

Microseismic event classification with time, frequency and wavelet domain Convolutional Neural Networks

Jiaxin Jiang, Vladimir Stankovic, Lina Stankovic, Emmanouil Parastatidis, Stella Pytharouli

Abstract—Passive seismics help us understand subsurface processes, e.g. landslides, mining, geothermal systems etc. and help predict and mitigate their effects. Continuous monitoring results in long seismic records that may contain various sources, which need to be classified. Manual detection and labeling of recorded seismic events is not only time consuming but can also be inconsistent when done manually, even in the case where it is done by the same expert. Therefore, an automated approach for classification of continuous microseismic recordings based on a Convolutional Neural Network (CNN) is proposed, with a multiclassifier architecture that classifies earthquakes, rockfalls and low signal to noise ratio quakes. Furthermore, we propose three CNN architectures that take as input time series data, Short Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) maps. The suitability of these three networks is rigorously assessed over five months of continuous seismometer recordings from the active Super-Sauze landslide in France. We observe that all three architectures have excellent and very similar performance. Furthermore, we evaluate transferability to a geographically distinct seismically active site in Larissa, Greece. We demonstrate that the proposed network is able to detect all 86 catalogued earthquake events, having only been trained on the Super-Sauze dataset and shows good agreement with manually detected events. This is promising as it could replace painstaking manual labelling of events in large recordings.

Index Terms—Microseismic event classification, Short time Fourier transform, Continuous wavelet transform

I. INTRODUCTION

Deformation of slow-moving clay-rich landslides can cause endogenous seismicity [1], [2], [3], including microseismicity, such as quake, rockfall and tremor-like signals. Quakes are defined as seismic events originating as a result of landslide processes, for example, formation of surface fissures. Researching micro-seismic events leads to improved understanding of subsurface processes with applications from predicting landslides to planning of mining activities and geothermal exploration. This requires accurate detection of the waveform of micro-seismic events from traces of seismic recordings. Manually detecting events and labelling them is time-consuming and a subjective task, prone to errors and bias. Thus, methods that automatically detect and classify seismic events are needed. However, micro-seismic events are of low magnitude and highly attenuated, which makes the detection and classification tasks challenging.

Prior work has focused on application of traditional machine learning algorithms to classification of (micro-)seismic events, supported by various signal processing tools for denoising and detection of events. Most classification approaches have been based on well-known algorithms, such as Hidden Markov Models (HMMs), Support Vector Machine (SVM) and Random decision Forest (RF) [4], [5], [6], [7]. For example, Provost et al. [8] propose a classification method using an RF supervised classifier to classify micro-seismic events on slow-moving landslides. The method uses the STA/LTA algorithm for detection, then calculates 71 seismic attributes as features inputted into a supervised RF to classify each event into one of four pre-determined classes (earthquake, quake, rockfall and Natural/Anthropogenic noise). More recently, an end-to-end automated system is proposed in [9] consisting of signal denoising, event detection via statistical Neyman-Pearson based thresholding, feature selection, and graph-based classification.

In contrast to traditional pipeline-based approaches, e.g., [4], [5], [6], [8], [9], deep learning provides an integrated approach to detection, feature representation and classification, with competitive performance under the assumption that a good representative dataset is available for training. Though there have been many attempts to use various deep learning architectures for seismic signal detection and classification (e.g., [10], [11], [12], [13], [14], [15]), classification of micro-seismic endogenous landslide events based on deep learning is rarely studied. Moreover, transferability of deep learning classification models to different monitoring network geometries is rarely discussed.

In this paper, after providing a detailed literature review, we formulate the microseismic classification problem as a multi-class classification task, and propose three convolutional neural network (CNN) models. As in [8], we classify four types of events, namely earthquakes, rockfalls, seismic sources related to landslide processes, e.g., fissure formation (thereafter referred to as quakes) and anthropogenic noise. To capture a variety of time-domain and frequency features at different scale, besides using filtered raw time-series waveforms as input to the network, we also use Short-time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) coefficients. This leads to three different architectures, each optimised and designed for one input type. As typical with time-series data analysis, we slide a fixed length window over the input to achieve multi-classification on continuous data.

To ensure reproducibility of the work, we train and test the proposed models on publicly accessible Résif dataset

The authors are from University of Strathclyde, Glasgow, UK. Emails: {jiaxin.jiang, vladimir.stankovic, lina.stankovic, emmanouil.parastatidis, stella.pytharouli}@strath.ac.uk.

[16] that contains 6-channel seismic recordings at sampling rate of 250Hz over a period of roughly 4 months with over 1000 (micro)seismic events recorded. In addition, we test our pre-trained model on a geologically-distinct dataset from the region of Larissa in mainland Greece, using two different array geometries, to evaluate its transferability. This dataset was made available to us by the Aristotle University of Thessaloniki, and is available upon request through [17]. In summary, our main contributions are:

- Three CNN-based multi-classifier models for three different inputs (time series, STFT maps and CWT maps) for classification of three different micro-seismicity types plus anthropogenic noise on continuous recordings.
- Detailed evaluation and analysis of classification performance of the three models, including reliability of the results, and analysis of correctly and incorrectly classified examples to shed light into the most important features of the input signal and reasons for mis-classification.
- Evaluation of transferability of the proposed model by testing the CNN model pre-trained on Résif dataset on a geographically-distinct dataset and analysis of how different array typologies affect seismic signal classification
- Validation of the classified events via a combination of known catalogue of earthquakes, as well as manual and commercial software corroboration for additionally classified non-catalogued events.
- Release of validated labelled dataset (additional catalogue¹) on public repository comprising earthquakes, rockfalls and quake events, as well as three proposed models for reproducing results².

