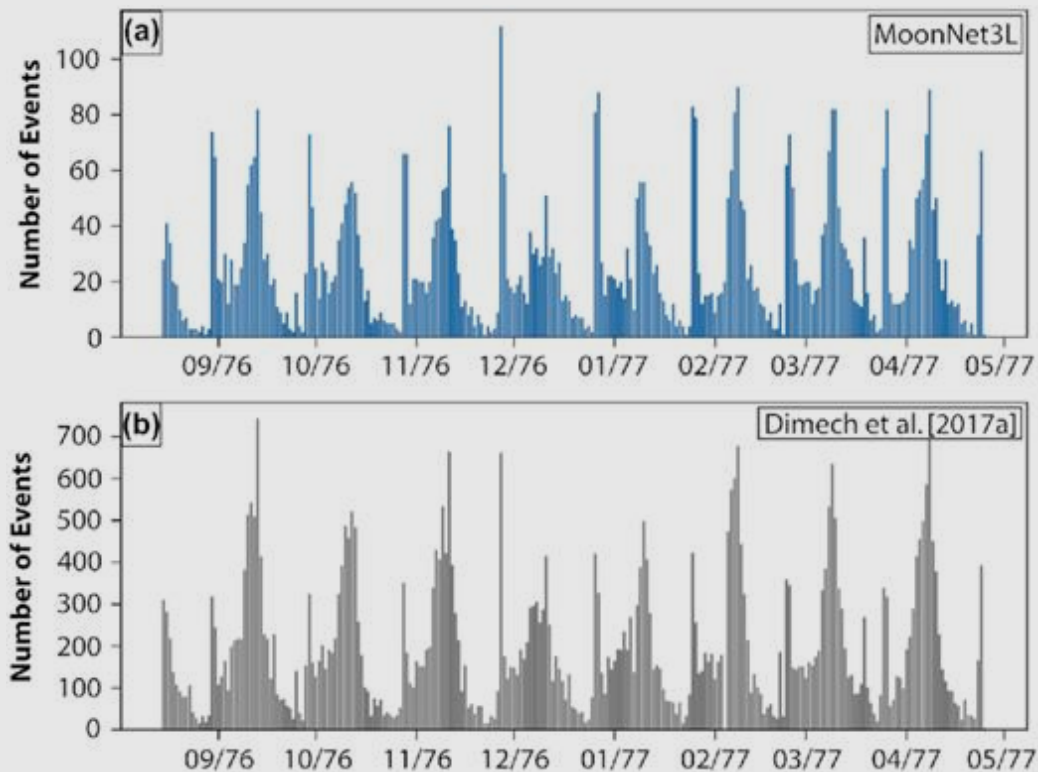


Figure 9. Cumulative daily moonquake arrivals using (a) the MoonNet3L model for LSPE Geophone 1 and (b) the detections from the Dimech *et al.* (2017) study. Time axis is in month/year format.

5 DISCUSSION

5.1 PSE assessment

We have demonstrated that computer-vision deep learning classifiers can be used for planetary seismicity detection despite not being trained on a local data set. We were able to obtain an average classification accuracy between 89 and 99 per cent on a 200 element subset of the Apollo PSE Grade-A events using two- to five-layer convolution neural

networks and a 300-s arrival leniency window (Table 2 and Tables S1–S4). The extra-arrival accuracy for the models ranged between 57 and 88 per cent. This accuracy was determined by assessing which arrivals corresponded to extras in the leniency window, the Grade-A event coda, or the window and coda of non-Grade-A events (Fig. 7b). It is likely that the Grade-A accuracy is higher than the values listed in Table 2 due to overgeneralized signal reference times in the PSE catalogue (Nakamura *et al.* 1981), which result in the deep learning arrival being classified as a coda arrival instead of falling within the leniency window of the event itself (e.g. Fig. S3). This is especially present in MoonNet3L, which has a lower Grade-A accuracy than the other models (78.9 per cent) but a high amount of Grade-A coda arrivals (68) and few unassigned arrivals (9) (Fig. 7). This high accuracy was obtained using a training set from a single Earth seismic station composed of events greater than magnitude 3 with varied spectrogram characteristics but was not located in a Moon-analogue geological area. In order to increase the number of training examples, data was augmented by taking 20 spectrograms across the noise and earthquakes segments for each record.

We found that a highpass filter was necessary to prevent the deep learning models from assessing the low-frequency lunar noise as seismic events. Models with a highpass filter lower than 1 Hz were tested, but provided poorer accuracy for most models and stations. This was particularly true for S14, which tends to have worse signal-to-noise ratio compared to the others. The ‘unassigned’ arrivals (Fig. 7, purple bars) are primarily due to instrument glitches, which are reflected in the spectrograms as short bursts of energy across all frequencies (Fig. S10).

