Geophysical Journal International

Geophys. J. Int. (2025) **240,** 483–501
Advance Access publication 2024 November 9
GJI Applied and Marine Geophysics

https://doi.org/10.1093/gji/ggae400

Classification of images derived from submarine fibre optic sensing: detecting broadband seismic activity from hydroacoustic signals

Ioannis Matthaiou,[1] Ali Masoudi,[1] Eiichiro Araki,[2] Shuichi Kodaira,[2] Stefano Modafferi[3] and Gilberto Brambilla[1]

[1] *Optoelectronics Research Centre, University of Southampton, Southampton SO17 1BJ, UK.*
E-mail: I.Matthaiou@soton.ac.uk
[2] *Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa 237-0061, Japan*
[3] *Digital Health and Biomedical Engineering, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK*

Accepted 2024 November 4. Received 2024 October 29; in original form 2023 August 21

SUMMARY

Distributed acoustic sensing (DAS) is an optoelectronic technology that utilizes fibre optic cables to detect disturbances caused by seismic waves. Using DAS, seismologists can monitor geophysical phenomena at high spatial and temporal resolutions over long distances in inhospitable environments. Field experiments using DAS, are typically associated with large volumes of observations, requiring algorithms for efficient processing and monitoring capabilities. In this study, we present a supervised classifier trained to recognize seismic activity from other sources of hydroacoustic energy. Our classifier is based on a 2-D convolutional neural network architecture. The 55-km-long ocean-bottom fibre optic cable, located off Cape Muroto in southwest of Japan, was interrogated using DAS. Data were collected during two different monitoring time periods. Optimization of the model's hyperparameters using Gaussian Processes Regression was necessary to prevent issues associated with small sizes of training data. Using a test set of 100 labeled images, the highest-performing model achieved an average classification accuracy of 92 per cent, correctly classifying 100 per cent of instances in the geophysical class, 80 per cent in the non-geophysical class and 96 per cent in ambient noise class. This performance demonstrates the model's effectiveness in distinguishing between geophysical data, various sources of hydroacoustic energy, and ambient noise.

Key words: Image processing; Machine learning; Distributed acoustic sensing.

1 INTRODUCTION

In seismology, there is a growing demand for real-time sensing and monitoring capabilities over large distances. Certain types of seismic events, such as teleseismic earthquakes-defined by the U.S. Geological Survey as seismic events occurring at distances greater than 1000 km from the sensing instrumentation are of significant importance to the geophysical community. Consequently, there is an increasing demand for monitoring systems that are sensitive enough to capture seismic activity even at such vast distances. Expanding sensing coverage presents numerous challenges, particularly given that approximately 71 per cent of the Earth's surface is covered by oceans (Grassl 2001).

To date, seismologists have primarily deployed single-point instrumentation, such as ocean-bottom seismometers and hydrophones. The time-domain records from these instruments are processed into short-time over long-time averages to facilitate the identification of seismic activity by closely examining the processed signals. Computing the ratio of ground-motion averages over two different fixed-time window lengths can effectively discriminate signals from background noise (Trnkoczy 2009).

Reliable sensing capabilities across remote and harsh environments not only enable the early detection of natural hazards (Sladen *et al.* 2019) but also offer better imaging of the Earth's subsurface structure. For these reasons, DAS technology is seen as instrumental in expanding sensing coverage and, thus, enhancing our understanding of important geological systems. Essentially, DAS transforms a dark optical fibre cable into a highly dense array of interconnected and synchronized sensors that measure the strain field over large distances. This is achieved by using an interrogator unit that sends laser pulses into the optical fibre and measures the phase shift in the Rayleigh backscattered light between adjacent points along the fibre—a technique known as phase-sensitive optical time-domain interferometry (ϕ-OTDR).

ϕ-OTDR is a relatively mature sensing technology that has already been deployed in the field for a wide range of sensing and condition monitoring applications. Examples include: (i) subsea cable damage detection (Masoudi *et al.* 2019); (ii) traffic monitoring and vehicle classification (Corera *et al.* 2023); (iii) train tracking and localization (Kowarik *et al.* 2020), (iv) monitoring changes in the dynamic behaviour of railway tracks (Milne *et al.* 2020) and (v) leak detection on gas pipelines (Muggleton *et al.* 2020). From a geophysical perspective, ϕ-OTDR-based distributed optical fibre sensing technology is particularly attractive because it can measure subtle changes in the strain field over a wide range of frequencies-something that cannot be achieved by other DAS technologies, for example Brillouin-OTDR (Masoudi & Newson 2016).

Using DAS, seismologists can carry out large-aperture experiments, cost-effectively and efficiently, with high sampling rates both spatially and temporally, by connecting an optoelectronic device (the interrogator unit) to an already installed dark fibre optic cable. The acquisition of coherent seismic waves, as they travel through the Earth's multilayered structure, can be captured and studied in greater detail (Biondi *et al.* 2021). For these reasons, DAS has recently been explored in a number of research studies to monitor and characterize geological features. Some of these studies are discussed in this section.

In Jousset *et al.* (2018) DAS was used to explore geological features, like the crustal structure, and in Jousset *et al.* (2022) the authors identified volcanic activity. In Agostinetti *et al.* (2022) the potential of DAS was investigated for mapping heterogeneities in shallow subsurface using a 8.9-km fibre optic cable on the Brady geothermal field. Using DAS, the earthquake fault ruptures were studied in Cochran (2018), while in Li *et al*. (2023a) the high-spatial sampling of DAS made it possible to determine *P*-wave polarities by cross-correlating earthquake pairs. The paper has shown improvements in the quality of computed focal mechanisms for each earthquake event examined. In Li *et al*. (2023b), imaged the high-frequency rupture radiators for megathrust earthquakes using the 100-km Long Valley DAS array. Similarly, in Cheng *et al.* (2021) a 20 -km fibre optic cable at Moss Landing Monterey Bay was deployed to examine submarine structural characteristics, including shallow fault zones while in Lior *et al.* (2023) a rigorous magnitude estimation and shaking intensity prediction methodology was developed specifically for DAS data. To enable conventional seismological analysis with DAS, for example for earthquake magnitude estimation, the strain field measurements were converted into a quantity called 'deformation' in Trabattoni *et al.* (2023). Additionally, hydrological near-surface characterization and seismic activity detection have been demonstrated using a 27-km-long fibre optic cable (with a spatial spacing of 2 m) at West Sacramento, recording 7 months of continuous data (Ajo-Franklin *et al.* 2019). In Baba *et al.* (2023) the authors detected tectonic tremors (a type of slow earthquake) for the first time using the same 50-km fibre optic cable as the one used in this study. These tremors were estimated to be located around a subducted seamount peak. In several studies, including (Agostinetti *et al.* 2022) and (Sladen *et al.* 2019), seismic activity monitoring using DAS has been found to be well-correlated with conventional seismic sensors and arrays of geophones, offering similar broadband sensing capability (Matsumoto *et al.* 2021). A comprehensive overview of DAS for seismology from a geophysical perspective can be found in Lindsey & Eileen (2021) while (Fernández-Ruiz *et al.* 2022) offer a similar overview but from a digital signal processing point-of-view.

The development of new and improved methodologies to analyse seismic records has seen a tremendous rise in recent years (Mousavi & Beroza 2023). This is especially true in fibre optic seismology where data are generated in large volumes, for example about 1 TB per day as reported in Lellouch *et al.* (2019). Given such an immense scale of data, researchers have been exploring ways to utilize them for seismological applications. To improve the signal-to-noise ratio, self-supervision was used in van den Ende *et al.* (2021) to suppress incoherent noise. To remove different types of noise on images of the strain field, that is as a function of time and cable distance, studies in Yang *et al.* (2023a) and later in Yang *et al.* (2023b) used 2-D convolutional neural networks in a U-Net architecture. Similarly, in Li *et al.* (2022) a 2-D convolutional neural network in an auto-encoder architecture was trained with the aid of a synthetic noisy data set (and associated annotations) to remove six different types of noise on similar image data.

Convolutional neural networks with 1-D input data were widely adopted in recent studies to take into account the large-volumes of seismic waveforms generated from seismological experiments. In Majstorović *et al.* (2021) a binary classifier was trained using seismometer data over a 30-yr period, where earthquake signals were discriminated against ambient noise, demonstrating a substantial increase in the number of detected earthquakes. Additionally, in Majstorović *et al.* (2023) the feature maps extracted from convolutional neural networks were investigated, reporting major improvement (in terms of binary classification) on both the amplitude and waveform frequency of the seismic signals. Similarly, in Jiang *et al.* (2023) a binary classification model was trained to discriminate tectonic tremors and teleseismic earthquakes. Moreover, an encoder–decoder architecture was used to separate ambient noise from seismic signals in Yin *et al.* (2022) and in Liu *et al.* (2021) a binary classifier was trained to discriminate tectonic and non-tectonic tremor signals. In Hernández *et al.* (2021) a 1-D convolutional neural network was trained using a large database of time-domain seismic waveforms from seismometers, which was then used to detect earthquakes using a different data set of DAS records. To expand DAS data synthetically, Shiloh *et al.* (2019) and Shiloh *et al.* (2020) explored generative adversarial networks.

Convolutional neural networks with 2-D input data were also used in recent years. For instance, in Liu *et al.* (2022) man-made microseisms were recognized using the guided wave energy as features. In Huot *et al.* (2022a,b), a binary classifier with optimized hyperparameters was trained where microseismic events and background noise were discriminated using around 7000 DAS records. Inputs were Wavelet scalograms. In Nakano *et al.* (2019) the authors trained a classifier to distinguish between tectonic tremors and local earthquakes with spectral images as inputs. In Mousavi *et al.* (2019) an unsupervised classification framework was presented for discriminating between local and teleseismic events using Short-Time Fourier Transform spectrograms as inputs.

Other studies that used convolutional neural networks and are worth mentioning include the ones in Ren *et al.* (2023a) for picking dispersion curves using amplitude spectra (inputs) and Scholte waves (outputs) and in Chen *et al.* (2023) an analysis of clusters was presented using techniques such as hierarchical clustering for a set of similar input data. Hypocentre location using DAS records was estimated in Mousavi & Beroza (2022), magnitude estimation in Ren *et al.* (2023b), velocity model improvement in Muller *et al.* (2023) and inversion of teleseismic *P*-wave receiver function and surface wave dispersions in Gan *et al.* (2023).

DAS technology for monitoring seismic activity is still relatively new and major challenges, including: (i) low signal-to-noise ratio, (ii) sensing range, (iii) incidence angle sensitivity for incoming seismic waves and (iv) large-scale data management, have to be addressed for its widespread adoption. A low signal-to-noise ratio can be caused due to poor mechanical coupling between Earth and fibre optic cable. This issue has been investigated in recent years [for instance, see Matsumoto *et al.* (2021) and Harmon *et al.* (2022)]. On the other hand, DAS is mainly sensitive to the axial strain of the fibre optic cable which limits the sensing capability to 1-D measurement of the strain field.

In this study, we explored 2-D convolutional neural networks for tackling the issue of large-scale data management by constructing a multi-class classifier that can discriminate between geophysical and non-geophysical sources of hydroacoustic energy and ambient noise. The structure of this paper is as follows: in Section 2 the data set is presented including the signal processing steps used for data transformation. In Section 3 the classifier's main characteristics and architecture are discussed. In Sections 4 and 5 the classifier's performance for DAS records is analysed, and in Section 6 the conclusions of the study are presented.

2 DATA SET DESCRIPTION

2.1 Monitoring seismic activity in the Nankai Trough

The Muroto fibre optic cable is located in the southwestern part of Japan within the Nankai subduction zone, where the Philippine Sea Plate is subducting beneath the Amur Plate. Since 1997, the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), has been using the Comprehensive Seafloor Monitoring System offshore Cape Muroto, which now connects a seismic land station to six single-mode fibre optic cables. The total length of the Muroto fibre optic cable is 128 km. Between 0.35 and 2.1 km from the land station, the cable is buried under the seafloor, while from 2.1 km and onwards it rests on the seafloor. In their study, Karrenbach *et al.* (2021) demonstrated the use of DAS for seismic monitoring using the Muroto fibre optic cable.

The Nankai Trough is a well-known seismogenic zone, where megathrust earthquakes with moment magnitudes (M_w) greater than 8 occur every 100–150 yr. The last megathrust earthquake that originated in the Nankai Trough happened in 1946, and therefore, the area has been extensively monitored over the last decade. For improving understanding of fault slip events and developing better monitoring systems for megathrust earthquakes, a dedicated array of interconnected seismographs, known as the Dense Oceanfloor Network system for Earthquakes and Tsunamis (DONET) is being actively used for monitoring and recording continuous measurements.

Using a dense network of seismometers, it is possible to identify and analyse a wide range of slow earthquakes in the Nankai Trough. Such events include: long-term and short-term slow slip events (time durations of days to years), very low frequency earthquakes (time durations between 10 and 100 s) and tectonic tremors Yamamoto *et al.* (2022). In Nakano *et al.* (2018) the authors used the vertical components of ocean-bottom seismometers to show that very low frequency earthquakes are temporally correlated with tectonic tremors in the Nankai Trough. At the same time, very low frequency earthquakes were found to have been triggered by large earthquake events. For instance, the authors in Wallace *et al.* (2021) reported swarms of very low frequency earthquakes and tectonic tremors (due to plate ruptures), just after the main shock of the M_w 6.0 Mie-ken Nanto-oki earthquake on 01 April 2016. Additionally, in Takemura *et al.* (2022) the authors studied very low frequency earthquakes in the Nankai Trough (using DONET records between April 2004 and March 2021) and showed that these type of events occured around the western edge of the subducted oceanic ridge.

In our study, a total of 571 geophysical events were recorded from our DAS system. Furthermore, by closely examining the spatial and temporal characteristics of each record, different earthquake types were identified:

(i) *Local earthquakes*: A total of 447 local earthquakes were recorded between 30 January 2022 and 23 March 2022. Another 16 local earthquakes between 17 August 2021 and 02 October 2021.