We present how our work builds upon and goes beyond the state-of-the-art in Section II, including summary Table I. Our proposed methodology including the proposed network structures is presented in Section III, after which the last three contribution points including transferability and feature maps visualization are discussed in depth in Section IV, before concluding in Section V.

II. BACKGROUND

We provide an overview of approaches for seismic signal classification based on deep learning (DL). Most DL-based seismic classifiers are binary classifiers (with one class representing the event of interest and the other class groups everything else) usually detecting earthquake events only. There are some emerging seismic multi-classifiers which output more than two categories. We also include DL approaches whose end result is classification, although they also perform detection or phase-picking. It is worth noting that an additional detection step is not necessary if performing classification on continuous recordings, parsed into windows. For non-DL approaches, an up-to-date review can be found in [9].

A. Binary classification

In [12], seismic data are sampled and parsed into time windows of 20 msec duration, which are then fed into a

CNN. The CNN model, consisting of one 1-D convolutional layer, one pooling layer and 3 fully connected (FC) layers, acts as a binary classifier and classifies the input window into earthquake event waveform or noise waveform. Perol et al. [10] propose earthquake detection techniques for multi-channel 1-D data using CNN with 3 channels and sampling rate of 100Hz. The raw recordings were split into 10sec seismic waveform windows that are fed into a trained network consisting of eight 1-D convolutional layers to extract features for earthquake detection, followed by a FC layer to perform the earthquake/noise classification and location estimation by features outputted by convolutional layers.

Besides feeding raw signals, various methods are proposed that take as inputs spectrograms [18], [19]. For example, Dokht et al. [20] propose a CNN model to classify the input spectrogram into earthquake or noise. The authors use 3-component spectrograms of 10sec seismic data as input to a CNN architecture consisting of 4 convolutional and 2 FC layers. Each convolutional layer is followed by a max-pooling layer. The final output layer is a two-neuron FC layer with softmax activation which outputs the probability distribution of 2 classes (earthquake and noise). Liao et al. [18] use the method of transfer learning to detect whether the CWT map contains first break of earthquake. In particular, the authors use pre-trained CNNs used for image classification tasks, namely, GoogLeNet, AlexNet and SqueezeNet, to perform transfer learning from image data to seismic signals and classify the CWT output into first-break waveform and not first-break waveform. Linville et al. [21] use long-short-term memory (LSTM) and CNN models to classify seismic events as either quarry blasts or earthquakes. The LSTM model routes input spectrograms to output classes (0 or 1) which represents quarry blasts or earthquakes through 4 bidirectional layers as a many-to-one learning scenario which takes input from many time steps to make one binary classification output. The CNN architecture with 4 convolutional, 4 max-pooling and 2 FC layers outperforms RF, SVM and residual neural network.

Mousavi et al. [22] propose a sequence-to-sequence learning model, ‘EQ-transformer’, for phase picking and earthquake classification using a multi-task structure, that outputs 3 sequences of probabilities, representing presence of earthquake, P-phase picking and S-phase picking. The deep network structure consists of an encoder that converts the raw input signal into features through 1-D convolution, max-pooling, residual convolution, and LSTM layers, and 3 separate decoders. In [23], the authors propose a vision transformer (ViT)-based system for earthquake detection and its magnitude prediction. The system consists of two separate ViT networks: the first one detects earthquake events from the picked P-wave; the second network predicts the magnitude of the detected earthquakes. In [13], spectrograms of 30 seconds 3-component seismograms are used as input to a CNN-RNN Earthquake Detector (CRED), that consists of convolutional, recurrent and dense layers, in a residual structure. A 2D convolution layer extracts features from the input spectrograms. Then, bi-directional LSTM performs sequence learning. Finally, dense layers classify the extracted features and output a sequence of predicted probabilities, for classification of earthquakes and

¹<https://doi.org/10.15129/589f7af3-26b3-4a93-b042-fbc8100fc977>

²<https://github.com/kanata2020/microseismic-classification>

noise. [24] perform a 7-level CWT on 3-component 30s-long time window with 100Hz sampling frequency to construct a CWT map as input to an encoder-decoder network with residual learning to classify earthquake signals using 2 classes - earthquake signal or not.

B. Classification of more than one event type

A CNN is proposed in [11] to classify seismic events into 3 categories - tectonic earthquakes, mining-induced events, and mining blasts, based on 90sec long spectrograms as input. The model consists of 4 2-D convolutional and a 3-node softmax activated dense layer. A ‘deepquake’ CNN architecture is proposed in [25] that classifies 3-component 20sec input data into earthquake, other events and noise. Two CNNs are built for two different input types: time series and spectrograms. The two architectures consist of 6 convolution layers for feature extraction and one dense layer for classification. An attention-based CNN architecture is proposed in [15] using multi-task learning. This architecture first acts as a binary classifier and classifies the seismic waveform into earthquake or noise; then as a multi-class classifier, it classifies the seismic waveform into micro-earthquake, macro-earthquake or noise. The input data is a 10-sec raw seismic waveform with 100Hz sampling rate. 8 1-D convolutional layers (with Relu activation) with an attention module to extract features, and 2 task-specific layers with 2 FC layers (softmax activation) classify the features.