5.2 LSPE comparisons

The MoonNet3L model was able to obtain detections in the LSPE data to greater accuracy than a recent study using Hidden Markov Models with an order of magnitude fewer false detections (Fig. 8). This result is especially promising for our methodology because the Hidden Markov Models algorithm was trained on lunar data (Dimech *et al.* 2017). The displayed trace is the same hour used for the STA/LTA method in Fig. 2. It is observable that the MoonNet3L model produced more accurate

arrival results than either STA/LTA parameters tested. Furthermore, the quality of the results show that the algorithms are flexible enough to analyse diverse instrument types, as both the training set and the PSE instruments are long-period but the LSPE used short-period Geophones (Figs 3 d, e, f). As hand-picked catalogues of the LSPE thermal moonquakes do not exist, it is difficult to quantify the extent of how the models perform compared to other studies.

As with the PSE data, the LSPE data set had occasional glitches which registered as false detections. Most interestingly, the periodicity and relative number of events in time is visually similar between MoonNet3L and the Dimech *et al.* (2017) model, suggesting that the same mechanism is being observed with consistently fewer additional detections in the deep learning methodology. Similar to that study, we hypothesize that the moonquakes are a thermally driven response to the lunar day/night cycle.

The purpose of this work is to create a robust seismic detection algorithm that is lightweight enough to be incorporated on a lander to prioritize telemetry data, not classification as is typically done with Hidden Markov Models (e.g. Knapmeyer-Endrun & Hammer 2015; Dimech *et al.* 2017). However, this algorithm can be altered to fit such a purpose if a synthetic or local training set is used whose event categories are distinguishable through spectrograms.

5.3 Practical considerations

Deep learning models can be developed with a philosophy based on generality (the model can be applied to numerous data sets at the expense of accuracy) or specificity (the model has high accuracy but is only suitable for particular data sets or instruments). The models in this manuscript were developed to be as general as possible yet tuned to minimize false triggers. This tuning comes at a cost, as certain small or rare events might be missed. This is a trade-off that cannot be avoided, only mitigated on a case-by-case basis.

Seismic detections are computed using the cross-entropy loss function, which is used to measure the differences between an input spectrogram image with that of the generalized classification. Therefore, features

such as maximum time-domain amplitude can be poor indicators of whether or not the model will trigger. In order to quantify whether or not only large events would be detected, we developed an extra-arrival accuracy metric and found that non-Grade A events were still detected.

We decided to expand this assessment to 200 randomly selected Grade-B events to test how the algorithm performed for data with worse signal-to-noise ratio (Table S5 and Figs S11–S14). The subset of Grade-B events used are listed in the [supplementary materials](#). Unlike with the Grade-A events, we did not visually inspect and exclude certain events to simulate a potential ‘worst case scenario’ where glitches are prominent. We found that the average station accuracy for Grade-B events ranged from 59 to 84 per cent (Table S5), a drop down from the 89 to 98 per cent for Grade-A events (Table 2). The extra-arrival accuracy for Grade-B events was between 63 and 81 per cent.

It is important to note that decreased accuracy for Grade-B events does not mean that the algorithm is tuned for large events. After all, the algorithm was trained on Earth seismograms, which are many orders of magnitude larger in amplitude and have higher signal-to-noise ratios compared to moonquakes. The accuracy is directly based on the comparisons of the spectrogram segments within the lunar record to the generalized spectrogram images of the noise class (little to no energy in the spectrum) and seismic class (energy at various parts of the spectrum). We hypothesize that results could be improved by using a dynamic colourbar for the spectrogram plots that scaled depending on the amplitude of the event in the time domain. This would not be possible to do on a lander in real-time, but could be carried out for segments of data stored in a buffer in conjunction with other time-domain analysis methods. A simpler solution at cost of increased processing power could be to create several spectrograms with discrete colour-bar maximums corresponding to weak, intermediate and strong events. Unlike the dynamic scale, this technique could be run in real-time on a lander. Either solution would still only require a single machine learning model.

Tuning the models to obtain a more accurate arrival time might increase false detections and hinder the main objective of lander-side detection for telemetry. For this purpose, we are less interested in single

accurate arrival times of seismicity and are not concerned with additional arrivals within the coda of an event. This may be an area to further expand this code in the future, as knowing the full extent of the seismic coda could provide an end boundary for the data window to be sent for telemetry. The simple solution of this problem is to send the data for telemetry starting a few minutes prior to the first detection. Finally, it is important to note that tuning the model for a mission where data has not yet been retrieved is merely an approximation and would greatly benefit from iteration after deployment. It's possible that a large segment of data could initially could be sent back from a lander upon arrival and used to adapt and retrain models on Earth, after which a new trained model could be uploaded to the lander.