(ii) *Tectonic tremors*: A total of 29 tectonic tremors were recorded between 30 January 2022 and 02 August 2022.

(iii) *Teleseismic earthquakes*: A total of 79 teleseismic earthquakes were recorded between 30 January 2022 and 02 August 2022.

Note that the above geophysical events were identified by experts by inspecting the DAS records.

To validate and cross-reference the identified seismic events using our DAS system, we compared our data set with the official earthquake catalogue as provided by the Japan Meteorological Agency (JMA) for the period between 30 January 2022 and 23 March 2022. During this time-frame, the JMA recorded a total of 38 570 earthquakes across Japan. The JMA catalogue offers comprehensive details for each seismic event-including date, time, geographic coordinates, depth and magnitude-making it one of the most accurate and informative references for seismic activity in the region. Fig. 1 shows the topographic map of Japan with the epicentres of each cross-referenced geophysical event recorded by DAS. As mentioned previously, we identified three types of earthquakes: local earthquakes, tectonic tremors and teleseismic earthquakes, which are labelled on the map accordingly.

The cross-referencing procedure was performed by first aligning the date and time stamps of the DAS records with those listed by the JMA catalogue. Subsequently, a magnitude filter was used to exclude events below a certain threshold-specifically, earthquakes with a vertical component magnitude of less than 0.5—as such low-magnitude events are unlikely to produce strain signals detectable by DAS. We prioritized events with higher vertical magnitudes to focus on those most likely to be captured by the DAS system. Moreover, we used the Haversine formula to calculate the geodesic distance between the epicentre of each earthquake and every point along the fibre optic cable. By iteratively applying this calculation along the entire set of cable coordinates, we identified the minimum possible distance from the cable to each earthquake epicentre. Based on these computed distances, we labelled earthquakes with a minimum distance greater than 1000 km from any point along the fibre optic cable as teleseismic events, in accordance with standard seismological definitions (e.g. as used by the U.S. Geological Survey).

The highest magnitude recorded within this period was 4.3, corresponding to a teleseismic event located 1643 km from the optical fibre cable, originating off the coast of Taiwan (visible at the bottom left in Fig. 1). Tectonic tremor events, were generally of low magnitude ranging from 0.7 to 2.3. In contrast, the magnitude of local earthquakes varied more widely, from 0.9 to 4.0 in the vertical component. The median magnitude of the recorded DAS events was 1.4, while the maximum recorded depth was 420.9 km. Note that the magnitudes listed in the JMA catalogue refer to the vertical component of the seismic signal, which is relevant for DAS detection capability.

Fig. 2, shows a typical DAS signal (amplitude normalized) representing a local earthquake event as measured at a distance of about 15 km from the fibre optic cable. Figs 2(a) and (b) show the time-domain strain-rate for about 55 seconds, for the filtered (bandpass filter applied between 0.005 and 20 Hz) and raw signals, respectively. In Fig. 2(a), the body waves are clearly observed, while in the raw signal only the *P*-wave component can be distinguished from

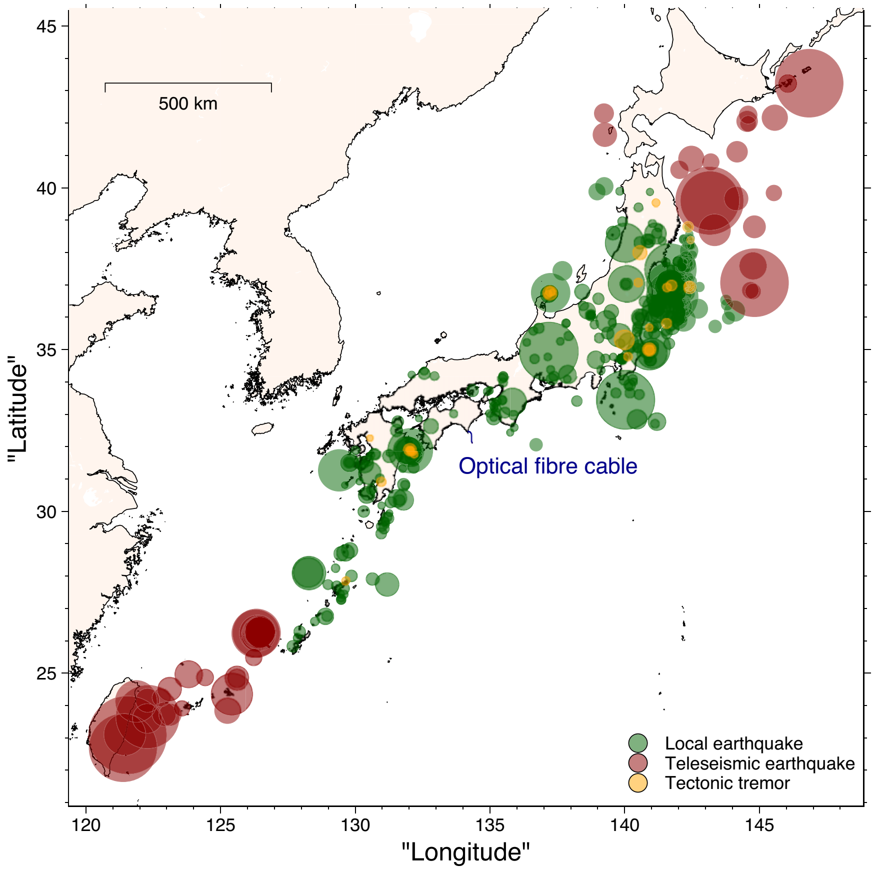


Figure 1. Topographic map of Japan showing 555 geophysical events recorded using DAS during monitoring period A, that is between the period 30 January 2022 and 23 March 2022. The three geophysical event categories as shown on the map are: 'Local earthquake' (in green), 'Teleseismic earthquake' (in red) and 'Tectonic tremor' (in yellow). The size of each circle indicates the magnitude of each event: ranging from 0.5 (smallest circle) up to 4.3 (largest circle). The location of the DAS events recorded were cross-referenced with the earthquakes recorded by the Japan Meteorological Agency (www.jma.go.jp).

noise. The corresponding power spectrum of the filtered signal, shown in Fig. 2 (c), indicates that there is significant seismic energy between about 2.5 and 11 Hz. On the other hand, the Wavelet scalogram in Fig. 2(d) shows that *P*-wave energy has a broader spectrum, which ranges from sub-Hz up to about 50 Hz. While, *S*-wave energy is mainly concentrated in the lower frequencies, that is about less than 20 Hz. As also indicated in Fig. 2(d), there is a substantial level of noise present at frequencies above 25 Hz.

The tectonic non-volcanic tremor observed in the Nankai Trough was identified as a slow slip event by the JMA (Katsumata & Kamaya 2003). This type of geophysical event has a range of features that are very distinct from local earthquakes. The majority of tectonic tremor signals recorded by either DAS or conventional seismometers indicate significant depletion in *P*-wave energy. The lack of impulsivity in the recorded data means that signals representing tectonic tremors are less broadband in nature, as compared to local earthquake signals that are highly energetic. In Fig. 3 strain field measurements representing a single tectonic tremor event are shown on four different locations along the Muroto fibre optic cable. More specifically, the first row of subplots, that is Figs 3(a)–(d) show the bandpass filtered waveforms in units of nanostrain per second (ns s^{-1}) at four different locations along the cable. In channels where signal-to-noise ratio was relatively low in comparison to the rest of the channels, robust *z*-score standardization was applied throughout the full data set. As observed, Figs 3(a)–(c) have a much lower dynamic range of strain-rate values, as compared to Fig. 3(d). Rescaling all channels within the statistical range is beneficial to avoid 'masking' low signal-to-noise ratio channels. Figs 3(e)–(h) show the corresponding rescaled signals, while the power spectral densities are also shown in Figs 3(i)–(l).

As it is evident from these plots, tectonic tremor signals have significant body wave energy at lower frequencies (about less than 3 Hz). With the exception of low signal-to-noise ratio channels as the one shown in Figs 3(c), (g) and (k), this statement holds true. This low signal-to-noise ratio channel is shown to contain energy at a broader range of frequencies, with no distinct signal components, as the rest of the channels shown in this example. A low signal-to-noise ratio in certain channels is a common feature in DAS data acquisition due to numerous factors including optical fading and/or poor ground-cable mechanical contact/coupling.

In two different studies, the authors in Baba *et al.* (2023) and Shelly *et al.* (2007) also analysed tectonic tremor signals obtained in the Nankai Trough, and suggested that most of their energy exists in the frequency range between about 1 and 8 Hz. In particular, the tectonic tremor DAS records (29 events in total) used in this study were previously analysed in Baba *et al.* (2023) and their analysis agrees with the previous claims made regarding signal characteristics.

Using DAS along the first 55 km of the fibre optic cable more than 9800 signals (equally spaced) were recorded, similar to the ones that were shown in Fig. 3. Hence, a 2-D representation can be plotted, showing the strain field both as function of total elapsed

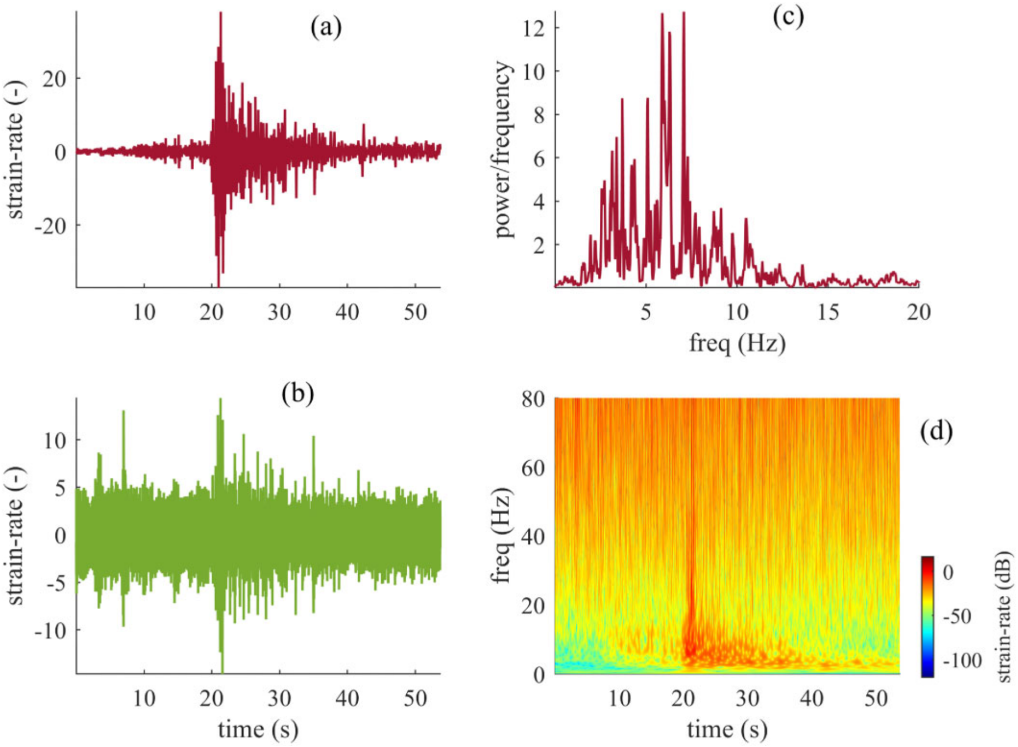


Figure 2. Local earthquake signal recorded by DAS at a minimum distance of about 15 km from the optical fibre cable: (a) filtered and rescaled signal (using robust *z*-score standardization), (b) raw signal, (c) power spectral density of the filtered signal and (d) wavelet scalogram of the raw signal.

time and distance along the cable. Fig. 4 shows an example of a 2-D representation for the previously presented tectonic tremor event (occurred on 02 February 2022). As can be seen from this plot, there are several locations along the fibre optic cable where signal-to-noise ratio is high enough to reveal the event clearly, while there are other locations where the signal was poor, for example compare signals at a distance of 50 km with ones at around 35 km. This is mainly due to the differences in physical contact between seabed/ground and fibre optic cable. Also, in the first few kilometres, ocean waves can be seen very clearly as they are capable of exciting the cable very effectively due to their broadband energy. Note that the signals prior to generating the image in Fig. 4 were bandpass filtered between 1 and 8 Hz prior to assembling them in a 2-D matrix, which was further processed using median spatial filtering (3 × 3 window). In Section 2.2 such processing steps are discussed in greater detail.

All teleseismic earthquakes used in our study were provided after bandpass filtering in the 2–10 Hz frequency range. This frequency range differs from the conventional approach in seismological studies, where teleseismic events are typically analysed within the sub-Hz range, where most of their energy is concentrated. For instance, Ajo-Franklin *et al.* (2019) applied a bandpass filter between 0.01 and 0.1 Hz to analyse signals from teleseismic events.

Therefore, the use of the 2–10 Hz frequency range in this set of DAS measurements enables us to monitor teleseismic earthquakes using the higher-frequency components of the signals, which are less commonly studied. While the rationale for applying this specific filter prior to data provision is beyond the scope of our study, it presents an opportunity to investigate the capabilities of DAS in detecting and analysing the higher-frequency part of seismic waves.

In our observations, the seismic phases—including P-phases and T-phases—lasted several minutes. T-phases, which propagate through the water column at velocities ranging from 1.4 to 1.5 $\mathrm{km\,s^{-1}}$, were effectively captured within this frequency range. This suggests that the DAS system is sensitive to these higher-frequency components, providing valuable data for our analysis.

DAS records were also obtained from a range of non-geophysical hydroacoustic energy sources between the same dates, that is between 30 January 2022 and 23 March 2022. The exact data distribution for a total of 315 hydroacoustic energy sources of non-geophysical nature are as follows: 11 ship-based airgun shots, 245 ship vessels sailing near the fibre optic cable and 59 marine mammal movements. Ship-based airgun shots were used to explore the sensing capability of DAS, as explained in more detail in Matsumoto *et al.* (2021). DAS measurements were compared with hydrophone signals obtained from airgun shots where the majority of wave energy is within the frequency range of 5 and 10 Hz. In that study, the authors concluded that DAS offers a similar sensing capability to hydrophones.