A 3D-CNN/RNN-based architecture is proposed in [26] to classify earthquake magnitudes. Each segment of 60-sec waveform is split into 6×10 sec clips, which are then processed and transformed into a 2-D Log-Mel map. Thus, the input data of the model is the 3-D matrix of 6 Log-Mel maps stack that is first inputted to a 3-D CNN and then to RNN. FC layer is used to classify the extracted features into five categories (greater than or equal to 0.0, 1.0, 2.0, 3.0 and 4.0 on Richter scale). In [27], a CNN is used to detect and classify seismic events into microseismic event, single-phase event, and ambient noise events. The model’s input is a 22×2000 seismogram image obtained from 22 seismometer channels with 2000 sampling points. The proposed architecture consists of 6 convolutional, 6 pooling, and 2 FC layers. The models are first trained and tested on synthetic data and then used to detect microseismic events from a field data set. It is showed that training on synthetic data and testing on the field data leads to poor performance. The results are then improved by labelling the field data and using it for training the model and testing.

C. Summary

Table I summarises the state-of-the-art in DL-based approaches whose end result is classification, as per the aim of this paper. In the table, ‘transfer learning’ denotes using a trained model to test on an unseen dataset collected at a different location. In [11], [21], [26], [15], [27], [35], only one type of feature is used as input without any feature selection study. It can be seen that most proposed architectures are based on CNN, and mainly use temporal or spectral features with 1 to 6 channels. The CNN model of [25] reports classification results that are among the best in the literature, without

complex feature engineering. Although using more complex input features and network structures, such as [26], [13], [22], [24], [23], can lead to high performance, the complex pre-processing steps and deep models are computationally demanding. In addition, sequence-to-sequence learning requires clearly labelled start and end times of each event. Limited by the absence of such a dataset for rockfall and quakes, as is usually the case in practice, we adopt a sequence-to-point classification architecture.

The uniqueness of our approach lies in its ability to classify three important types of landslide micro-seismicity, which has not been addressed in prior work on DL, as well as anthropogenic noise. Moreover, our architecture demonstrates successful transfer learning for endogenous landslide seismicity from one seismic dataset to another collected at a different location, showing that differences in structural terrain do not necessarily affect attributes that the model learned during training. Finally, this paper offers three types of architectures to effectively exploit both temporal and frequency features and analyses how these features affect detection and classification performance. It is worth noting that the proposed approach does not require a P-wave picking or detection step (manually or via an algorithm) since it performs classification on continuous recordings, as sliding windows, fed directly into the network.

III. METHODOLOGY

We perform pre-processing on the recordings, namely filtering and segmenting the filtered raw signal into windows, before performing multi-classification.

A. Denoising

A detailed review of various denoising methods is reported in [9]. In this paper, we use band-pass filtering to remove high frequency measurement noise and low frequency noise originating from humans, vehicles, rain, and animals. Specifically, having in mind the frequency range of events of interest, the raw recordings are filtered by a 3-order 5-60Hz Butterworth band-pass filter.

B. Data Processing

We use 3 different model inputs: raw temporal data, STFT and CWT maps. Figure 1 shows examples of three classes of seismic events (earthquake, quake and rockfall). For time-series raw signals, we use 10 seconds window as input. Given the sampling frequency of 250 Hz, the length of the input window is thus 2500 samples, which is usually sufficient to capture the entire seismic event, and short enough to ensure manageable complexity. Thus, for 6-channel recordings, the input time series signal is of dimension 2500×6 . To perform STFT, we use Hann window with heuristically set length of 128 samples with 75% overlap, generating output of dimension of $65 \times 75 \times 6$. For CWT, as in [24], as mother wavelet, we use the Morlet wavelet with 8 cycles. We construct scalograms using 80 scales spanning frequency range between 5Hz and 60Hz. Thus, CWT model’s input dimension is $80 \times 2500 \times 6$.

TABLE I: Summary of prior work on using deep learning models for seismic signal classification. UUSS stands for University of Utah Seismic Stations, UUEB for Unconstrained Utah Event Bulletin, NCEDC stands Northern California Earthquake Data Center. NECIS stands for National Earthquake Comprehensive Information System, IRIC for Incorporated Research Institutions for Seismology, STEAD for Stanford Earthquake Dataset and KNMI for Royal Netherlands Meteorological Institute.