5.4 Effect to mission design

Planetary seismology missions require striking a delicate balance between obtainable results, instrument cost and weight and computation and power requirements of both lander operation and telemetry. In other words, planetary studies try to achieve the maximum research possible with the most limited amount of resources. Due to the cost and practicality of planetary research some compromises in mission design need to be made, especially since research teams of multiple disciplines will strive to have their own work take priority.

For seismology, this trade-off is primarily made with instrument design, in particular by the sensitivity and sample-rate of the seismometer. As previously mentioned in the manuscript, it may not be possible to have both high-resolution and continuous data for the entire mission duration due to power costs. However, continuous high-resolution data may not be needed for all studies. For example, studies such as Larose *et al.* (2005), which used continuous records of Apollo 17 ambient seismic noise to determine a velocity model of the near-surface, typically use data at lower sample-rate than earthquake tomography (e.g. Bensen *et al.* 2007). Therefore an instrument could be run continuously in a low sample-rate mode and in triggered mode for detected seismic events. Furthermore, this algorithm expands the possibility for future missions to record very high resolution data in

short bursts, something that is rarely considered in mission design due to the difficulty of lander-side analysis.

The inclusion of a lower sample-rate channel should help mitigate confirmation bias due to expected events. In other words, unexpected signals which were not triggered by the algorithm would hopefully be observable in the low sample-rate data. If this is the case, an additional classification algorithm could be coded, trained on Earth, and uploaded to the instrument to catch these unexpected signals in the highest possible resolution. It may be useful to run the data in high-sample mode in repeated intervals at the start of the mission to better understand the noise profile and characterize unexpected signals or glitches. The proposed algorithm could be part of a suite of seismic detection techniques for the lander to increase the likelihood of signal discovery and retrieval. Ultimately, the trade-offs of this methodology will need to be evaluated on a case-by-case basis for each mission according to the primary science goals.

Finally, algorithms for future missions can be improved by using synthetic data to increase flexibility of training sets. We tested the effect of adding synthetic noise to the Earth seismic data used in the training set and built additional two- to five-layer MoonNet models. We added Gaussian noise for each training example containing a variance equal to 5 per cent the maximum amplitude of the event (Fig. S15). Results for the models with added noise are listed Table S6 and in Figs S16–S19.

The noise-added MoonNet2L had approximately the same Grade-A accuracy as the original (97 per cent) but a worse extra-arrival accuracy (72 per cent versus 80 per cent). The noise-added MoonNet3L model had a 4 per cent improvement (96 per cent versus 92 per cent) at the cost of a 2 per cent decrease in extra-Arrival accuracy (69 per cent versus 71 per cent). Adding noise to MoonNet4L decreased the accuracy by 6 per cent compared to the original model (91 per cent versus 97 per cent) but had a substantial increase in the extra-arrival accuracy (77 per cent versus 61 per cent). The noise-added MoonNet5L had the worst performance of any model, with a Grade-A detection accuracy of only 4 per cent. The reason that the accuracy of the noisy MoonNet5L model is so low is because the detections for each spectrogram (done

across the record every second) are very inconsistent in their classification, and do not pass the 70 per cent input confidence (parameter c in Section 3.3) in the detection-to-arrival algorithm. We hypothesize that the bad model performance is due to poor generalization caused by overfitting the Earth-based training set. To facilitate the capture of differently sized events it may be better to change the input spectrogram as discussed in the adaptive and discrete maximum spectrogram amplitude strategy in Section 5.3 instead of adding synthetic noise to training data.

6 CONCLUSION

A CNN model was built using solely a training set of noise and earthquakes from a single Earth seismic station and used to detect seismic events within the Apollo PSE and LSPE data sets (Fig. 3). Two- to five-layer convolution models were created using a data set of 20, 20-s windows for both noise and seismic activity from 716 earthquake records and obtained over 99 per cent accuracy in their validation set (Table 1). An algorithm was developed to sort detections from each model into arrivals which removed false picks and extra consecutive detections. Each model was then tested against a 200 element subset of the Grade-A events within the Apollo PSE data and obtained between 89 and 96 per cent accuracy across all stations (Table 2, Fig. 5). An ‘extra-arrival accuracy’ metric was developed to more robustly describe the accuracy of excess arrivals in the PSE data and used to select the three-layer model as the best-fitting model.

The three-layer model was applied to the LSPE data and found to obtain more accurate results and fewer false or multiple detections compared to a recent study using Hidden Markov Models (Figs 8). The cumulative number of daily moonquakes show a monthly periodicity that we interpret as the lunar day and night cycle, which is the mechanism driving the generation of thermal moonquakes, the primary phenomena observed in the LSPE data (Fig. 9). We have demonstrated that accurate seismic detections of planetary seismicity can be made using deep learning generalized algorithms despite a lack of local training data. Although we tested these algorithms solely using seismic

data, we anticipate that they can easily be adapted to other types of planetary data sets. Future work will also assess the detection capability of these types of algorithms to marsquakes recorded by the InSight mission (Clinton *et al.* 2018; Giardini *et al.* 2020) and proposed or anticipated missions such as the Lunar Geophysical Network (Neal *et al.* 2020) and Dragonfly (Lorenz *et al.* 2018).