A similar work to the one discussed above, Ide *et al.* (2021), showed that strain measurements across the fibre optic cable were highly correlated with ocean-bottom conventional seismographs. However, at certain locations of the Muroto fibre optic cable, there were no signal records due to poor coupling between the seabed/ground and cable (as already shown in Fig. 4). In addition, monitoring very low frequency earthquakes in the ranges of 0.02–0.05 Hz, DAS signals were reported to have a much lower signal-to-noise ratio than seismographs.

2.2 Transforming signals into straingrams for classification

Two different DAS interrogators were installed on the Muroto fibre optic cable to record hydroacoustic signals related to seismic

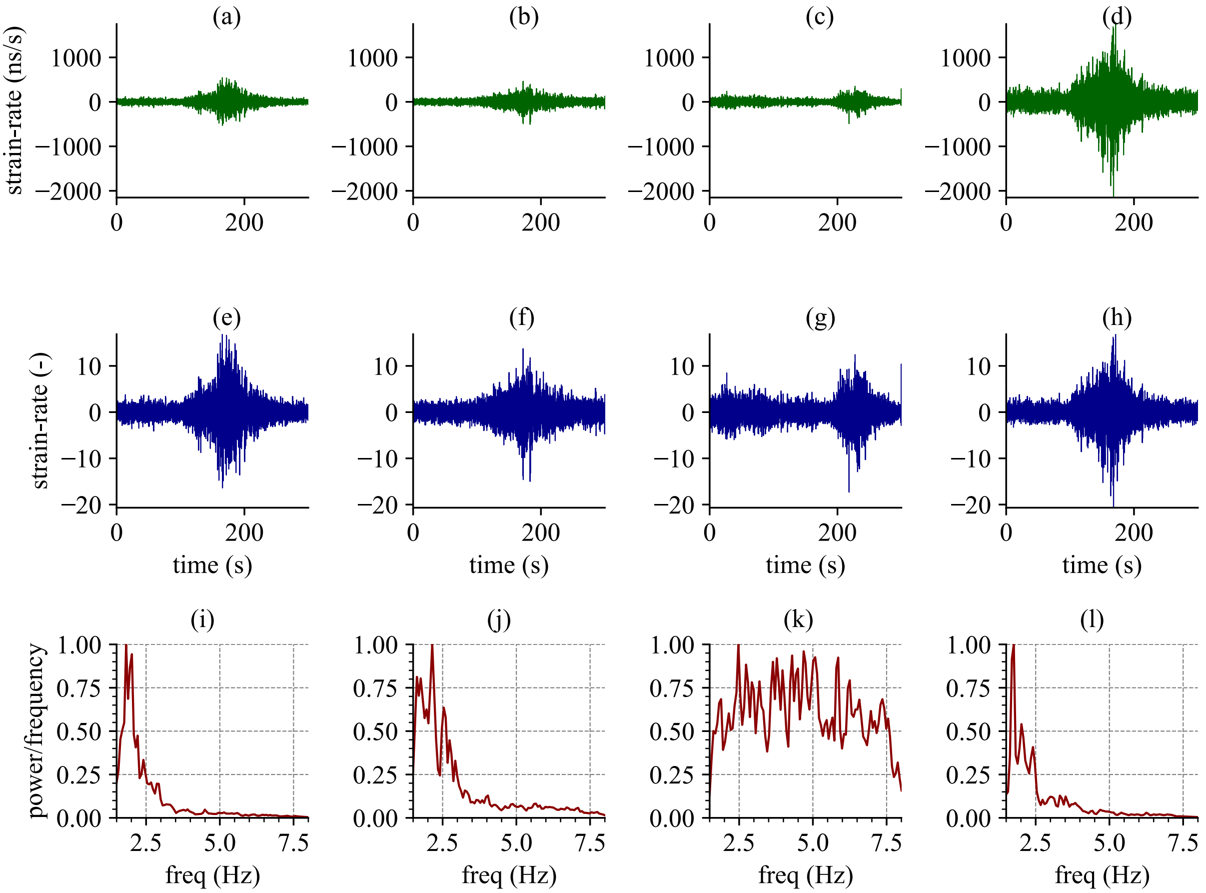


Figure 3. Four tectonic tremor signals recorded on the Muroto fibre optic cable using DAS on 02 February 2022: (a)–(d) strain-rate at different locations along the fibre optic cable, (e)–(h) rescaled strain-rate (robust z-score standardized) and (i)–(l) corresponding normalized power spectral densities.

activity in the region. Both systems used the first 55 km (from the coast) of the cable as their total effective sensing ranges. The first DAS system was developed by AP Sensing GmbH, while the second one was developed by the Distributed Optical Fibre Sensing (DOFS) research group at the University of Southampton in the United Kingdom. The latter system had a higher spatial and temporal resolution. The main specifications of the two DAS interrogators used in this study, including their respective monitoring periods, are given in Table 1.

For monitoring period A (30 January 2022 up to 23 March 2022) the interrogator recorded about 9800 waveforms of strain-rate or strain field along the Muroto fibre optic cable. The recorded temporal sampling rate was 500 Hz, while the spatial resolution was 5.1 m. Spatial downsampling (by averaging in-between samples) was applied to reduce data storage requirements. Also, for that purpose, temporal resolution was decreased to 100 Hz. Moreover, a bandpass filter was applied to each of the 980 waveforms individually, so that frequencies outside the 2–10 Hz band were attenuated [similar to the work presented in Baba *et al.* (2023)]. Plots such as the one shown previously in Fig. 4 were constructed by concatenating all the pre-processed 980 waveforms into a single 2-D matrix. A fixed-time window with a total duration of 3 min converted the continuous stream of data into the 2-D matrices (hereafter ‘straingrams’). In this study, the straingram is an effective transformation of our raw waveform data that allows efficient characterization and discrimination of the identified events both qualitatively and quantitatively. Using straingrams one can visualize the evolution of strain across different parts of the fibre optic cable as a function of time for different hydroacoustic signals.

On the other hand, the raw data from monitoring period B (17 August 2021–02 October 2021) were subjected to a different digital signal and image processing treatment due to the availability of raw signals. Instead of bandpass filtering each waveform between 2 and 10 Hz, each waveform in monitoring period B comprises of a broader range of frequencies, that is between 0.005 and 20 Hz. This is to ensure that different types of earthquakes can be effectively detected and identified by DAS. For instance, in Jiang *et al.* (2023) a bandpass filter encompassing the frequency range of 5–25 Hz was applied on raw DAS waveforms in order to obtain only local microseismic activity. In that way, teleseismic earthquakes that occur in lower frequency regions were effectively filtered out from their DAS records. While in Sladen *et al.* (2019) the authors applied bandpass filtering between 1 and 15 Hz to each waveform recorded by DAS. In our study, using the frequency band of 0.005–20 Hz, the expected seismic activity in the region can be more effectively monitored. Broad-band signals due to local earthquakes of 2–20 Hz, teleseisms of 0.005–1 Hz and tectonic tremors of 1–8 Hz can all be measured in monitoring period B. Due to instrumentation limitations, only 16 straingrams representing local earthquakes were constructed in this monitoring period. Typical seismic waveforms on three locations along the Muroto cable that represent a local earthquake in monitoring period B are shown in Fig. 5. These waveforms are shown as the robust z-score scaled values of the strain field. Using this

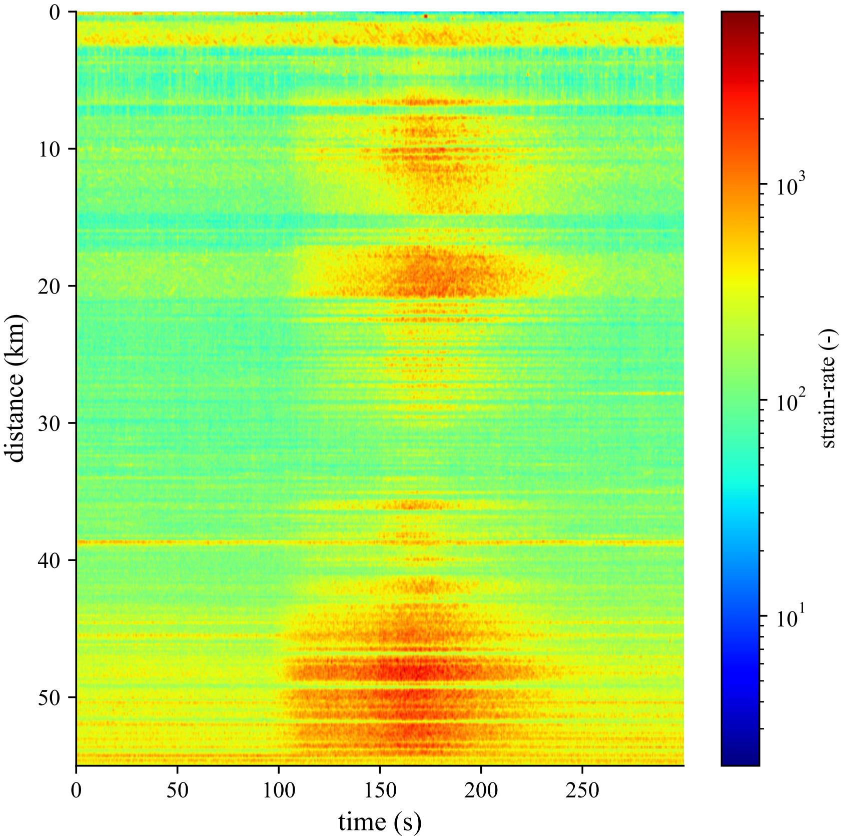


Figure 4. Strain field measurement as a function of time elapsed and distance along the fibre optic cable for a tectonic tremor that occurred on 02 February 2022. Pre-processed strain waveforms were assembled in a 2-D matrix and median spatial filtering was applied as part of the processing stage to generate this image and enhance the visibility of strain variation across the cable.

Table 1. Main specifications of the two DAS interrogators used in this study to monitor hydroacoustic activity in the Nankai Trough, while being connected to the Muroto fibre optic cable.

MONITORING PERIOD A	
Developer	AP Sensing GmbH
Technology	ϕ-OTDR
Spatial resolution	5.1 m
Temporal sampling rate	500 Hz
Resulting number of channels	≈ 9800
Monitoring period	30 January 2022 to 23 March 2022
MONITORING PERIOD B	
Developer	DOFS research group, University of Southampton
Technology	ϕ-OTDR
Spatial resolution	1 m
Temporal sampling rate	1 kHz
Resulting number of channels	≈ 55 000
Monitoring period	17 August 2021 to 02 October 2021

standardization or normalization procedure, we subtract each sample with the median value of the waveform and divide by the mean absolute deviation. The advantage of such a normalization procedure is in the identification of extreme values in each waveform, which corresponds to impulses and other outlying events in the seismic waves of DAS records. Hence, both *P* and *S* waves can be more easily discriminated than the remaining signal components, as shown in these plots. Also, the time difference of each strain transient event, that is the beginning of the *P* wave, which is related to the light reaching each channel along the fibre, is also apparent on these plots. With regard to robust measures of location and scale, many techniques have been studied in the past and implemented on different problems. For instance, in Matthaiou (2022) robust statistical outlier analysis on multivariate and high-dimensional input feature domains was demonstrated on a range of different vibration-based sensing data sets. While, in Schreurs *et al.* (2021) an anomaly detection method was presented for input spaces that are not constrained to be elliptical.

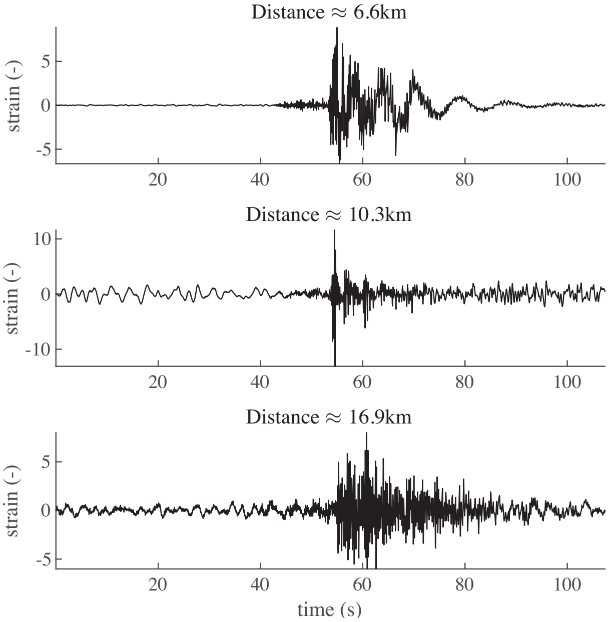


Figure 5. Strain field measured (robust *z*-score standardized) at three different locations across the DAS fibre optic cable. In these three plots, the waveform of a local earthquake is captured during monitoring period B. *P* and *S* waves are clearly visible on the three waveforms.

Moreover, the transformation of monitoring period B records into straingrams, suitable for classification purposes, includes a 2-D spatial filtering stage that enhances its signal-to-noise ratio value. In particular, a 2-D zero-mean symmetric Gaussian of the form $g[i, j] = \exp(-(i^2 + j^2)/2\sigma_u^2)$ [with standard deviation $\sigma_u \approx [1, 5]$ Szeliski (2022)] was applied on the bandpass filtered 2-D matrix. This filter was chosen since it has many suitable properties for smoothing these DAS matrices, including relatively low computational cost (e.g. as compared to the non-local means method). The enhancement in signal-to-noise ratio for the straingram is important in the subsequent steps of classifier training since discriminatory image features can be more effectively extracted. Additionally, learning noise-related features from the straingrams can be detrimental to classification accuracy. This is due to the fact that we wish to discriminate events from background noise. Additionally, the inclusion of noise in the straingrams is expected to require a more complex network architecture, necessitating larger data sets to train on. Note that excessive spatial smoothing is also expected to impact classification accuracy since important image features such as object edges and corners will be smoothed out. Pre-processing data to improve classification accuracy in convolutional neural networks has already been observed in many studies, for example Jernelv *et al.* (2020). In Matthaiou *et al.* 2023, we previously examined different image pre-processing techniques in order to improve the images derived from DAS signals in noisy environments.