Paper	Year	Aim	Best Architecture	Input data	Dataset & Transf. learning
[21]	2019	seismic classification (quarry blasts and earthquakes)	CNN & LSTM	Spectrograms (three channels)	UUSS
[11]	2019	seismic classification (tectonic earthquakes, mining-induced events, mining)	CNN	Spectrograms (three channels)	UUSS & UUEB [28]
[13]	2019	seismic detection (earthquakes and noise)	residual structure with convolutional, recurrent, and dense	Spectrograms (three channels)	North California [29] transfer learning to Arkansas [30]
[15]	2020	seismic detection (earthquake and noise) and classification (microearthquake, macroearthquake and noise)	attention-based CNN	Time series (three channels)	NECIS [31] & IRIS [32]
[22]	2020	seismic detection and phase picking (earthquake, P-phase and S-phase)	multi-task structure	Time series (three channels)	STEAD [33] transf learning to aftershock region of Tottori
[26]	2021	earthquake magnitude classification (greater than or equal to 0.0, 1.0, 2.0, 3.0 and 4.0 on Richter scale)	3D-CNN RNN	Log-Mel spectrogram (one channel)	STEAD
[24]	2021	Seismic detection (earthquake and noise)	U-NET	CWT map (3 channels)	Northern California transfer learning to Arkansas, Texas [34], Japan and Egypt
[27]	2021	microseismic classification (Dyke-roadway, Longwall, Low-frequency and noise)	CNN	seismogram (6 channels)	an underground coal mine
[35]	2021	Microseismic event classification (microseismic event, single-phase event, and ambient noise)	CNN	seismogram (22 channels)	synthetic data; unsuccessful transfer learning to field data
[23]	2022	seismic detection (P-wave and noise) and magnitude estimation	Vision Transformer	Time series (three channels)	STEAD
[25]	2022	seismic classification (earthquake, other events and noise)	CNN	Time series and seismogram (three channels)	KNMI [36]
This paper	2022	microseismic classification (earthquake, quake, rockfall and noise)	CNN	Time series, STFT maps and CWT maps (6 channels)	R��sif [16] Transfer learning to Larissa [17]

C. Network Architecture

The architectures of the proposed three networks, one for each type of signal input, are inspired by VGGNet [37] and adapted to the sampling rate, the size of the input seismic signal and its feature map. These are deep networks composed of convolutional layers (for feature representation and extraction), max pooling layers (for downsampling the extracted features to obtain the feature map of a small size) and FC layers (for classification). That is, after multiple convolutional layers, where the number of convolutional kernels increases with the layer number, and max-pooling layers, the input signals are compressed into small feature maps, that are then classified through 3 FC layers. The output layers have 4-node softmax classifiers, providing the probability distribution of 4 classes of events. The three architectures are shown in Figure 2.

Note that for the time waveform case, since the input 1-D time series contains positive and negative values, the activation functions of the first two convolutional layers are set to Linear, to ensure that a large number of neurons are not killed (go to zero) with ReLU, which would reduce the efficiency of learning.

The parameters for STFT and CWT models, were set to

match the complexity to that of time-series case, trading off complexity and performance. All convolutional layers are 2-D convolutions due to STFT/CWT feature maps being 2-D image signals. In contrast to STFT, in the CWT model, due to the fact that one dimension of the input CWT map is much larger than the other (80×2500 input), we set the stride of some convolutional layers to 1×2 , to allow the longer dimension to compress faster. Moreover, the size of some convolution kernels is set to (3, 9) to obtain a larger temporal field of view. Many deployed monitoring network configurations contain less than 6 channels. For such systems, we design single-channel models that take one channel at a time as inputs. The parameters of the single-channel model are the same as in the multi-channel model except that the input shape is changed. To classify single-channel data, the input size is 2500×1 for time series, $65 \times 75 \times 1$ for STFT maps and $80 \times 2500 \times 1$ for CWT maps. When these single-channel models are used with n -channel data, it will output n softmax vectors for each event. Then, to make a decision, we calculate the mean of each class for these n vectors.

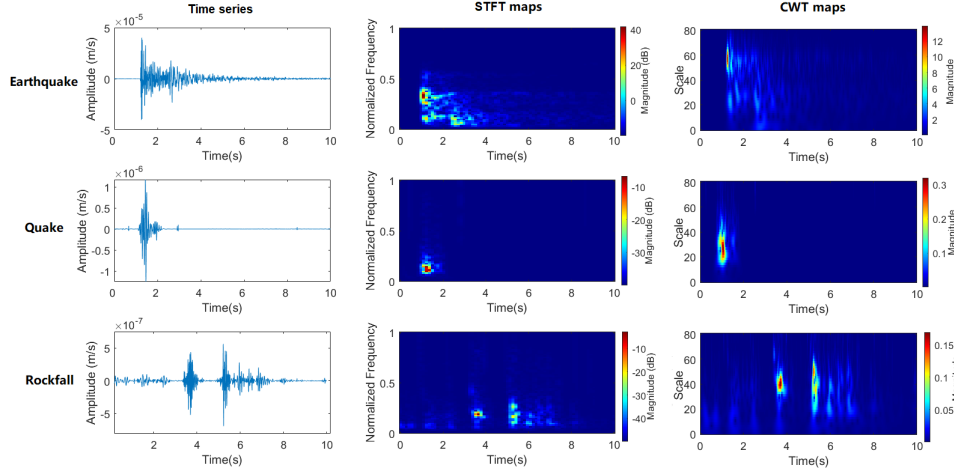


Fig. 1: Examples of seismic signals for 3 classes (earthquake, quake and rockfall) with 3 different input formats (temporal waveform, STFT and CWT maps) in Résif dataset.