SUPPORTING INFORMATION

- Table S1.** Architecture for the tested 2-layer model. It contains a total of 455 298 trainable parameters.
- Table S2.** Architecture for the tested 3-layer model. It contains a total of 459 778 trainable parameters.
- Table S3.** Architecture for the tested 4-layer model. It contains a total of 1 179 394 trainable parameters.
- Table S4.** Architecture for the tested 5-layer model. It contains a total of 4 407 554 trainable parameters.
- Table S5.** Grade-B and extra-arrival accuracy (marked as -B and -Ex, respectively) for each model using a 300-s leniency window. The column Avg. represents the average accuracy for each station while the rows Avg.-B and Avg.-Ex display the average accuracy values for each model. The values in the extra-arrival accuracy correspond to extra Grade-B window + Grade-B coda + non-Grade-B window + non-Grade-B coda / total additional arrivals.
- Table S6.** Grade-A and extra-arrival accuracy (marked as -A and -Ex, respectively) for each model built with noise using a 300-s-leniency window. The column Avg. represents the average accuracy for each station while the rows Avg.-A and Avg.-Ex display the average accuracy values for each model. The values in the extra-arrival accuracy correspond to extra Grade-A window + Grade-A coda + non-Grade-A window + non-Grade-A coda / total additional arrivals.
- Figure S1.** Loss with learning rate for the 2- to 5-layer models analysed in this study.

Figure S2. Training loss, validation loss and accuracy for 2- to 5-layer models.

Figure S3. PSE Grade-A event (S12 channel MH2, impact event on 1970–11–12) with poor catalogue arrival time. (a) Time-series of a Grade-A PSE event with catalogue arrival (red line), leniency window (magenta line) and coda (shaded). (b) Time-series of the same event with three arrivals from the MoonNet3L model, which were classified to be within the coda of the event (orange line). (c) Spectrogram of the PSE Grade-A event.

Figure S4. Accuracy of PSE arrivals for S14. (a) Grade-A event and (b) extra-arrival accuracy for MoonNet 2L through 5L for 1–10-min leniency windows. The 300-s leniency window is highlighted using a red line in both subplots.

Figure S5. Accuracy of PSE arrivals for S15. Subplot information same as Fig. S4.

Figure S6. Accuracy of PSE arrivals for S16. Subplot information same as Fig. S4.

Figure S7. Bar graphs showing the distribution of Grade-A and extra-arrival accuracy for PSE Station 14. (a) Grade-A arrivals distributed into Grade-A matched (blue) or not matched (orange). (b) Extra arrivals distributed into arrivals in the Grade-A 300-s window (blue), Grade-A coda (orange), 300-s window of a non-Grade-A event (green), non-Grade-A event coda (red) or none of the above (purple).

Figure S8. Bar graphs showing the distribution of Grade-A (a) and extra arrivals (b) for PSE Station 15. Subplot information same as Fig. 7.

Figure S9. Bar graphs showing the distribution of Grade-A (a) and extra arrivals (b) for PSE Station 16. Subplot information same as Fig. 7.

Figure S10. PSE Grade-A event (S14 channel MHE, impact event on 1971–10–20) with instrument glitch. (a) Time-series of Grade-A PSE event with catalogue arrival (red line) and coda (shaded). (b) Time-series of the same event with the matched A arrival time (green line) from the MoonNet3L model and an unassigned arrival (purple line) due to instrument glitch. (c) Spectrogram of the PSE Grade-A event.

Figure S11. Bar graphs showing the distribution of Grade-B (a) and excess arrivals (b) for PSE Station 12. Subplot information same as Fig. 7.

Figure S12. Bar graphs showing the distribution of Grade-B (a) and excess arrivals (b) for PSE Station 14. Subplot information same as Fig. 7.

Figure S13. Bar graphs showing the distribution of Grade-B (a) and excess arrivals (b) for PSE Station 15. Subplot information same as Fig. 7.

Figure S14. Bar graphs showing the distribution of Grade-B (a) and excess arrivals (b) for PSE Station 16. Subplot information same as Fig. 7.

Figure S15. Original trace (a) and FFT (b) of Earth seismic signal recorded at Piñon Flats Observatory station. Trace (c) and FFT (d) after adding Gaussian noise with a variance of 5 per cent the maximum amplitude.

Figure S16. Bar graphs showing the distribution of Grade-A (a) and excess arrivals (b) for PSE Station 12 with added noise. Subplot information same as Fig. 7.

Figure S17. Bar graphs showing the distribution of Grade-A (a) and excess arrivals (b) for PSE Station 14 with added noise. Subplot information same as Fig. 7.