A typical example illustrating the effectiveness of enhancing the signal-to-noise ratio for straingrams, using the above set of methods, is shown in Fig. 6. Given a relatively large range of strain values on the raw input 2-D DAS matrix representing this local earthquake [26 Septe2021 at 17:29 (UTC)], a logarithmic normalization (base 10) was applied (Fig. 6a). The same 2-D DAS matrix is constructed as shown in Fig. 6(b), after each individual waveform is bandpass-filtered between the range of 0.005–20 Hz. In addition, Fig. 6(c) shows the resultant DAS matrix, after convolving Fig. 6(b) with the 2-D Gaussian spatial smoothing filter, with $\sigma_u = 5$. While, Fig. 6(d) shows the final DAS matrix output from this procedure, that is the straingram, which is a result of applying robust *z*-score standardization on each of the pixel values (logarithmic normalization to the base 10 is also applied as a final step). This standardization procedure effectively normalizes the data in the same range of values, while highlighting extreme values (as discussed previously) on the DAS records: raising the signal-to-noise ratio of the DAS matrix. This is especially valuable at larger distances from the Muroto coast. Since the waveforms from the first few kilometres where the fibre optic cable was buried and ocean wave energy was significant in the DAS records were removed from the DAS matrix. Also, in monitoring period B, given that only local earthquakes were recorded, a fixed-time window of 50 s was applied on the continuous waveforms to generate the straingrams (mainly to decrease computational processing requirements associated with large DAS matrices). Note that on both DAS interrogators, significant attenuation of laser pulses at higher distances from the coast was observed for local earthquakes. As seen previously, in Fig. 5, the strain data at a distance of about 16.9 km from the coast of Cape Muroto exhibit a lower signal-to-noise ratio compared with the data collected from locations closer to the shore.

In total, six different types of DAS signals were identified: airgun shots, sailing ships, the passing of marine mammals, local earthquakes, tectonic tremors and teleseismic events. Straingrams for each of these signals were constructed according to the procedure that was outlined earlier. An example for each of the identified events is shown in Figs 7 and 8. Fig. 7 shows three straingrams representing three different geophysical events, where the *x*-axis is time (3 min in total) and the *y*-axis is the distance from the coast along the fibre optic cable (whole 55 km span). In particular, Fig. 7(a) shows a straingram of a local earthquake, while in Figs 7(b) and (c) the straingrams represent a teleseismic earthquake and a tectonic tremor, respectively. As it is shown, local earthquakes have sharper onsets of strain and occur for about less than a minute or so. On the other hand, teleseismic earthquakes last for minutes to hours, and thus can typically occupy all or a larger proportion of the straingram as they are enriched in lower-frequency energy. The propagation of seismic wave energy along the fibre optic cable will typically be much lower than for local earthquakes. This is apparent in the strain field variation patterns that appear more diagonal than local earthquakes. While patterns observed with teleseismic events in the straingrams are largely similar to tectonic tremors, tectonic tremor straingrams typically contain a single event, occupying a certain proportion of the image. In Fig. 8(a) a straingram of repeated airgun shots is shown, while in Figs 8 (b) and (c) show the straingram of a ship sailing and marine mammals' sounds (either due to crossing it or from other hydroacoustic sound waves emitted by it), respectively.

The total amount of straingrams constructed for geophysical and non-geophysical classes of events, were 517 and 315, respectively. Additionally, a truly automated processing capability for DAS records demanded discrimination of both of these two classes from background noise, as well. Hence, a third class of events was considered with 164 straingrams representing background noise (hereafter 'noise'). In summary, the total number of available (and annotated) straingram examples was 1050. Stratified random sampling was employed to split the dataset into 800 training samples, 150 validation samples, and 100 testing samples, ensuring that the class distribution was preserved across each subset.

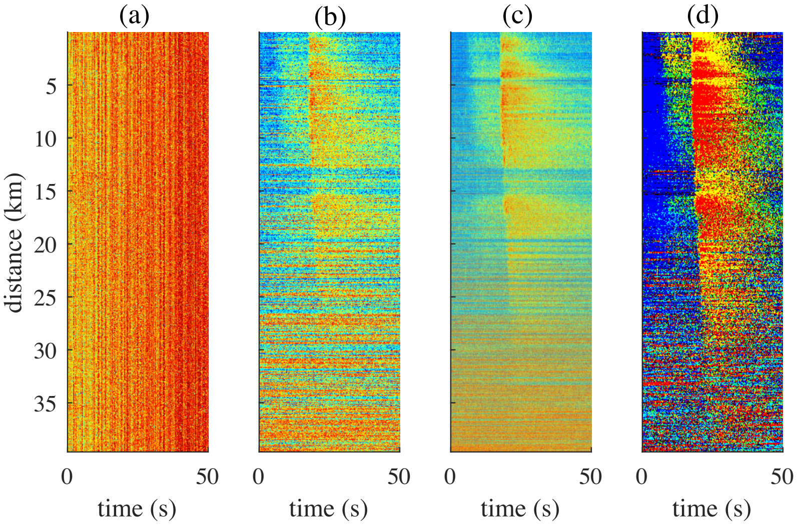


Figure 6. A straingram generated using a set of processing steps: (a) the original raw 2-D DAS matrix, (b) the bandpass filtered (0.005–20 Hz) 2-D DAS matrix, (c) 2-D spatial smoothing filter applied on the image in (b) and (d) robust *z*-score standardization applied on the image in (c). In each step (a)–(d) the strain field is log-normalized.

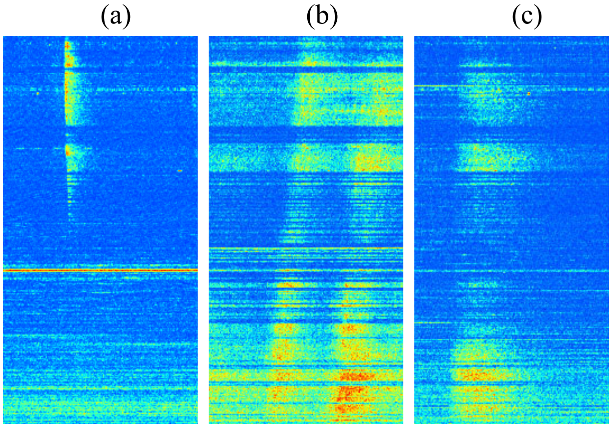


Figure 7. Straingrams (*x*-axis is time, *y*-axis is distance, *z*-axis or colour is strain amplitude), representing geophysical events captured during monitoring period A: (a) local earthquake, (b) teleseismic earthquake and (c) tectonic tremor. Straight horizontal lines are spurious features due to various noise sources across the fibre optic cable.

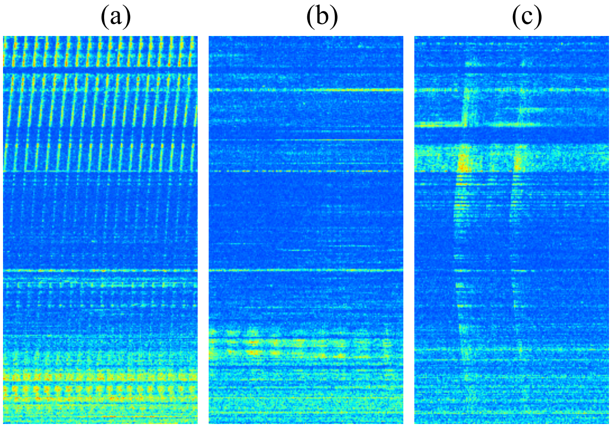


Figure 8. Straingrams (*x*-axis is time, *y*-axis is distance, *z*-axis or colour is strain amplitude), representing non-geophysical events captured during monitoring period A: (a) airgun shots, (b) ship vessels and (c) marine mammals. Straight horizontal lines are spurious features due to various noise sources across the fibre optic cable.

3 CONVOLUTIONAL NEURAL NETWORKS FOR CLASSIFYING STRAINGRAMS

As was discussed and observed in Section 2, the constructed straingrams representing each of the six events identified (in monitoring periods A and B) have a range of distinguishing features. From a computational perspective, these characteristics can be seen as differences in the edges: boundaries of a particular object in the straingram and its shape or feature pattern. Also, differences in contrast are expected to be an important discriminating factor, which will be attributed to the absolute value of the strain field with respect to the background noise. For instance, sailing ships, as seen in Fig. 8(b), have a weaker signal in comparison to a local earthquake in Fig. 7(a). It is evident from the straingrams that sailing ships are far less distinguishable from the background noise levels (varies along the fibre), which is due to a lower dynamic range of their strain field. Similarly, the teleseismic event recorded in Fig. 7(b) is also of low dynamic range, where straight horizontal lines (noise either due to optical fading or poor physical contact of the cable) are profound in the image. Therefore, other characteristics will also need to be extracted from these images in order to achieve a robust discrimination of the classes. Spatial relationships between neighbouring pixels within a given window and pattern regularity will be different between geophysical and non-geophysical events. As observed, the local earthquake example can be seen as an abrupt and straight line pattern while airgun shots contain multiple such lines but at a smaller gradient value. Generally, it is expected that patterns from geophysical sources will be of more regular shape, while non-geophysical events can be less structured. Additionally, colour distribution and variation can be explored as a potential feature: teleseisms and tremors seem to cause higher disturbances on the cable at its far-end from the coast (possibly due to a better mechanical coupling between the seabed and cable at those locations).

As an overview, the diagram in Fig. 9 shows the main processes involved in earthquake catalogue creation using DAS. As shown, a DAS system interrogates the fibre optic cable, generating N_f strain waveforms (i.e. at different locations) each having M_f number of samples. A 2-D DAS matrix X is constructed and the straingram is generated using the signal and image processing steps discussed

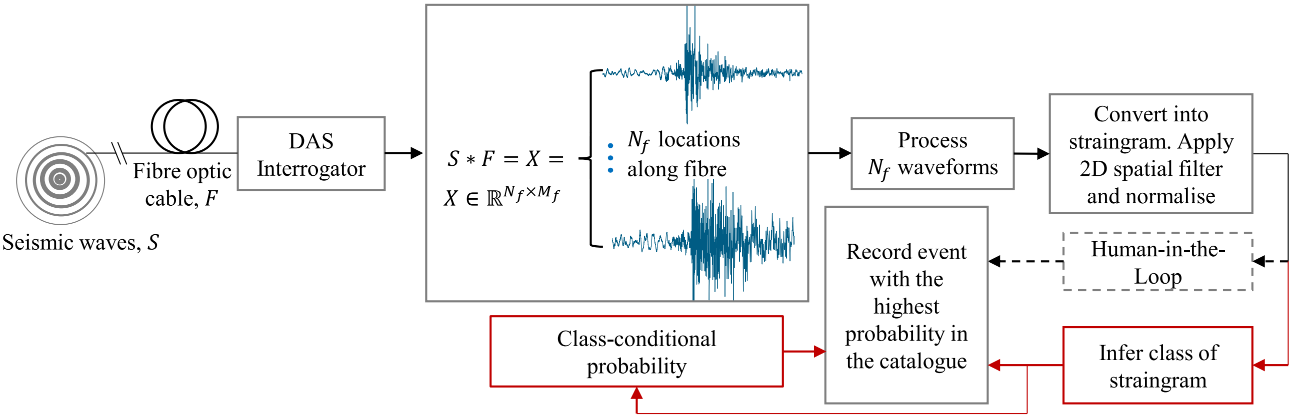


Figure 9. Diagram illustrating the main processing steps undertaken to classify raw DAS signals into one of the three classes: geophysical, non-geophysical or noise. The classifier also gives a class-conditional probability output.

previously in Section 2. In this study, the classifier is used to infer the class that each constructed straingram belongs to. The classifier can therefore replace the human expert's manual classification procedure. Additionally, the classifier will produce prediction outputs with a class-conditional probability associated with each image. This prediction output (or score) can be combined with a suitable threshold to provide a more informed classification of the straingrams.

3.1 Fundamental aspects of the classifier

The spatial hierarchical features on the straingrams, that is from edges to patterns, for each of the three classes, were identified using convolutional neural networks. Generally, the development of data-driven models using neural network architectures allows one to be very flexible and versatile in the training phase. Given a suitable architecture and hyperparameters or through transfer learning, neural networks can deal with various challenges including modelling with insufficient training examples (Goodfellow *et al.* 2016). Furthermore, neural networks using 2-D convolutional layers (hereafter 'convolutional layers') in order to automatically extract image features for classification offer several advantages over their predecessors (more details in LeCun *et al.* 2015, 1998; Krizhevsky *et al.* 2017) including:

(i) *Weight sharing*: the same set of parameters that are subjected to optimization in the training phase (hereafter 'free weights') are used across the whole image. Whereas, in conventional neural networks each layer is connected to its adjacent one using a free weight that is only used once in the network.

(ii) *Hierarchical feature learning*: convolutional neural networks use several layers, whereby its free weights are used to extract useful image features for a given task. These features are then used in the next layer to carry on an equivalent operation, thereby extracting features on features. By using deeper feature hierarchies, these models are capable of learning whole image objects, starting from simple edge detection at the higher network levels.

(iii) *Translation invariance*: in an image a feature (e.g. an edge) is guaranteed to be identified no matter its spatial location, since the same kernel is used throughout its full spatial length. For our study, an event such as a local earthquake in a straingram may be shifted in both time and distance. In order to discriminate it from the other two classes, the translation invariance property is necessary.