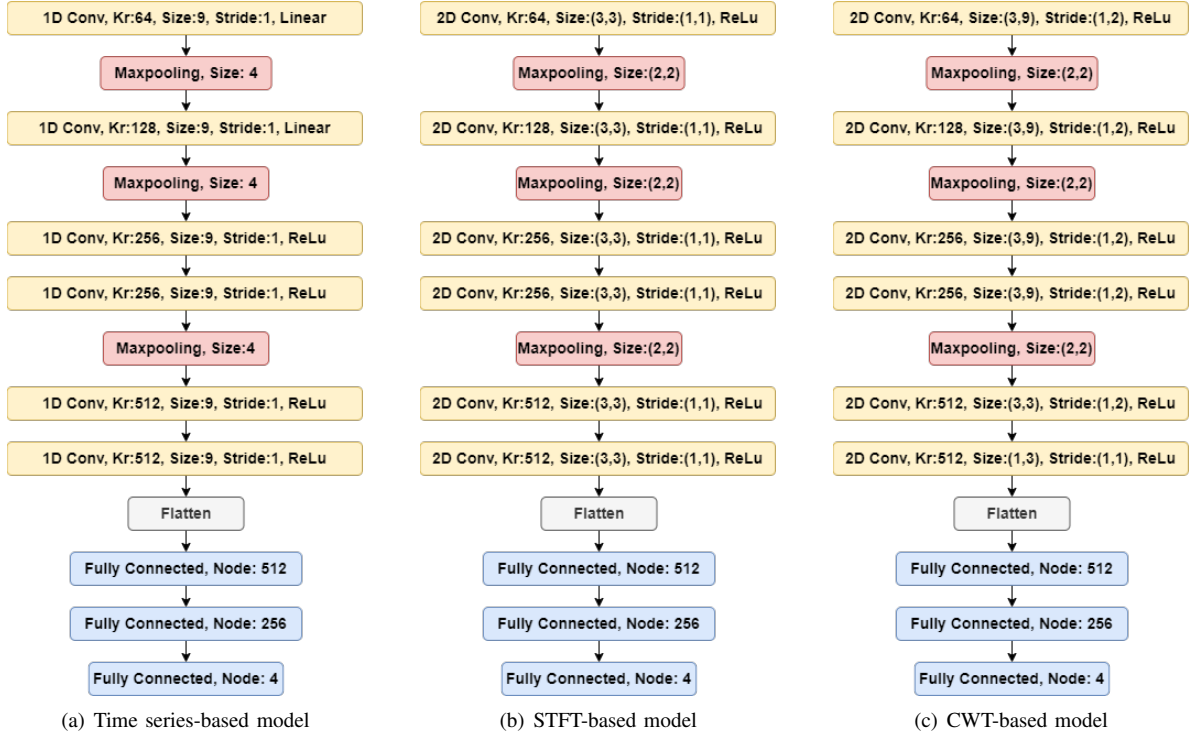


Fig. 2: The network structure for three different inputs.

IV. EXPERIMENTAL FINDINGS

A. Dataset

The dataset used to train all the models is openly accessible from the Résif Seismological Data Portal, acquired by the French Landslide Observatory OMIV (Observatoire Multi-disciplinaire des Instabilités de Versants) [16]. The data is acquired by MT network, specifically, Super-Sauze C (SZC) stations which are installed at the east and west sides of the Super-Sauze landslide in Southeast France, (Latitude: 44.34787, Longitude: 6.67805). See [38] for more details about the sensors used and the terrain. The signals were recorded over 3 periods, namely, from 11 Oct. to 19 Nov. 2013;

from 10 to 30 Nov. 2014; and from 9 June to 15 Aug. 2015. The seismometers consist of one three-component sensor and three vertical one-component sensors (organized as equilateral triangle); thus, the seismometers provide a 6-channel seismic signal, each at a sampling rate of 250Hz.

The raw seismic data is accompanied by a catalogue of labelled events. There are four different types of seismic events in this dataset: earthquakes, quakes, rockfalls and natural/anthropogenic (N/A) noise signals [8]. The total number of events is shown in Table II, as well as the number of events used for testing. Rockfalls mainly occur at the main scarp of the landslide, where the rigid block falls from the steep slope (height > 100m). The quake is likely to be triggered by mate-

rial damage, surface cracks and openings. The earthquakes represent regional seismic events in this area and the teleseisms. N/A noise events include all anthropogenic and environmental noise, due to, e.g., transportation, pedestrian walking, heavy rain, animals, strong wind, etc. For more details about the endogenous seismicity at Super-Sauze landslide, readers are referred to [39], [40], and [2]. In Figure 3, we show examples of seismic signals, namely, earthquakes, quakes and rockfalls, with high, medium and low SNR illustrating a variety of noisy events present in the dataset.

To validate our classifiers, first we use only labelled events, that is, we removed all sections in the dataset that were not catalogued. We note that the catalogue includes natural/anthropogenic noise segments, hence the classifiers are trained to distinguish this type of noise as well as the other three micro-seismic events. We split the dataset of labelled events into training (60%), validation (10%), and testing (30%) sets according to the time of the event (earliest to latest). To increase learning efficiency, we standardise the dataset by subtracting the mean and dividing by the standard deviation after denoising. Since the original dataset is very unbalanced (see Table II), the training set was balanced to avoid training bias, by generating new events by shifting the seismic events in the window as well as adding background noise from non-catalogued eventless waveforms to the catalogued events. After data augmentation, each class of events has about 600 training samples in the training set. We did not balance the test set, each event in the test set is unique.