Figure S18. Bar graphs showing the distribution of Grade-A (a) and excess arrivals (b) for PSE Station 15 with added noise. Subplot information same as Fig. 7.

Figure S19. Bar graphs showing the distribution of Grade-A (a) and excess arrivals (b) for PSE Station 16 with added noise. Subplot information same as Fig. 7.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

ACKNOWLEDGEMENTS

Many thanks to the NASA Postdoctoral Program for sponsoring this research. Conversations with Mark Panning and Sharon Kedar helped improve this work in its initial stages. Comments from the editor and two anonymous reviewers improved this paper.

REFERENCES

Allen R., 1982. Automatic phase pickers. Their present use and future prospects, *Bull. seism. Soc. Am.*, 72(6B), S225–S242.

[Google Scholar](#) [WorldCat](#)

Bates J., Lauderdale W., Kernaghan H., 1979. Alsep termination report, Tech. rep., NASA Reference Publication.

Bensen G., Ritzwoller M., Barmin M., Levshin A., Lin F., Moschetti M., Shapiro N., Yang Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, 169, 1239–1260.

[10.1111/j.1365-246X.2007.03374.x](https://doi.org/10.1111/j.1365-246X.2007.03374.x).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Bereuther M., Barsch R., Krischer L., Megies T., Behr Y., Wassermann J., 2010. Obspy. A python toolbox for seismology, *Seismol. Res. Lett.*, 81(3), 530–533.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Bottou L., Curtis F., Nocedal J., 2018. Optimization methods for large-scale machine learning, Tech. Rep., preprint (arXiv:1606.04838).

Brown J., Beroza G., Shelly D., 2008. An autocorrelation method to detect low frequency earthquakes within tremor, *Geophys. Res. Lett.*, 35(16),

doi:10.1029/2008GL034560.doi: [10.1029/2008GL034560](https://doi.org/10.1029/2008GL034560)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Bulow R., Johnson C., Shearer P., 2005. New events discovered in the Apollo lunar seismic data, *J. geophys. Res.*, 110(E10),

doi:10.1029/2005JE002414.doi: [10.1029/2005JE002414](https://doi.org/10.1029/2005JE002414)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Canup R., 2004. Simulations of a late lunar-forming impact, *Icarus*, 168(2), 433–456. [10.1016/j.icarus.2003.09.028](https://doi.org/10.1016/j.icarus.2003.09.028).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Caruana R., 1995, *Learning Many Related Tasks at the Same Time with Backpropagation*, MIT Press, pp. 657–664. [10.1.1.940.142](https://doi.org/10.1.1.940.142).

[Crossref](#)

Chollet F. et al., 2015. Keras, <https://keras.io>.

Clinton J. et al., 2018. The Marsquake service. Securing daily analysis of SEIS data and building the Martian Seismicity Catalogue for InSight, *Space Sci. Rev.*, 214(133), doi:10.1007/s11214-018-0567-5. doi: [10.1007/s11214-018-0567-5](https://doi.org/10.1007/s11214-018-0567-5)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Cybenko G., 1989. Approximation by superpositions of a sigmoidal function, *Math Control Signals*, 2(4), 303–314. [10.1007/BF02551274](https://doi.org/10.1007/BF02551274).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Dai H., MacBeth C., 1995. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophys. J. Int.*, 120(3), doi:10.1111/j.1365-246X.1995.tb01851.x. doi: [10.1111/j.1365-246X.1995.tb01851.x](https://doi.org/10.1111/j.1365-246X.1995.tb01851.x)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Dainty A., Toksöz M., 1981. Seismic codas on the Earth and the Moon, a comparison, *Phys. Earth planet. Inter.*, 26(4), 250–260. [10.1016/0031-9201\(81\)90029-7](https://doi.org/10.1016/0031-9201(81)90029-7).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Delalleau O., Bengio Y., 2011. Shallow vs. deep sum-product networks, in *Advances in Neural Information Processing Systems 24*, eds Shawe-Taylor J., Zemel R.S., Bartlett P.L., Pereira F., Weinberger K.Q., pp. 666–674., Curran Associates, Inc, <http://papers.nips.cc/paper/4350-shallow-vs-deep-sum-product-networks.pdf>.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

de Pater I., Lissauer J.J., 2010. *Planetary Sciences*, 2nd edn, Cambridge Univ. Press.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [Google Preview](#)
[WorldCat](#) [COPAC](#)

Dimech J., Knapmeyer-Endrun B., Phillips D., Weber R., 2017. Preliminary analysis of newly recovered Apollo 17 seismic data, *Results Phys.*, 7, 4457–4458.
[10.1016/j.rinp.2017.11.029](https://doi.org/10.1016/j.rinp.2017.11.029).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Duennebier F., 1976. Thermal movement of the regolith, in *Proceedings of the 7th Lunar Sci. Conf.*, Houston, TX, March 15–19, Vol. 1 (A77-34651 15-91), Pergamon Press, Inc., pp. 1073–1086.