In convolutional neural networks for images, each set of parameters will be a 2-D square matrix of relatively small size, for example 3×3 or 7×7. This matrix is convolved with an image by sliding it both horizontally and vertically in order to extract features suitable for classification. This matrix is known as the 'kernel', while the corresponding set of features extracted from a single convolution operation is called a 'feature map'. In each layer of the network, different kernels will be used so that multiple feature maps (as the number of kernels used) can be extracted from a single image. This feature map will then be used in the next layer to perform the same convolution operations (hierarchical structure), and so on. A key difference with other classification techniques, for example support vector machines, for input image data (or other spatially correlated data), is that convolutional neural networks compute a set of unique features relevant to the task. All kernel matrices that are used in the convolutional layers are optimized in the training phase so that features are synthesized instead of handcrafted.

For completeness, some of the most important computational processes involved in convolutional neural networks and their associated technologies will also be discussed more formally in this section. At the same time, the diagram in Fig. 10 helps to visualize these processes at the inference (or prediction) stage for each straingram, as drawn from the test set. In the diagram, a straingram representing a local earthquake of pixel size 256×256 is convolved with 112 kernels, where each kernel is a different 5×5 matrix (parameter values range from 0 to 100). Therefore, 112 feature maps are generated after this first layer of the network. Three feature maps are shown in Fig. 10(c) (out of 112), while the corresponding kernels are seen in Fig. 10(b). Most commonly, there are three different computational processes involved in the extraction of feature maps: 2-D linear convolution with a kernel matrix K, a non-linear activation function $f_a()$ of each pixel in the feature map and 2-D spatial downsampling (although in the example in Fig. 10 there is no spatial downsampling after the first convolutional layer).

During a 2-D convolution operation with a $p \times p$ kernel matrix K and an input straingram $X \in \mathbb{R}^{I \times J \times 1}$ (second-order tensor), a feature map Z will be computed at various spatial coordinates i, j

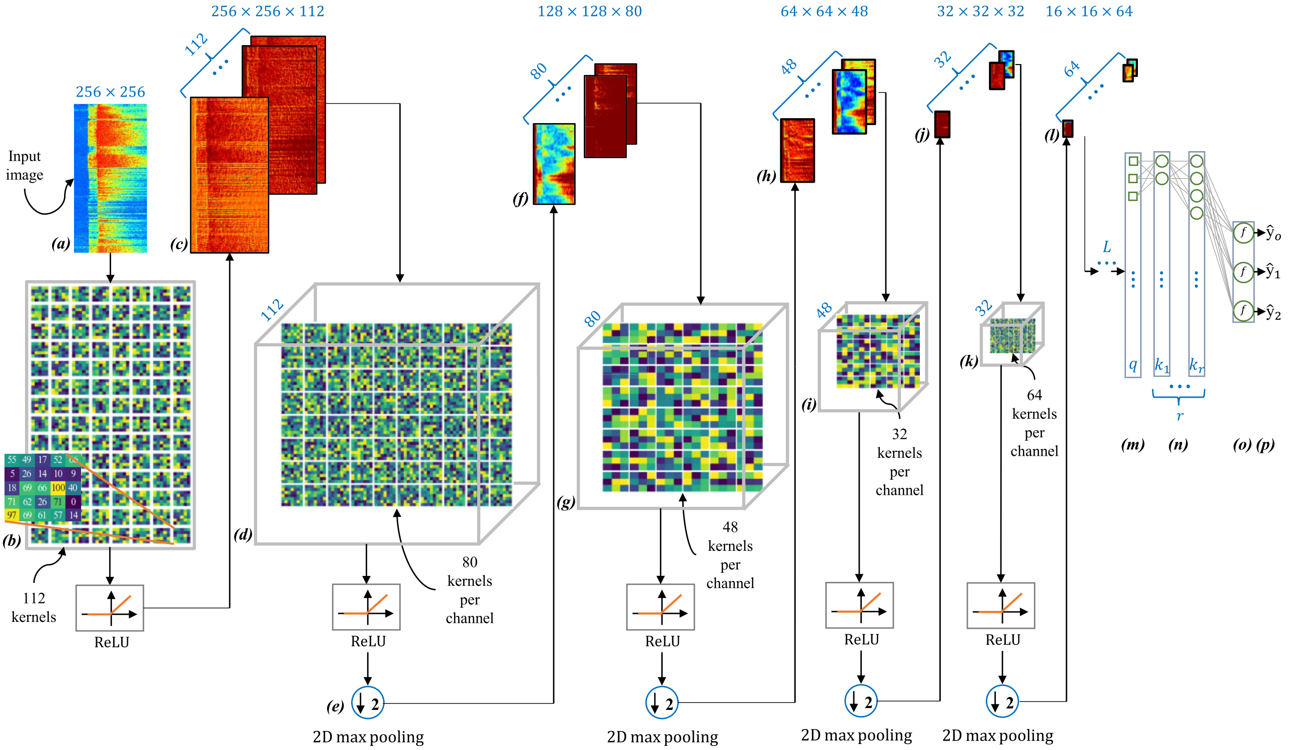


Figure 10. Diagram illustrating the hierarchical process of generating prediction outputs from straingrams using a trained 2D convolutional neural network.

as follows:

$$Z[i, j] = \sum_{a=0}^{p-1} \sum_{b=0}^{p-1} X[i + a, j + b]K[a, b]. \qquad (1)$$

The operation in eq. (1) will generate a single feature map. In practice, there will be several such kernel matrices, and as such the operation will result in sets of feature maps after each convolutional layer. Note that all feature maps generated will jointly contribute to the classification accuracy, and so one cannot consider them as separate features. For instance, corners in one feature map and contrast levels in another are combined in a feature vector, which is then used to identify each class. This is the result of simultaneously optimizing all kernel matrices together to extract features that may be shared across the feature maps, as well (Zhang *et al.* 2021). And, this is one of the reasons that convolutional neural networks are highly successful in image processing tasks, like classification. Also, from Fig. 10(d) and onwards, each convolutional layer will comprise C number of channels, which is equal to the number of feature maps computed in the previous convolution layer, and so on. Therefore, we will be looking to expand our single-input and single-output computation in eq. (1) to one that incorporates multiple inputs and multiple outputs in each convolutional layer. For this example, each one of the 112 channels in the second convolutional layer will be cross-correlated with each feature map (equal to 112), and the result at various spatial coordinates i, j will be summed to generate a single feature map as follows:

$$Z[i, j] = \sum_{c=1}^{C=112} \sum_{a=0}^{p-1} \sum_{b=0}^{p-1} Y[i + a, j + b, c]K[a, b, c]. \qquad (2)$$

Note that in the second convolutional layer, there are 80 kernels per channel, and so there will be 80 feature maps generated, as seen in Fig. 10(f). Before generating the second set of feature maps there are two more computational processes after each convolutional layer that must be discussed: non-linear activation and 2-D spatial downsampling, as mentioned previously.

The purpose of applying any type of activation function on the feature maps, for example $f_a(Z[i, j] = \max\{0, Z[i, j]\}, \ \forall\, i, j$, is to induce flexibility during the training phase. This means that the model can extract features from the straingrams that are not only complex but also non-linearly correlated (i.e. in the case of non-linear activation functions). This broad-band capability is not possible with just linear convolution operations. Rectified Linear Unit (ReLU) is an example of a non-linear activation function that has already demonstrated a number of benefits in learning tasks, for example the introduction of a sparse set of weights (since it sets values to 0) and avoidance of the vanishing gradient seen in other activations (e.g. sigmoid, Krizhevsky *et al.* 2017; Goodfellow *et al.* 2016; LeCun *et al.* 2015). On the other hand, using 2-D spatial downsampling layers one can effectively reduce the size of the input image by sliding a fixed-size spatial window over the image and extracting suitable information. Following this procedure, coarser feature maps can be calculated after each convolutional layer, which will result in the recognition of more global features. This ultimately yields in recognizing whole objects in the image, for example the straingram pattern of the local earthquake (Fig. 7a). Due to downsampling, small translations and transformations in the input image will not influence the computation of feature maps. This is another important consideration in classifying the straingrams, since small shifts in the pattern for each event will be effectively filtered out. As an example, in Fig. 10(e) 2-D maximum pooling with a window size of 2 × 2 is used to obtain the maximum pixel value within that window. This operation exactly halves each feature map, that is after the first 2-D maximum pooling: $f_a(Z) \in \mathbb{R}^{I\times J} \rightarrow \mathbb{R}^{I/2\times J/2}$, and so on. Therefore, in the last extracted feature map (Fig. 10l) the pixel size of each straingram will be decreased by 2^{N_p}, where N_p is the total number of 2-D maximum pooling layers used. Hence, no matter the pixel size of each straingram and given sufficient computational resources, one may use any input image size to train the convolutional neural network, as the 'field-of-view' of the classifier is taken care of by the downsampling layers. After the final convolutional layers, higher-level features (e.g. textures) are extracted by combining more localized ones in the earlier layers (e.g. edges).

Given the relatively small set of labelled straingrams at our disposal (i.e. 1050 for training-validation-testing phases), two regularization techniques were investigated. The first involves the addition of dropout layers after each 2-D maximum pooling layer, which was inspired by other neural network architectures as seen in Srivastava *et al.* (2014), Zeiler & Fergus (2013) and Wu & Gu (2015). The process of dropout layers enables each free weight to be self-sufficient during the training phase of the classifier, such that it prevents 'co-adaptation' behaviour (Hinton *et al.* 2012). Essentially, a fraction (or the dropout rate) of free weights is randomly set to zero at the beginning of each epoch. More formally, an independently sampled Bernoulli variable associated with one of the free weights is set to 0 with probability p_d ($\beta \sim$ Bernoulli (p_d)). This random variable β is multiplied by one of the free weights so that it eliminates its contribution during the training phase when $\beta = 0$. In order to combat potential overfitting due to small-sized data sets, image augmentation is another approach that was used (see Shorten *et al.* (2019)). Using that approach, the data set is expanded with the addition of synthetically generated copies of the straingrams by applying different types of label-preserving affine transformations as shown in previous studies, for example Krizhevsky *et al.* (2017) and Perez & Wang (2017). Both dropout and image augmentation layers do not have free weights that need to be optimized during the training phase.

Furthermore, in Fig. 10(n) the convolutional neural network incorporates r fully connected dense layers. In k_1 layer, each neuron in that layer is connected to each feature value extracted from the last convolutional layer (Fig. 10l). This is done by first vectorizing the tensor to 1-D (Fig. 10m). Given that there are a total of 64 feature maps in the last convolutional layer, each of size 16×16, there will be a total of $k_1 = 16 \times 16 \times 64$ free weights that need to be trained in the first fully connected dense layer (see LeCun *et al.* 2015, for more details).

For training this type of classifier when the number of classes is greater than 2 the categorical cross-entropy loss function is a typical choice (Goodfellow *et al.* 2016; Zhang *et al.* 2021). Among other things, cross-entropy loss avoids issues such as the 'learning slowdown' that characterizes quadratic losses. Given t_{n_t,n_c} and $\hat{y}_{n_t,n_c}$ represent the target and predicted classes, respectively, so that for each n_t^{th} training straingram there will be a binary vector associated with the target n_c^{th} class. For instance, for the n_t^{th} straingram the true class $t_{n_t,n_c} = [1, 0, 0]$, which can correspond to the geophysical class, while $t_{n_t,n_c} = [0, 0, 0]$ for both non-geophysical and noise classes. Then, the average categorical cross-entropy loss overall training straingrams N_t, and the number of classes N_c is,

$$J = -\sum_{n_t=1}^{N_t}\sum_{n_c=0}^{N_c}\left(t_{n_t,n_c}\ln\left[\hat{y}_{n_t,n_c}\right]\right). \tag{3}$$

Note that $J \to 0$ when target and predicted classes are similar, such that the cross-entropy loss is minimized over all N_t in the data set. For each straingram during inference, an output score vector is computed $\hat{\mathbf{y}}$ (Fig. 10p), which can be converted into a class-conditional probability for a more meaningful interpretation of the results. This is done using a softmax activation function (Goodfellow *et al.* 2016), which is used after the last fully connected dense layer, as in the diagram in Fig. 10(o).

Batch gradient descent is a technique that is used to minimize J by adjusting all the free weights of the classifier simultaneously by using the backpropagation algorithm (Murphy 2022). Free weights include all $p \times p$ weights to parametrize each kernel K for feature map extraction in all convolutional layers and all the weights that are used to connect each fully connected dense layer to its adjacent one. For instance, in layer k_1 a typical number of neurons will be 256 each connected to another 256 neurons in layer k_2, and so on. While in the first convolutional layer with each K being 5×5, there will be $5 \times 5 \times 112$ free weights to optimize. On the other hand, in the second convolutional layer, there is a 4-D tensor of free weights with a size of $7 \times 7 \times 112 \times 80$.

In dealing with large-sets of data, a mini-batch gradient descent approach is used. This is because, in batch gradient descent J is evaluated only once for all N_t training data, as seen in eq. (3). This in turn leads to a single update for each free weight used in the training phase of the classifier (also known as one 'epoch'). However, in mini-batch gradient descent, J is evaluated multiple times for each epoch, which yields a more efficient estimation of the loss (is the average of all mini-batches), albeit more noisy. In Masters & Luschi (2018), it was shown that small mini-batch sizes, i.e., less than 32 instances, improves generalization performance. This finding was also supported by the findings in Keskar *et al.* (2016), showing that large mini-batches tend to converge to sharp minimizers.

Note that, in our study each straingram is assigned to each mini-batch at random, after each mini-batch evaluation (completion of one epoch). The Adaptive Moment Estimation Algorithm (ADAM, Kingma & Ba 2014) is a common technique that implements mini-batch gradient descent in an efficient and robust way. ADAM combines the capabilities of RMSProp (Zhang *et al.* 2021) and AdaGrad (Duchi *et al.* 2011), where in these algorithms a decaying exponential gradient is used to adapt the learning rate.