TABLE II: The number of labelled events in Résif catalogue.

Class	Total No. events	No. of events for testing
Earthquake	388	116
Quake	234	70
Rockfall	401	120
Noise	351	105

B. Classification Results

To evaluate classification performance, we use standard classification performance measures, i.e., precision, recall and F1 score as in [9]. All networks are implemented using the framework of Keras, and trained over 100 epochs. We used ADAM as the optimiser and cross-entropy as the loss function. The initial learning rate was set to 0.0007 and decreased by 10% with every five epochs.

The classification performance results for the three models are shown and compared in Table III. Each model was trained and tested 5 times using the same training and testing sets to ensure repeatability. The results are presented in the form of “mean \pm standard deviation” where the mean and standard deviation are calculated using the results obtained after 5 trainings and tests.

It can be seen from the table, that all three models provide highly accurate and similar classification performance, with the F1-score ranging from 0.85 for quakes to 0.98 for earthquakes. The CWT model has marginally better classification accuracy (averaged over all classes) of 90.82%, followed closely by the STFT model with 90.62%, and the time series model with 89.97%. However, the relatively lower overhead of

the time-series model over the others makes it more desirable in practice.

The results using the proposed single-channel models when only one channel is available are shown in Table IV. Comparing with the multi-channel models, the single-channel models performs slightly worse for all classes of events. This can be explained by the fact that the channels with very poor SNR affect the output due to averaging, which is not the case with the multi-channel model.

Our results are aligned with [8], where an RF classifier is used on a subset of randomly chosen events from the same dataset, and correctly classified 94% of earthquakes (vs. 98% with the proposed method in Table III), 94% of rockfalls (vs. 92% in Table III), 92% of noise (vs. 86% in Table III) and 93% of quakes (vs. 89% in Table III).

The corresponding confusion matrices are shown in Table V, for time waveform, STFT, and CWT models, respectively. From the confusion matrices, we can see that few quake events are misclassified, due to much shorter duration (<5seconds), smaller SNR and the fact that quakes are usually localised events and hence not detected by all channels, which is not the case with earthquakes. Quake classification results are better for STFT and CWT than time series inputs, with fewer misclassified events, since CWT and STFT features take into account both frequency and time duration. However, the recall (sensitivity) of quake is still slightly inferior to the 93% reported by [8]. We think there are three reasons for this. First, Provost et al. [8] use 71 constructed features including nine network geometry attributes (such as mean and std correlation lag between the stations; stations with max/min amplitudes, etc.), that assume knowledge of positions of the deployed sensors. After removing these features, the performance of [8] was reduced to 90% in average over all classes, ranging between 86% and 94%. Secondly, the number of labelled quake events is much smaller than the number of other events. For example, there are 401 labeled rockfalls and only 234 quakes, and this has negative impact on the generalisation ability of the networks. Thirdly, in [8] the channel with the highest SNR is chosen for feature extraction. In our models, all 6 channels’ data were inputted, and some channels with low SNR may impact classification results.

C. Interpretation of Misclassified Events

As can be seen from the confusion matrices, the main cause of a drop in recall is quake events being misclassified as noise, which is not surprising given their low SNRs. Therefore, we looked closely at cases when the quake events are misclassified as noise (10 events in total across all three multi-class models). Out of these 10 misclassified quake events, 4 events were misclassified by 2 or 3 models. We focus on these events and show a representative example for the time-series model in Figure 4. From the figure, it can be seen the main reason for the wrong classification is that the energy of the seismic signal is too weak compared to background noise. Indeed, in Figure 4 (left column), the wave peak of the correctly classified quake event reaches roughly 1×10^{-6} m/s while the wave peak of the missed quake event (right column) is only around 8×10^{-8} m/s, that is, two magnitudes lower.