[Google Scholar](#) [WorldCat](#)

Duennebier F., Sutton G., 1974. Thermal moonquakes, *J. geophys. Res.*, 79(29), doi: [10.1029/JB079i029p04351](https://doi.org/10.1029/JB079i029p04351) [10.1029/JB079i029p04351](https://doi.org/10.1029/JB079i029p04351).

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Dysart P., Pulli J., 1990. Regional seismic event classification at the NORESS array. Seismological measurements and the use of trained neural networks, *Bull. seism. Soc. Am.*, 80(6B), 1910–1933.

[Google Scholar](#) [WorldCat](#)

Friedler S., Lottick K., Susai S., 2019. *Energy-Usage*,
<https://github.com/responsibleproblemsolving/energy-usage>.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

GEOSCOPE - French Global Network of broadband seismic stations, 1982. Institut de Physique du Globe de Paris & Ecole et Observatoire des Sciences de la Terre de Strasbourg (EOST), doi:10.18715/GEOSCOPE.G.doi: [10.18715/GEOSCOPE.G](https://doi.org/10.18715/GEOSCOPE.G)

[WorldCat](#) [Crossref](#)

Giardini D., et al. , 2020. The seismicity of Mars, *Nat. Geosci.*, 13, 205–212.
[10.1038/s41561-020-0539-8](https://doi.org/10.1038/s41561-020-0539-8).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Gibbons S., Ringdal F., 2006. The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.*, 165(1), 149–166.
[10.1111/j.1365-246X.2006.02865.x](https://doi.org/10.1111/j.1365-246X.2006.02865.x).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Glorot X., Bordes A., Bengio Y., 2011. Deep sparse rectifier neural networks, in *Proceedings of the Artificial Intelligence and Statistics Conference (AISTATS) 2011*, pp. 315–323.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Goodfellow I., Bengio Y., Courville A., 2016. *Deep Learning*, MIT Press,
<http://www.deeplearningbook.org>.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Goodfellow I., Warde-Farley D., Mirza M., Courville A., Bengio Y., 2013. Maxout networks, in *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*.

[Google Scholar](#) [WorldCat](#)

Heffels A., Knapmayer M., Oberst J., Haase I., 2017. Re-evaluation of Apollo 17 Lunar Seismic Profiling Experiment data, *Planet. Space Sci.*, 135, 43–54.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Hinton G., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R., 2012. *Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors*, preprint (arXiv:1207.0580v1).

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Howard J. et al., 2018. fastai, <https://github.com/fastai/fastai>.

Jarrett K., Kavukcuoglu K., Ranzato M., LeCun Y., 2009. What is the best multi-stage architecture for object recognition?, in *Proceedings of the IEEE 12th International Conference on Computer Vision*, pp. 2146–2153., Kyoto, Japan ,
doi:10.1109/ICCV.2009.5459469.doi: [10.1109/ICCV.2009.5459469](https://doi.org/10.1109/ICCV.2009.5459469)

[Google Scholar](#) [Crossref](#) [Search ADS](#) [Google Preview](#)
[WorldCat](#) [COPAC](#)

Karpathy A., 2018. Stanford cs231n, <http://cs231n.github.io/convolutional-networks/>.

Knapmeyer-Endrun B., Hammer C., 2015. Identification of new events in Apollo 16 lunar seismic data by Hidden Markov Model-based event detection and classification, *J. geophys. Res.*, 120, 1620–1645. [10.1002/2015JE004862](https://doi.org/10.1002/2015JE004862).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Kovach R., Watkins J., 1973. Apollo 17 seismic profiling. probing the lunar crust, *Science*, 180(4090), 1063–1064. [10.1126/science.180.4090.1063](https://doi.org/10.1126/science.180.4090.1063).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [PubMed](#) [WorldCat](#)

Krizhevsky A., Sutskever I., Hinton G.E., 2012. Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems* 25, eds Pereira F., Burges C.J.C., Bottou L., Weinberger K.Q., pp. 1097–1105., Curran Associates, Inc, <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [Google Preview](#)
[WorldCat](#) [COPAC](#)

Larose E., Khan A., Nakamura Y., Campillo M., 2005. Lunar subsurface investigated from correlation of seismic noise, *Geophys. Res. Lett.*, 32(L16201) [10.1029/2005GL023518](https://doi.org/10.1029/2005GL023518) , doi:10.1029/2005GL023518.

[Google Scholar](#) [WorldCat](#) [Crossref](#)

LeCun Y., 1989. Generalization and network design strategies, Tech. Rep. CRG-TR-89-4, University of Toronto.