3.2 Tuning of classifier's main hyperparameters and architecture

Hyperparameters in convolutional neural networks refer to any parameter that is not part of the gradient descent updating rule, that is all parameters that are not free weights and need to be specified prior to training. Table 2 lists the search space of values for 14 different hyperparameters that were considered. In order to identify a suitable set of hyperparameters, given the relatively large search space shown in Table 2, Bayesian Optimization is a popular choice. The objective function g used in Bayesian Optimization is the loss computed using the validation set of 150 straingrams, and varies with hyperparameter changes $g(\mathbf{h}_p)$. Using Gaussian Processes Regression (GPR, Murphy 2022) one can construct a surrogate model for g. More importantly, the GPR framework in the context of Bayesian Optimization, is used to evaluate $g(\mathbf{h}_p) \in \mathbb{R}^d$ by searching for points in this $d-$dimensional space that provide a trade-off between exploration and exploitation of the surrogate model's posterior distribution. This is done using an acquisition function that balances exploiting regions where the posterior of $g(\mathbf{h}_p)$ is high and exploring regions where the uncertainty in predicting $g(\mathbf{h}_p)$ is high.

Table 2. Search space of the main hyperparameters used to optimise the 2D convolutional neural network.

Hyperparameters	Search space
Input image pixel size	$\{128 \times 128, 224 \times 224, 256 \times 256, 312 \times 312\}$
Number of convolutional layers	$\{3, 4, 5, 6, 7, 8\}$
Dimension $p \times p$ of convolutional kernel	$\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9\}$
Activation function $f()$ binary choice	{ReLU, Sigmoid}
2-D 2×2 maximum pooling binary choice	{True, False}
Number of maximum pooling layers	$\{1, 2, 3, 4\}$
Dropout layer binary choice	{True, False}
Image augmentation layer binary choice	{True, False}
Brightness and contrast rates $(0 - 1)$	$\{0, 0.01, ..., 0.19, 0.2\}$
Vertical flip binary choice	{True, False}
Dropout rate $(0 - 1)$	$\{0, 0.01, ..., 0.39, 0.4\}$
Number of neurons in fully connected dense layers	$2^i, \forall i = \{6, 7, 8, 9, 10\}$
Learning rate η	$2^{-i}, \forall i = \{6, 7, ..., 11, 12\}$
Number of straingrams used in each mini-batch	$2^i, \forall i = \{3, 4, 5, 6, 7\}$

Table 3. Bayes model architecture, hyperparameters and output tensor shape from each network layer.

Layer type	Value chosen	Tensor output shape
Input layer	$256 \times 256 \times 1$	$256 \times 256 \times 1$
2-D convolution | activation	Kernel size $= 5 \times 5 \,\|\, ReLU$	$256 \times 256 \times 112$
Dropout	Dropout rate $= 0.245$	$256 \times 256 \times 112$
2-D convolution | activation	Kernel size $= 7 \times 7 \,\|\, ReLU$	$256 \times 256 \times 80$
2-D maximum pooling	Kernel size $= 2 \times 2$	$128 \times 128 \times 80$
Dropout	Dropout rate $= 0.255$	$128 \times 128 \times 80$
2-D convolution | activation	Kernel size $= 3 \times 3 \,\|\, ReLU$	$128 \times 128 \times 48$
2-D maximum pooling	Kernel size $= 2 \times 2$	$64 \times 64 \times 48$
Dropout	Dropout rate $= 0.225$	$64 \times 64 \times 48$
2-D convolution | activation	Kernel size $= 3 \times 3 \,\|\, ReLU$	$64 \times 64 \times 32$
2-D maximum pooling	Kernel size $= 2 \times 2$	$32 \times 32 \times 32$
Dropout	Dropout rate $= 0.02$	$32 \times 32 \times 32$
2-D convolution | activation	Kernel size $= 5 \times 5 \,\|\, ReLU$	$32 \times 32 \times 64$
2-D maximum pooling	Kernel size $= 2 \times 2$	$16 \times 16 \times 64$
Dropout	Dropout rate $= 0.255$	$16 \times 16 \times 64$
2-D convolution | activation	Kernel size $= 7 \times 7 \,\|\, ReLU$	$16 \times 16 \times 48$
Dropout	Dropout rate $= 0.235$	$16 \times 16 \times 48$
Flatten	Inherited from previous layer	$12288 \times 1 \times 1$
Fully connected dense | activation	Number of neurons $= 256 \,\|\, ReLU$	$256 \times 1 \times 1$
Fully connected dense | activation	Number of neurons $= 3 \,\|\, Softmax$	$3 \times 1 \times 1$

It is therefore a 'guided' procedure, and one that minimizes the requirement for obtaining actual observations of the validation set loss, which are expensive to obtain for a range of different hyperparameters. In this study, the upper confidence bound is used as an acquisition function. See Snoek *et al.* (2012), Shahriari *et al.* (2015) and Brochu *et al.* (2010) for more details on Bayesian Optimization using Gaussian Processes.

To implement hyperparameter tuning and training-validation-testing of the classifier Keras (O'Malley *et al.* 2015) and TensorFlow (Abadi *et al.* 2016) libraries were used, respectively. Python version 3.10 was used on Ubuntu 22.04 on a PC with 32GB of RAM and an Nvidia V100.

4 RESULTS

An image resizing layer was included as part of the classifier's architecture so that each greyscale straingram was resized to a fixed-size image of 256 × 256. Additionally, a mini-batch of size 32 was also chosen via the Bayesian Optimization routine used. In order to further constraint the optimization procedure when tuning the network architecture, 'blocks' of layers (further limiting the total number of blocks of layers between 3 and 8) were considered in the following sequence: convolutional, 2-D maximum pooling and dropout. The optimization procedure was 'free' to choose between using 2-D maximum pooling and/or dropout or just convolutional layer in every block of layers, while at the same time respecting the constraints set in Table 2.

In order to visualize the final classifier architecture, Table 3 shows such an optimal choice of hyperparameters that was derived using Bayesian Optimization. An important feature of this particular model architecture (hereafter 'Bayes') is the extensive use of dropout layers, which are included after every convolutional layer. In 5 out of 6 'blocks' (2-D convolution-2-D maximum pooling-dropout), the dropout rate is more than 20 per cent of the total number of free weights used in each convolutional layer. At the same time, only ReLU activation functions (no Sigmoid functions) were selected by the optimization routine. A total of 50 trials was specified in each optimization run, which yielded 50 classifiers of varying validation losses, having a range of different hyperparameters and architectures. Many such optimization runs were performed to further evaluate our results and to ensure robustness in the solu-

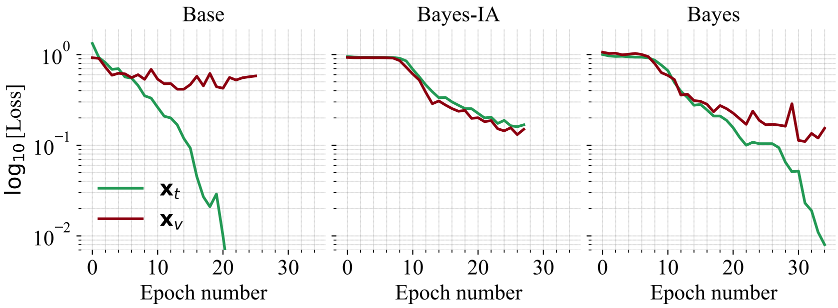


Figure 11. Categorical cross-entropy loss (log-scale) with epoch number for the three classifiers compared in this study. The two overlapping lines correspond to the loss calculated on the training and validation sets, $\mathbf{x}_t$ and $\mathbf{x}_v$, respectively.

tions.

The main features of two of the most promising classifiers, as derived using Bayesian Optimization (chosen as the top classifiers from two different optimization runs) are as follows:

(i) *Bayes*: As shown in Table 3, Bayes is comprised of 6 convolutional layers with 4 2-D maximum pooling layers after each convolutional operation (starting from layer 2 to 6). It has a single fully connected dense layer with 256 neurons, and the total number of free weights being equal to 3 839 091.

(ii) *Bayes-IA*: This classifier uses two image augmentation layers. One is a random vertical flip and the other applies random contrast with a maximum level of 10 per cent. The model is comprised of six 2-D convolution layers with four 2-D maximum pooling operations being followed by dropout layers in each one (similar blocks of layers as in Bayes model). This classifier has a fully connected dense layer with 640 neurons, and the total number of free weights is equal to 664 531. The reason that this model has a smaller number of free weights, in comparison to the other two models, is due to the fact that it incorporates fewer kernels in each convolutional layer, so fewer feature maps are generated and so on.

To facilitate comparison, a baseline classifier (hereafter 'Base') was also trained using hyperparameters and model architecture that were derived heuristically. The base has only three convolutional layers and incorporates 2-D maximum pooling after each one. Additionally, it has a fully connected dense layer with 128 neurons, yielding a total number of free weights equal to 4 239 267. Hence, Base has the highest number of free weights of the three classifiers considered, which is partly attributed to less spatial downsampling of its feature maps (only three 2-D maximum pooling layers). Finally, the learning rate of Base was manually chosen to be 0.003, while for Bayes and Bayes-IA it was selected by the optimization routine to be equal to, 0.0005 and 0.0002, respectively.

In Fig. 11 the variation in categorical cross-entropy loss as a function of epoch number is shown for the three classifiers. Note that the two overlaid lines correspond to the losses computed using both the training $\mathbf{x}_t$ and validation $\mathbf{x}_v$ sets. All images assigned to training and validation sets were randomly selected (stratified random sampling) prior to each training phase from the data set of 1050 straingrams. As shown in these plots, all three models were trained on different numbers of epochs, that is 32 for Bayes, 27 for Bayes-IA and 25 for Base. This is because an early stopping criterion was used during the training phase which reduces the chances of model overfitting. This is done by terminating the training phase after the loss function on the validation set has seen no particular decrease (a threshold is typically set at a validation loss of around 0.01) for five consecutive epochs, starting from epoch number 10. After early stopping the model with the lowest loss function on $\mathbf{x}_v$ is selected. Note that the specification for the early stopping criterion was done empirically, but, was the same for the three models.

The gradients of the losses in Bayes and Bayes-IA on both $\mathbf{x}_t$ and $\mathbf{x}_v$ data sets, increase dramatically starting from epoch number 8 and onwards. In Bayes-IA, both of the losses decrease at approximately the same (slow) rate until the training phase is terminated at epoch number 27 due to the early stopping criterion being satisfied. On the other hand, in Bayes the loss continues to decrease fast on $\mathbf{x}_t$, while after about epoch number 27, the decrease in loss on $\mathbf{x}_v$ is much smaller, resulting in early stopping at epoch number 34. Note that Bayes shows a large deviation between validation and training losses after epoch number 20, but, it does not overfit the model due to the extensive use of dropout layers. Overall, Bayes obtains a higher decrease in loss on $\mathbf{x}_v$, as compared to Bayes-IA. The addition of image augmentation layers to synthesize a larger straingram data set for this particular classification task results in a slower convergence of the loss function. Moreover, the heuristically tuned hyperparameters and model architecture of the convolutional neural network, that is the Base model, proved to be insufficient for this classification task. As can be seen, the loss function on $\mathbf{x}_t$ keeps decreasing dramatically (falls below the minimum value of the *y*-scale axis) since the beginning of the training phase. At the same time, the loss on $\mathbf{x}_v$ does not decrease, which can be a sign of model overfitting.

To further evaluate the three classifiers, in terms of accuracy and robustness in predicting the correct labels for each straingram, Fig. 12 shows both the cross-entropy loss and classification accuracy on $\mathbf{x}_v$. As seen, Bayes achieves the highest classification accuracy of around 97 per cent after about epoch number 30. Small discrepancies between loss and accuracy values can be observed, which is a consequence of computing classification performance using two slightly different measures. On the one hand, the loss function calculation provides a direct probability measure (between 0 and 1) regarding the divergence of predictions from the target class distribution (Goodfellow *et al.* 2016). While classification accuracy on $\mathbf{x}_v$ is a simplistic way of examining model performance through an intuitive ratio: the amount of 'correct' predictions (given a suitable threshold, e.g. 0.5) to the total number of images in the validation set. An image may be predicted as belonging to one of the three classes, even if the probability is low (i.e. even when it is close to the threshold of 0.5). Although this will increase the number of 'correct' predictions, the loss function may see no particular decrease in its value when the confidence is relatively insignificant.

Nevertheless, looking at both the loss and accuracy plots for Bayes-IA, one notices that after about epoch number 20 this classifier reaches an overall improved performance (around 95 per cent in

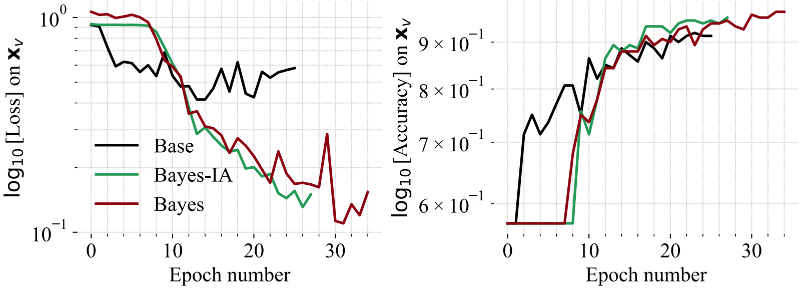
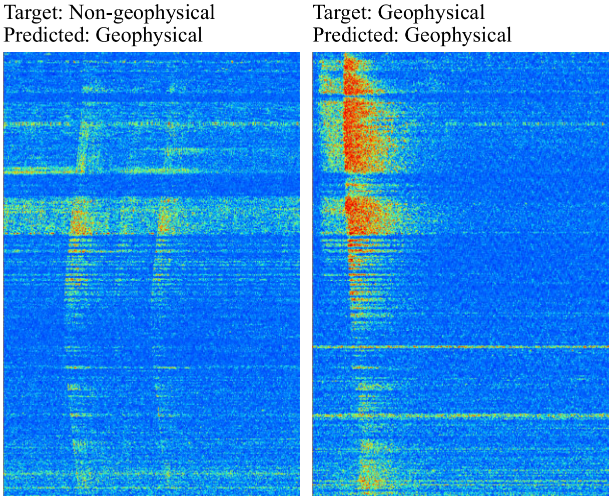


Figure 12. Categorical cross-entropy loss function (left-hand panel) and classification accuracy (right-hand panel) as a function of epoch number on the validation set. The overlapping lines correspond to the three classifiers compared in this study.