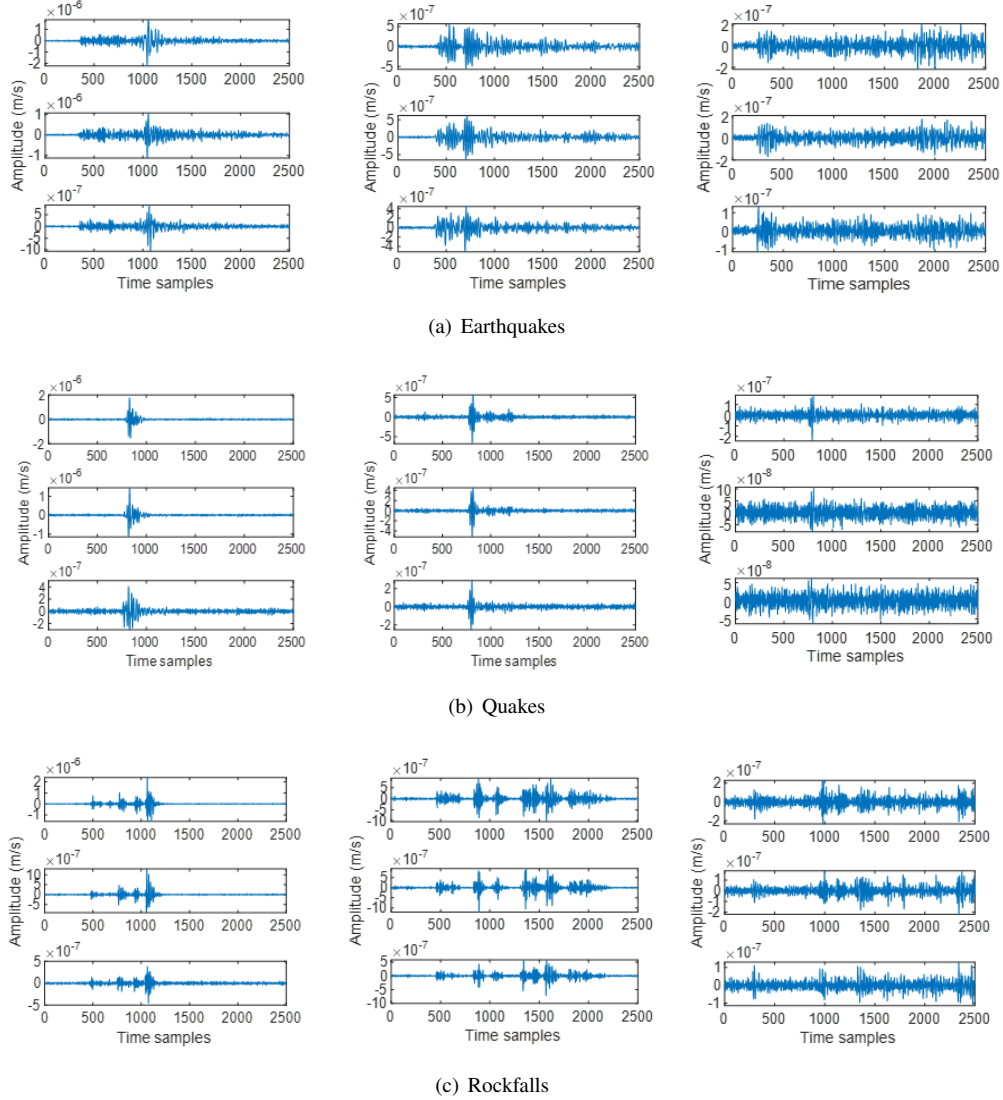


Fig. 3: Examples of seismic events: (a) Earthquake (b) Quake (c) Rockfall. Three different events are shown for each class, including high, medium and low SNR cases. In all cases, we show three waveforms, representing, respectively, the East, North and vertical direction of the three-component sensors from top to bottom.

TABLE III: The classification performance results for the three six-channel models.

	Input: Time series			Input: STFT maps			Input: CWT maps		
	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score
Earthquake	0.96 ± 0.016	0.97 ± 0.009	0.96 ± 0.008	0.96 ± 0.019	0.98 ± 0.004	0.97 ± 0.011	0.97 ± 0.017	0.98 ± 0.008	0.98 ± 0.007
Rockfall	0.91 ± 0.023	0.90 ± 0.011	0.90 ± 0.010	0.88 ± 0.009	0.92 ± 0.012	0.90 ± 0.007	0.91 ± 0.004	0.91 ± 0.005	0.91 ± 0.005
Quake	0.86 ± 0.032	0.84 ± 0.015	0.85 ± 0.013	0.90 ± 0.039	0.87 ± 0.019	0.88 ± 0.023	0.90 ± 0.023	0.88 ± 0.015	0.89 ± 0.011
Noise	0.85 ± 0.011	0.86 ± 0.013	0.86 ± 0.004	0.89 ± 0.008	0.85 ± 0.013	0.86 ± 0.009	0.86 ± 0.015	0.86 ± 0.017	0.86 ± 0.013

TABLE IV: The classification performance results for the three single-channel models.

	Input: Time series			Input: STFT maps			Input: CWT maps		
	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score
Earthquake	0.95 ± 0.010	0.93 ± 0.026	0.95 ± 0.016	0.96 ± 0.007	0.97 ± 0.007	0.96 ± 0.005	0.96 ± 0.004	0.97 ± 0.012	0.96 ± 0.005
Rockfall	0.86 ± 0.018	0.83 ± 0.013	0.85 ± 0.014	0.89 ± 0.012	0.88 ± 0.009	0.89 ± 0.008	0.91 ± 0.009	0.81 ± 0.016	0.86 ± 0.002
Quake	0.85 ± 0.012	0.83 ± 0.018	0.84 ± 0.016	0.90 ± 0.021	0.84 ± 0.024	0.86 ± 0.020	0.87 ± 0.015	0.87 ± 0.013	0.87 ± 0.002
Noise	0.80 ± 0.027	0.84 ± 0.011	0.82 ± 0.013	0.84 ± 0.014	0.85 ± 0.009	0.84 ± 0.014	0.81 ± 0.011	0.88 ± 0.017	0.84 ± 0.006

TABLE V: The confusion matrix for the three six-channel models.

	Input: Time series				Input: STFT maps				Input: CWT maps			
	Earthquake	Rockfall	Quake	Noise	Earthquake	Rockfall	Quake	Noise	Earthquake	Rockfall	Quake	Noise
Earthquake	112	1	0	3	113	2	0	1	115	0	0	1
Rockfall	0	110	4	6	3	110	0	7	1	109	2	8
Quake	4	1	58	7	2	3	61	4	1	3	62	4
Noise	3	5	7	90	2	13	5	85	1	8	11	85

In general, quake events are of low amplitude and short duration. Due to the energy of seismic activity and the distance between the event location and the monitoring station, a small number of these events have extremely low amplitude, and the duration of seismic events is often also very short. This makes the model prone to misclassifying such events as noise, as out-of-distribution samples.