LeCun Y., Bengio Y., Hinton G., 2015. Deep learning, *Nature*, 521(7553), 436–444. [10.1038/nature14539](https://doi.org/10.1038/nature14539).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [PubMed](#) [WorldCat](#)

Li D., Harris J., 2018. Full waveform inversion with nonlocal similarity and model-derivative domain adaptive sparsity-promoting regularization, *Geophys. J. Int.*, 215, 1841–1864. [10.1093/gji/ggy380](https://doi.org/10.1093/gji/ggy380).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Lognonné P., 2005. Planetary seismology, *Annu. Rev. Earth Planet. Sci.*, 33, 571–604. [10.1146/annurev.earth.33.092203.122604](https://doi.org/10.1146/annurev.earth.33.092203.122604).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Lognonné P., Banerdt W., Giardini D., Coauthors, 2019. SEIS: insight's seismic experiment for internal structure of Mars, *Space Sci. Rev.*, 215(12), doi: [10.1007/s11214-018-0574-6](https://doi.org/10.1007/s11214-018-0574-6) [10.1007/s11214-018-0574-6](https://doi.org/10.1007/s11214-018-0574-6).

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Lognonné P., Gagnepain-Beyneix J., Chanut H., 2003. A new seismic model of the Moon. Implications for structure, thermal evolution and formation of the Moon, *Earth planet. Sci. Lett.*, 211(1–2), 27–44. [10.1016/S0012-821X\(03\)00172-9](https://doi.org/10.1016/S0012-821X(03)00172-9).

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Lorenz R., 2015. Energy cost of acquiring and transmitting science data on deep-space missions, *J. Spacecr. Rockets*, 52(6),

doi:10.2514/1.A33298.doi: [10.2514/1.A33298](https://doi.org/10.2514/1.A33298)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Lorenz R. et al. , 2018. Dragonfly: a rotorcraft lander concept for scientific exploration at Titan, *Johns Hopkins APL Tech. Digest*, 34(3), 374–387.

[Google Scholar](#) [WorldCat](#)

Meier M. et al., 2019. Reliable real-time seismic signal/noise discrimination with machine learning, *J. geophys. Res.*, 124, 788–800. [10.1029/2018JB016661](https://doi.org/10.1029/2018JB016661).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Mimoun D., Murdoch N., Lognonn  P., Hurst K., Pike W., Hurley J., N but T., Banerdt W., 2017. The noise model of the SEIS seismometer of the InSight mission to Mars, *Space Sci. Rev.*, 211, 383–428. [10.1007/s11214-017-0409-x](https://doi.org/10.1007/s11214-017-0409-x).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Miura A. et al. , 2000. ISAS Data Archive and Transmission System (DARTS) Astronomical Data Analysis Software and Systems IX, 216, 180.

[WorldCat](#)

Mont far G., Pascanu R., Cho K., Bengio Y., 2014. On the number of linear regions of deep neural networks, in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 2924–2932., MIT Press, NIPS 14.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Musil M., Ple inger A., 1996. Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps, *Bull. seism. Soc. Am.*, 86(4), 1077–1090.

[Google Scholar](#) [WorldCat](#)

Nair V., Hinton G., 2010. Rectified linear units improve restricted boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, Israel.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Nakamura Y., 2003. New identification of deep moonquakes in the Apollo lunar seismic data, *Phys. Earth Planet. Inter.*, 139(3–4), 197–205.

[10.1016/j.pepi.2003.07.017](https://doi.org/10.1016/j.pepi.2003.07.017).

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Nakamura Y., 2005. Farside deep moonquakes and deep interior of the Moon, *J. geophys. Res.*, 110(E01001), doi:10.1029/2004JE002332. doi: [10.1029/2004JE002332](https://doi.org/10.1029/2004JE002332)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Nakamura Y., Latham G., Dorman H., Harris J., 1981. Passive seismic experiment long period event catalog, Tech. rep., University of Texas Institute for Geophysics, <http://hdl.handle.net/2152/65671>.

Nakamura Y., Latham G., Dorman H., Ibrahim A., Koyama J., Horvath P., 1979. Shallow moonquakes: depth, distribution and implications as to the present state of the lunar interior, in *Proceedings of the Lunar and Planetary Science Conference, 10th*, Houston, TX, March 19-23, Vol. 3 (A80-23677 08-91), Pergamon Press, Inc., pp. 2299–2309.

[Google Scholar](#) [WorldCat](#)

Neal C. et al., 2020. The lunar geophysical network mission, Tech. rep., submitted to the Planetary Mission Concept Study program, NASA.

Nielsen M., 2015. *Neural Networks and Deep Learning*, Determination Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Nunn C. et al., 2020. Lunar seismology: a data and instrumentation review, *Space Sci. Rev.*, 216(89), doi: [10.1007/s11214-020-00709-3](https://doi.org/10.1007/s11214-020-00709-3) [10.1007/s11214-020-00709-3](https://doi.org/10.1007/s11214-020-00709-3).