Table 4. Confusion matrices of three models: Base, Bayes-IA and Bayes. Values are presented as percentages (e.g., 95 represents 95 per cent).

Target class	Predicted class		
	Geophysical	Noise	Non-geophysical
BASE			
Geophysical	95	0	5
Noise	0	79	21
Non-geophysical	27	0	73
BAYES-IA			
Geophysical	100	0	0
Noise	4	80	16
Non-geophysical	24	6	70
BAYES			
Geophysical	100	0	0
Noise	0	96	4
Non-geophysical	17	3	80


Target: Non-geophysical
Predicted: Geophysical

Target: Geophysical
Predicted: Geophysical

Figure 13. Two straingram examples drawn from the test set with target and predicted class labels as computed by Bayes model.

accuracy). Due to the image augmentation layers, Bayes-IA results in a more stable minimization of the loss function, since the training data set is now expanded (essentially increasing the ratio of training examples to the total number of free weights). On the other hand, looking at the accuracy of the Base model, its performance can be seen to be comparable to the other two classifiers. As discussed in the previous paragraph, accuracy on its own can be a misleading indicator of performance. In the case of Base, the model could not achieve a decrease in its loss function on the validation set, starting from epoch number 5 and onwards. Although the Base model has a high number of 'correct' predictions (around 91 per cent) its confidence remains relatively low, given that the loss is relatively high. For instance, if the threshold is changed to a higher one (e.g. 80 per cent) to consider greater confidence in the results, then Base will have a very low classification accuracy. This is a sign that this model did not 'learn' adequately the feature patterns required for the correct classification of the classes.

The performance of the three classifiers was further evaluated on a test set of 100 images. This data set was generated as part of the overall stratified sampling procedure that was followed to separate the data set into three sets: training, validation and testing. The confusion matrices for the three models are shown in Table 4, where on each row is the target class and on each column the predicted one. As can be seen, all three classifiers accurately predict the labels for the geophysical straingrams, with Base scoring 95 per cent and the other two 100 per cent. Bayes is the most accurate classifier in predicting the correct target class for both noise and non-geophysical straingrams, with 96 per cent and 80 per cent, respectively. On the other hand, Bayes-IA is marginally more accurate than Base in these two classes. The overall accuracy (considering all the classes) of Bayes-IA is 83 per cent, which is just 1 per cent higher than Base's overall accuracy. On the other hand, the corresponding classification accuracy for Bayes is about 92 per cent on this test set. Predicting the correct target class for the straingrams from the non-geophysical class resulted in the largest number of errors. The top-performing classifier, that is Bayes, only managed to predict 80 per cent of the non-geophysical straingrams correctly.

5 DISCUSSION

Besides the relatively low classification score in predicting non-geophysical straingrams correctly (80 per cent), Bayes is relatively accurate with 92 per cent overall accuracy on our test set. Most importantly, Bayes is fairly accurate in discriminating geophysical straingrams from the other two classes, which means that the respective features and patterns were 'learned' by the model. Although this is a relatively small-scale test data set, as more straingrams become available we will be more confident in our model's generalization capability.

In terms of model architecture, Bayes is comprised of 'blocks', that is convolution-2-D maximum pooling-dropout sequences, as seen in Table 3. Such architecture was inspired by previous models, for example the 'VGG-types' developed by the Virtual Geometry Group at Oxford (see Simonyan & Zisserman 2014) and others used in earthquake detection by Huot *et al.* (2022b). Note that 'VGG-type' models have proven very successful in image classification and can be readily used using TensorFlow. However, such models require 1000s of training images, since the number of free weights is greater than 138 million. Given our limited data set of 1050 straingrams, this will lead to an overdetermined model. Instead, the proposed classifier is relatively simple, requiring the optimization of less than 4 million free weights. As a consequence, training Bayes is also very computationally inexpensive, and can be done on a CPU (≥ 4 cores with >3 GHz in clock-speed) or with a mid-range consumer GPU with CUDA cores (e.g. Nvidia RTX 30-series).

Bayesian Optimization was critical in designing our Bayes model architecture and tuning its hyperparameters. As seen, the optimization routine chose small values for η on the gradient descent updating rule. Additionally, the use of dropout layers with a relatively moderate rate, with around 1/5 of free weights zeroed in each convolutional layer, offered robustness capability for our small-scale straingram data set. At the same time, ReLU activations were also chosen, which supports the fact that they are widely used in current state-of-the-art neural network architectures. Although image augmentation, that is the Bayes-IA model, introduces robustness during the model training phase (as it reduces the possibility of overfitting), the slow convergence rate of the loss function is undesirable. Therefore, image transformations such as random vertical flips and contrast level changes applied to the original straingrams may not be a particularly successful regularization approach for our classification task. This is due to the fact that geophysical patterns are relatively regular and well-defined, that is body waves obey physical laws, and therefore, these patterns should not be artificially altered.

In Fig. 13, we show an example of predicting the correct target class and another incorrect prediction by Bayes. In particular, we see that the geophysical straingram (on the right) was correctly predicted, while the non-geophysical was not. Given that the latter is composed of patterns that resemble a local earthquake, spanning a large proportion of the straingram, Bayes's confidence is around 70 per cent in this example image. A portion of geophysical straingrams on our data set have such similar patterns, and this can explain the reason as to why our classifiers wrongly predicted non-geophysical straingrams as geophysical ones. Local earthquakes are generally more well-represented in our data set as compared to this small proportion of non-geophysical straingrams, which results in better 'learning' of the geophysical patterns.

To tackle such issues, one may adopt a higher classification threshold, for example 80 per cent in order to ensure better decision-making in a practical context. An important limitation of our classification framework is the reliance on correctly annotated straingrams that are required to train our models. Given a small fraction of wrongly annotated straingrams, can result in a significant degradation of classification accuracy. In the literature, self-supervised learning is a well-known technique to generate pseudo-labels directly from the structure of straingrams (e.g. Jing & Tian 2020). Semi-supervision has also been used (e.g. Yalniz *et al.* 2019) to improve performance in convolutional neural networks, by exploring a much smaller set of well-annotated images. In fibre optic seismology this is particularly important as there is an abundance of data sets, but annotations are relatively limited (Thrastarson *et al.* 2022).

One of the limitations associated with convolutional neural network-based architectures is the fact that these models can be insensitive to very localized patterns. From the non-geophysical straingram in Fig. 13, angle changes on both of the lines may have been completely disregarded during the training phase, even though this could serve as an important discriminating feature between the two classes (geophysical straingrams do not have such angle changes). Using spatial downsampling layers a certain degree of geometric transformation invariance is introduced, since they summarize a window in the image, for example to extract the maximum or average pixel values. In Sabour *et al.* (2017), a new class of models (known as Capsule Networks) was introduced that can preserve such localized spatial features.

Finally, another important consideration that needs to be discussed has to do with how well each geophysical event is represented by its corresponding straingram. Or in other words, the proportion of seismic waves being recorded by a geophysical event. Given that straingrams are constructed from a continuous stream of DAS signals, there is a high possibility of partially capturing an event, that is creating an 'incomplete' snapshot. In case an event is 'sufficiently' represented by the straingram (all other variables considered relatively constant), then the classifier is expected to be as accurate as the one presented in this study. Other other hand, if an event is not 'sufficiently' represented, for example body waves partially captured at the beginning of the fibre optic cable, then the classifier may misclassify this event. This is because the classifier has been trained with straingrams that fully (or almost fully) captured the seismic waves of each event. Therefore, any new straingram must contain a 'sufficient' proportion of seismic waves for the feature maps to be effective in classifying each class.

Having said that, it should also be noted that convolutional neural networks are invariant to geometric transformations (e.g. rotations) in the images and translations (e.g. temporal shifts), so that straingrams are not restricted to show an event on a certain part of the image. In other words, no matter the spatial location of an event on the straingram (top-left, bottom-right, etc.) the relevant feature maps required for classification will be extracted by convolutional neural networks (i.e. given the appropriate architecture and hyperparameters).

6 CONCLUSIONS

In this paper, we tackled the problem of processing large-scale observations from ocean-bottom fibre optic cables, being interrogated by DAS. In particular, we developed a digital signal and image processing method that converts a highly dense array of DAS strain data measured on a 55-km fibre optic cable into straingrams (plots of strain field variation as a function of time and location along the cable). Using these representations we have trained 2-D convolutional neural networks to classify straingrams representing seismic observations from other sources of hydroacoustic energy, for example sailing boats and marine mammal movements. The non-parametric Bayesian framework of Gaussian Processes Regression was used to tune the classifier's hyperparameters and particular architecture. We have seen that the optimization routine has systematically selected dropout layers in order to tackle the issue of overfitting, given our relatively small data set of straingrams. Two classifiers were compared, one using dropout layers and another using image augmentation as regularization approaches. These two classifiers were compared with a baseline model, whose hyperparameters and architecture were chosen heuristically. On a test set of around 100 straingrams, the classifier with the dropout layers proved to be the most promising for our classification task, having an overall accuracy of 92 per cent. Some limitations of the current classifier were also discussed, for example invariance to highly localized features, along with some suggestions/methods for future work.

Overall, we showed that the presented classifier can be used to significantly assist the seismologist, in analysing large-scale data sets using DAS technology. Moreover, as more data are generated from future experiments, we aim to develop a truly powerful image classifier, one that can be used to automate the process of discriminating between different types of seismic processes, for example tectonic tremors from teleseismic events.

DATA AVAILABILITY

We make publicly available the full dataset that was used to train, validate and test our models in the University of Southampton Institutional Repository (Matthaiou *et al.* 2023).

REFERENCES

Abadi, M. *et al.*, 2016. Tensorflow: a system for large-scale machine learning, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16),* pp. 265–283.

Agostinetti, N.P., Villa, A. & Saccorotti, G., 2022. Distributed acoustic sensing as a tool for subsurface mapping and seismic event monitoring: a proof of concept, *Solid Earth,* **13,** 449–468.

Ajo-Franklin, J.B. *et al.*, 2019. Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection, *Sci. Rep.,* **9,** 1328, doi:10.1038/s41598-018-36675-8.

Baba, S., Araki, E., Yamamoto, Y., Hori, T., Fujie, G., Nakamura, Y., Yokobiki, T. & Matsumoto, H., 2023. Observation of shallow slow earthquakes by distributed acoustic sensing using offshore fiber-optic cable in the Nankai Trough, Southwest Japan, *Geophys. Res. Lett.,* **50,** e2022GL102678, doi:10.1029/2022GL102678.

Biondi, B.L., Yuan, S., Martin, E.R., Huot, F. & Clapp, R.G., 2021. Using telecommunication fiber infrastructure for earthquake monitoring and near-surface characterization, *Distributed Acoustic Sensing in Geophysics: Methods and Applications,* Vol. **268,** pp. 131–148, Li, Y., Karrenbach, M. & Ajo-Franklin, J. B., John Wiley & Sons.

Brochu, E., Cora, V.M. & De Freitas, N., 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning, arXiv:1012.2599.

Chen, S., Han, J., Sui, Q., Zhu, K., Lu, C. & Li, Z., 2023. Advanced signal processing in distributed acoustic sensors based on submarine cables for seismology applications, *J. Lightwave Technol.,* **41,** 4164–4175.

Cheng, F., Chi, B., Lindsey, N.J., Dawe, T.C. & Ajo-Franklin, J.B., 2021. Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization, *Sci. Rep.,* **11,** 5613, doi:10.1038/s41467-018-04790-9.

Cochran, E.S., 2018. To catch a quake, *Nat. Commun.,* **9,** 2508, doi:10.1038/s41467-018-04790-9.

Corera, I., Piñeiro, E., Navallas, J., Sagues, M. & Loayssa, A., 2023. Long-range traffic monitoring based on pulse-compression distributed acoustic sensing and advanced vehicle tracking and classification algorithm, *Sensors,* **23,** 3127, doi:10.3390/s23063127.

Duchi, J., Hazan, E. & Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.,* **12,** 2121–2159.

Fernández-Ruiz, M.R. *et al.*, 2022. Seismic monitoring with distributed acoustic sensing from the near-surface to the deep oceans, *J. Lightwave Technol.,* **40,** 1453–1463.

Gan, L., Wu, Q., Huang, Q. & Tang, R., 2023. Quality classification and inversion of receiver functions using convolutional neural network, *J. geophys. Int.,* **232,** 1833–1848.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning,* MIT Press.

Grassl, H., 2001. Climate and oceans, in *International Geophysics,* Vol. **77,** pp. 3–9, Academic Press.

Harmon, N. *et al.*, 2022. Surface deployment of DAS systems: coupling strategies and comparisons to geophone data, *Near Surf. Geophys.,* **20,** 465–477.

Hernández, P.D., Ramírez, J.A. & Soto, M.A., 2021. Deep-learning-based earthquake detection for fiber-optic distributed acoustic sensing, *J. Lightwave Technol.,* **40,** 2639–2650.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors, preprint, arXiv:1207.0580.

Huot, F., Ariel, L., Paige, G., Bin, L., Robert, G. C., Tamas, N., Kurt, T. N. & Biondo, L. B., 2022a. Detection and characterization of microseismic events from fiber-optic DAS data using deep learning, *Seismol. Soc. Am.,* **93,** 2543–2553.

Huot, F., Biondi, B.L. & Clapp, R.G., 2024. Detecting local earthquakes via fiber-optic cables in telecommunication conduits under Stanford University campus using deep learning, *Computers & Geosciences,* **190,** 105625, doi:10.1016/j.cageo.2024.105625.

Ide, S., Araki, E. & Matsumoto, H., 2021. Very broadband strain-rate measurements along a submarine fibre-optic cable off Cape Muroto, Nankai subduction zone, Japan, *Earth, Planets Space,* **73,** 1–10.

Jernelv, I.L., Hjelme, D.R., Matsuura, Y. & Aksnes, A., 2020. Convolutional neural networks for classification and regression analysis of one-dimensional spectral data, arXiv:2005.07530.

Jiang, F., Dai, F., Zhou, J. & Jiang, R., 2023. AI-powered automatic detection of dynamic triggering of earthquake based on microseismic monitoring, *Soil Dyn. Earthq. Eng.,* **165,** 107723, doi:10.1016/j.soildyn.2022.107723.

Jing, L. & Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.,* **43,** 4037–4058.

Jousset, P. *et al.*, 2018. Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features, *Nat. Commun.,* **9,** 2509, doi:10.1038/s41467-018-04860-y.

Jousset, P. *et al.*, 2022. Fibre optic distributed acoustic sensing of volcanic events, *Nat. Commun.,* **13,** 1753, doi:10.1038/s41467-022-29184-w.

Karrenbach, M., Ellwood, R., Yartsev, V., Cole, S., Araki, E., Kimura, T. & Matsumoto, H., 2021. Turning the Muroto seafloor cable into a long DAS sensing array, in *Proceedings of the 14th SEGJ International Symposium,* Tokyo, Japan.

Katsumata, A. & Kamaya, N., 2003. Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan, *Geophys. Res. Lett.,* **30,** 20–1.

Keskar, N.S., Mudigere, D., Nocedal, J., Smelyanskiy, M. & Tang, P.T.P., 2016. On large-batch training for deep learning: Generalization gap and sharp minima, arXiv:1609.04836.

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

Kowarik, S., Hussels, M.T., Chruscicki, S., Münzenberger, S., Lämmerhirt, A., Pohl, P. & Schubert, M., 2020. Fiber optic train monitoring with distributed acoustic sensing: conventional and neural network data analysis, *Sensors,* **20,** 450, doi:10.3390/s20020450.

Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks, *Commun. ACM,* **60,** 84–90.

LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning, *Nature,* **521,** 436–444.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998. Gradient-based learning applied to document recognition, *Proc. IEEE,* **86,** 2278–2324.

Lellouch, A., Yuan, S., Ellsworth, W.L. & Biondi, B., 2019. Velocity-based earthquake detection using downhole distributed acoustic sensing—examples from the San Andreas Fault Observatory at Depth, *Bull. seism. Soc. Am.,* **109,** 2491–2500.

Li, J., Kim, T., Lapusta, N., Biondi, E. & Zhan, Z., 2023b. The break of earthquake asperities imaged by distributed acoustic sensing, *Nature,* **620,** 800–806.

Li, J., Zhu, W., Biondi, E. & Zhan, Z., 2023a. Earthquake focal mechanisms with distributed acoustic sensing, *Nat. Commun.,* **14,** 4181, doi:10.1038/s41467-023-39639-3.

Li, Y., Zhang, M., Zhao, Y. & Wu, N., 2022. Distributed Acoustic Sensing Vertical Seismic Profile Data Denoising Based on Multistage Denoising Network, *IEEE Trans. Geosci. Rem. Sens.,* **60,** 1–17.

Lindsey, N. J. & Eileen, R. M., 2021. Fiber-optic seismology, *Annu. Rev. Earth planet. Sci.,* **49,** 309–336.

Lior, I., Rivet, D., Ampuero, J.P., Sladen, A., Barrientos, S., Sánchez-Olavarría, R., Villarroel Opazo, G.A. & Bustamante Prado, J.A., 2023. Magnitude estimation and ground motion prediction to harness fiber optic distributed acoustic sensing for earthquake early warning, *Sci. Rep.,* **13,** 424, doi:10.1038/s41598-023-27444-3.

Liu, X., Ren, T., Chen, H. & Chen, Y., 2021. Classification of tectonic and non-tectonic seismicity based on convolutional neural network, *J. geophys. Int.,* **224,** 191–198.

Liu, Y., Huff, O., Luo, B., Jin, G. & Simmons, J., 2022. Convolutional neural network-based classification of microseismic events originating in a stimulated reservoir from distributed acoustic sensing data, *Geophys. Prospect.,* **70,** 904–920.

Majstorović, J., Giffard-Roisin, S. & Poli, P., 2023. Interpreting convolutional neural network decision for earthquake detection with feature map visualization, backward optimization and layer-wise relevance propagation methods, *J. geophys. Int.,* **232,** 923–939.

Majstorović, J., Giffard-Roisin, S. & Poli, P., 2021. Designing convolutional neural network pipeline for near-fault earthquake catalog extension using single-station waveforms, *J. geophys. Res.,* **126,** 1–20.

Masoudi, A. & Newson, T.P., 2016. Contributed review: distributed optical fibre dynamic strain sensing, *Rev. Sci. Instrum.,* **87,** 011501, doi:10.1063/1.4939482.

Masoudi, A., Pilgrim, J.A., Newson, T.P. & Brambilla, G., 2019. Subsea cable condition monitoring with distributed optical fiber vibration sensor, *J. Lightwave Technol.,* **37,** 1352–1358.

Masters, D. & Luschi, C., 2018. Revisiting small batch training for deep neural networks, arXiv:1804.07612.

Matsumoto, H. *et al.*, 2021. Detection of hydroacoustic signals on a fiber-optic submarine cable, *Sci. Rep.,* **11,** 2797, doi:10.1038/s41598-021-82093-8.

Matthaiou, I., 2022. On robust statistical outlier analysis for damage identification, *Doctoral dissertation,* University of Sheffield.

Matthaiou,, I., Araki,, E., Kodaira,, S., Masoudi,, A., Modafferi,, S. & Brambilla,, G., Distributed acoustic sensing spatiotemporal maps from Cape Muroto, 2023. Southampton ePrints, doi:10.5258/SOTON/D2841.

Matthaiou,, I., Masoudi,, A. & Brambilla,, G., 2023. Processing strain data generated from distributed acoustic sensing for monitoring tasks, in *28th International Conference on Optical Fibre Sensors,* Hamamatsu, Japan.

Milne, D., Masoudi, A., Ferro, E., Watson, G. & Le Pen, L., 2020. An analysis of railway track behaviour based on distributed optical fibre acoustic sensing, *Mech. Syst. Sig. Process.,* **142,** 106769, doi:10.1016/j.ymssp.2020.106769.

Mousavi, S.M. & Beroza, G.C., 2022. Deep-learning seismology, *Science,* **377**. 1–11.

Mousavi, S.M. & Beroza, G.C., 2023. Machine learning in earthquake seismology, *Annu. Rev. Earth planet. Sci.,* **51**. 105–129

Mousavi, S.M., Zhu, W., Ellsworth, W. & Beroza, G., 2019. Unsupervised clustering of seismic signals using deep convolutional autoencoders, *IEEE Geosci. Rem. Sens. Lett.,* **16,** 1693–1697.

Muggleton, J.M., Hunt, R., Rustighi, E., Lees, G. & Pearce, A., 2020. Gas pipeline leak noise measurements using optical fibre distributed acoustic sensing, *J. Nat. Gas Sci. Eng.,* **78,** 103293, doi:10.1016/j.jngse.2020.103293.

Muller, A.P., Costa, J.C., Bom, C.R., Klatt, M., Faria, E.L., de Albuquerque, M.P. & de Albuquerque, M.P., 2023. Deep pre-trained FWI: where supervised learning meets the physics-informed neural networks, *J. geophys. Int.,* **235,** 119–134.

Murphy, K.P., 2022. *Probabilistic Machine Learning: Advanced Topics,* MIT Press.

Nakano, M., Hori, T., Araki, E., Kodaira, S. & Ide, S., 2018. Shallow very-low-frequency earthquakes accompany slow slip events in the Nankai subduction zone, *Nat. Commun.,* **9,** 984, doi:10.1038/s41467-018-03431-5.

Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T. & Tsuboi, S., 2019. Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images, *Seismol. Res. Lett.,* **90,** 530–538.

O'Malley, T. *et al.*, 2019. Available online: github.com/keras-team/kerastuner. Date accessed: 7 June 2023.

Perez, L. & Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning, arXiv:1712.04621.

Ren, L., Gao, F., Wu, Y., Williamson, P., McMechan, G.A. & Wang, W., 2023a. Automated dispersion curve picking using multi-attribute convolutional-neural-network based machine learning, *J. geophys. Int.,* **232,** 1173–1208.

Ren, T., Liu, X., Chen, H., Dimirovski, G.M., Meng, F., Wang, P., Zhong, Z. & Ma, Y., 2023b. Seismic severity estimation using convolutional neural network for earthquake early warning, *J. geophys. Int.,* **234,** 1355–1362.

Sabour, S., Frosst, N. & Hinton, G.E., 2017. Dynamic routing between capsules, *Adv. Neural Inform. Process. Syst.,* **30,** 3859–3869.

Schreurs, J., Vranckx, I., Hubert, M., Suykens, J.A. & Rousseeuw, P.J., 2021. Outlier detection in non-elliptical data by kernel MRCD, *Stat. Comput.,* **31,** 66, doi:10.1007/s11222-021-10041-7.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P. & De Freitas, N., 2015. Taking the human out of the loop: a review of Bayesian optimization, *Proc. IEEE,* **104,** 148–175.

Shelly, D.R., Beroza, G.C. & Ide, S., 2007. Non-volcanic tremor and low-frequency earthquake swarms, *Nature,* **446,** 305–307.

Shiloh, L., Eyal, A. & Giryes, R., 2019. Efficient processing of distributed acoustic sensing data using a deep learning approach, *J. Lightwave Technol.,* **37,** 4755–4762.

Shiloh, L., Lellouch, A., Giryes, R. & Eyal, A., 2020. Fiber-optic distributed seismic sensing data generator and its application for training classification nets, *Opt. Lett.,* **45,** 1834–1837.

Shorten, C. & Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning, *J. Big Data,* **6,** 1–48.

Simonyan, K. & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556.

Sladen, A., Rivet, D., Ampuero, J.P., De Barros, L., Hello, Y., Calbris, G. & Lamare, P., 2019. Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables, *Nat. Commun.,* **10,** 5777, doi:10.1038/s41467-019-13793-z.

Snoek, J., Larochelle, H. & Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms, *Adv. Neural Inform. Process. Syst.,* **25,** 2951–2959.

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

Szeliski, R., 2022. *Computer Vision: Algorithms and Applications,* Springer Nature.

Takemura, S., Obara, K., Shiomi, K. & Baba, S., 2022. Spatiotemporal variations of shallow very low frequency earthquake activity southeast off the Kii Peninsula, along the Nankai Trough, Japan, *J. geophys. Res.,* **127,** 1–15.

Thrastarson, S., Torfason, R., Klaasen, S., Paitz, P., Sabuncu, Y.C., Jónsdóttir, K. & Fichtner, A., 2021. Detecting seismic events with computer vision: Applications for fiber-optic sensing, preprint, ESS Open Archive, doi:10.1002/essoar.10509693.1.

Trabattoni, A., Biagioli, F., Strumia, C., van den Ende, M., Scotto di Uccio, F., Festa, G., Rivet, D., Sladen, A., Ampuero, J.P., Métaxian, J.P. & Stutzmann, É., 2023. From strain to displacement: using deformation to enhance distributed acoustic sensing applications, *J. geophys. Int.,* **235,** 2372–2384.

Trnkoczy, A., 2009. Understanding and parameter setting of STA/LTA trigger algorithm, in *New Manual of Seismological Observatory Practice,* pp. 1–20, Deutsches GeoForschungsZentrum GFZ.

van den Ende, M., Lior, I., Ampuero, J.P., Sladen, A., Ferrari, A. & Richard, C., 2021. A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data, *IEEE Trans. Neural Networks Learn. Syst.,* **34,** 3371–3384.

Wallace, L.M. *et al.*, 2016. Near-field observations of an offshore Mw 6.0 earthquake from an integrated seafloor and subseafloor monitoring network at the Nankai Trough, southwest Japan, *J. geophys. Res.,* **121,** 8338–8351.

Wu, H. & Gu, X., 2015. Towards dropout training for convolutional neural networks, *Neural Networks,* **71,** 1–10.

Yalniz, I.Z., Jegou, H., Chen, K., Paluri, M. & Mahajan, D., 2019. Billion-scale semi-supervised learning for image classification, arXiv:1905.00546.

Yamamoto, Y., Yada, S., Ariyoshi, K., Hori, T. & Takahashi, N., 2022. Seismicity distribution in the Tonankai and Nankai seismogenic zones and its spatiotemporal relationship with interplate coupling and slow earthquakes, *Prog. Earth planet. Sci.,* **9,** 1–20.

Yang, L., Fomel, S., Wang, S., Chen, X. & Chen, Y., 2023b. Denoising distributed acoustic sensing (DAS) data using unsupervised deep learning, *Geophysics,* **88,** 1–12.

Yang, L., Fomel, S., Wang, S., Chen, X., Chen, W., Saad, O.M. & Chen, Y., 2023a. Denoising of distributed acoustic sensing data using supervised deep learning, *Geophysics,* **88,** 91–104.

Yin, J., Denolle, M.A. & He, B., 2022. A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms, *J. geophys. Int.,* **231,** 1806–1822.

Zeiler, M.D. & Fergus, R., 2013. Stochastic pooling for regularization of deep convolutional neural networks, arXiv:1301.3557.

Zhang, A., Lipton, Z.C., Li, M. & Smola, A.J., 2021. Dive into deep learning, preprint, arXiv:2106.11342.