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Oberst J., Nakamura Y., 1987. Distinct meteoroid families identified on the lunar seismograms, *J. geophys. Res.*, 92(B4), doi:10.1029/JB092iB04p0E769. doi: [10.1029/JB092iB04p0E769](https://doi.org/10.1029/JB092iB04p0E769)

[Google Scholar](#) [WorldCat](#) [Crossref](#)

Panning M., Stahler S., Huang H., Vance S., Kedar S., Tsai V., Pike W., Lorenz R., 2018. Expected seismicity and the seismic noise environment of Europa, *J. geophys. Res.*, 123, 167–179. [10.1002/2017JE005332](https://doi.org/10.1002/2017JE005332).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Phillips D., 2017, Analysis of rediscovered data from Apollo 17's Lunar Seismic Profiling Experiment: evidence for events associated with sunrise, *M.S. thesis*, The University of Alabama, Huntsville.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Poon H., Domingos P., 2011. Sum-product networks: a new deep architecture, in *The 27th conference on Uncertainty in Artificial intelligence (UAI 2011)*, Barcelona, Spain, <http://alchemy.cs.washington.edu/spn/poon11.pdf>.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [Google Preview](#)
[WorldCat](#) [COPAC](#)

Robbins H., Monro S., 1951. A stochastic approximation method, *Ann. Math. Stat.*, 22(3), 400–407. doi: [10.1214/aoms/1177729586](https://doi.org/10.1214/aoms/1177729586).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Ross Z., Meier M.-A., Hauksson E., 2018a. P wave arrival picking and first-motion polarity determination with deep learning, *J. geophys. Res.*, 123, 5120–5129. [10.1029/2017JB015251](https://doi.org/10.1029/2017JB015251).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Ross Z., Meier M.-A., Hauksson E., Heaton T., 2018b. Generalized seismic phase detection with deep learning, *Bull. seism. Soc. Am.*, 108(5A), 2894–2901. [10.1785/0120180080](https://doi.org/10.1785/0120180080).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Simmons G. et al. , 1970. Apollo 12 preliminary science report, Tech. Rep. SP-235, NASA, <https://www.hq.nasa.gov/alsj/a12/as12psr.pdf>.

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014. Dropout: a simple way to prevent networks from overfitting, *J. Mach. Learn. Res.*, 15, 1929–1958.

[Google Scholar](#) [WorldCat](#)

Trugman D., Shearer P., 2018. Strong correlation between stress drop and peak ground acceleration for recent M 1-4 earthquakes in the San Francisco Bay Area, *Bull. seim. Soc. Am.*, 108(2), 929–945. [10.1785/0120170245](https://doi.org/10.1785/0120170245).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Virtanen P., et al. , 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat. Methods*, 17, 261–272. [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [PubMed](#) [WorldCat](#)

Watters T., Weber R., Collins G., Howley I., Schmerr N., Johnson C., 2019. Shallow seismic activity and young thrust faults on the Moon, *Nat. Geosci.*, 12, 411–417. [10.1038/s41561-019-0362-2](https://doi.org/10.1038/s41561-019-0362-2).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Weber R., Lin P., Garnero E., Williams Q., Lognonné P., 2011. Seismic detection of the lunar core, *Science*, 331, 309–312. [10.1126/science.1199375](https://doi.org/10.1126/science.1199375).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [PubMed](#) [WorldCat](#)

Wieczorek M. et al. , 2012. The crust of the Moon as seen by GRAIL, *Science*, 339(6120), 671–675. [10.1126/science.1231530](https://doi.org/10.1126/science.1231530).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [PubMed](#) [WorldCat](#)

Zeiler M.D., Fergus R., 2014. Visualizing and understanding convolutional networks, in *Computer Vision – ECCV 2014*, eds Fleet D., Pajdla T., Schiele B., Tuytelaars T., pp. 818–833., Springer International Publishing.

[Google Scholar](#) [Crossref](#) [Search ADS](#) [Google Preview](#)
[WorldCat](#) [COPAC](#)

Zhan Z., Tsai V., Jackson J., Helmberger D., 2014. Ambient noise correlation on the Amery Ice Shelf, East Antarctica, *Geophys. J. Int.*, 196, 1796–1802. [10.1093/gji/ggt488](https://doi.org/10.1093/gji/ggt488).

[Google Scholar](#) [Crossref](#) [Search ADS](#) [WorldCat](#)

Published by Oxford University Press on behalf of The Royal Astronomical Society 2021.

This work is written by (a) US Government employee(s) and is in the public domain in the US.

Supplementary data

[ggab083_civilini_submission_supplementary](#) - pdf file

[ggab083_catalogs](#) - zip file