D. Comparison with state of the art DL-based methods

In addition to benchmarking against traditional machine learning methods [8], we compare the performance against the state-of-the-art CNN-based ‘deepquake’ network [25] that was shown to outperform similar DL architectures [25]. The ‘deepquake’ network has two models, one for time series input and another for STFT map, namely ‘arch-time’ and ‘arch-spect’, respectively. The network uses 20 seconds windows with 100Hz sampling frequency as input. Thus, we first down-sample our data to 100Hz (from 250Hz), use the same proposed pre-processing method to normalise data and extract 20 seconds of event waveforms as input window. The ‘deepquake’ model classifies inputs into three classes: earthquakes, other events and noise. Thus, we re-label both quakes and rockfalls as ‘other events’. We use ‘deepquake’ pre-trained models as initial weights of each layers. Then, we re-train these two models using the Résif dataset for an additional 80 epochs. The results are shown in Table VI. From the confusion matrix, it can be seen that ‘arch-spect’ has better performance than ‘arch-time’ on Résif data, which is expected and in accordance with our results, i.e., the STFT map input model outperforms the time series input model.

Comparing these results with Table V, we can see that our time series-based CNN model outperforms ‘arch-time’ for all classes. Indeed, 104 earthquake events are correctly classified by ‘arch-time’ while 112 earthquake were correctly classified by our time series-based CNN. 22 other events (rockfalls and quakes) are misclassified as noise by ‘arch-time’ while 6 rockfalls and 7 quakes (13 in total) are misclassified as noise by our model. For STFT maps as input, we can see that ‘arch-spect’ performs worse than our STFT-based CNN. Both models correctly classify 113 earthquakes. 14 other events are misclassified as noise by ‘arch-spect’ while 7 rockfalls and 4 quakes (11 in total) are misclassified as noise by our STFT-based model. This shows that our proposed model is in line with the state-of-the-art, with the advantage of additionally distinguishing endogenous landslide seismicity, including rockfalls and quakes.

E. Continuous detection results

To demonstrate applicability of our aforementioned trained time-series model on a continuous data feed, we test on an unseen period (25-28 Nov 2014) in the Résif dataset, during which 18 quakes, 23 earthquakes, 65 rockfalls were catalogued. Since the input of the model is a time window, and the output is a 4-class probability vector, we slide the input window on the continuous data to achieve a continuous series of probabilities, as in [41], [42], [43]. The input window is set, as before, to 10 sec. The input time window was slid

with 90% overlap, that is, the classification result is output each second. We set the decision threshold to 0.7: that is, a class probability greater than 0.7 will be considered as an event class of this time window. If there is no class with output probability greater than 0.7 in the output probability vector, this time window will be considered as noise. Our network correctly detected and classified 91% of earthquakes, 83% of quakes and 94% of rockfalls (TP). These results are similar to the time series-based model’s recall (sensitivity) in Tables III and V, and therefore we conclude that the network is robust to continuous detection and classification. In Figure 5, we show 2 and a half minutes (i.e., 150 sec) of the continuous waveform of an earthquake event that occurred at 4:05:39 on November 28, 2014. It can be seen from the figure that the model correctly detects the start of the signal. In addition, our network detected many other events that have not been catalogued, namely, 174 earthquakes, 260 quakes and 32 rockfalls.

F. Transferability to geologically distinct site with different monitoring network geometry

Using the same sliding window of continuous time series recordings as above, we evaluate transferability of our trained time series input model. The models described above, trained on the Résif dataset, are used to detect events on a microseismic dataset from Larissa region in Greece [44], [17]. There are inherent differences between the two sites. The Résif dataset was collected from the Sauze catchment basin in the Alps, characterized by limestone formations and black marl [38]. The region around Larissa in Greece is a seismically active area, characterized by gneiss and schists [45]. The two sites are geologically different, with different monitoring networks: different number of sensors and deployment geometries, sampling at different rates (250Hz Résif vs. 100Hz Larissa). Such differences, especially the different geological background, alter the characteristics (amplitude, frequency content) of the signals, therefore making the transferability problem very challenging.

To transfer the model from the 6-channel Résif dataset, we selected six channels from the Larissa dataset from HT network as the array: all three components (North, East, vertical) of station TYR1, vertical (Z) components of TYR3, TYR6 and TYRN [17]. The choice of these particular stations, referred in the following as Array 1 (A1), was based on their location and quality of recordings. The four chosen stations form an almost equilateral triangular array, with three stations at the three vertices and one inside the triangle - see Fig. 6. This is a commonly used geometry in microseismic monitoring surveys to maximise detection of microseismic events.

To assess the sensitivity of the results to sensor array deployment geometry, we perform the same analysis as above on another set of stations forming a more random geometry, referred to as Array 2 (A2): all 3 components of TYR1, Z-components of TYR2, TYR3 and TYR4. Compared to the previous array geometry, the sensors in this array are located further apart - see Fig. 6 - and hence, waveforms from the same event are likely to look different in the recordings of the different stations.

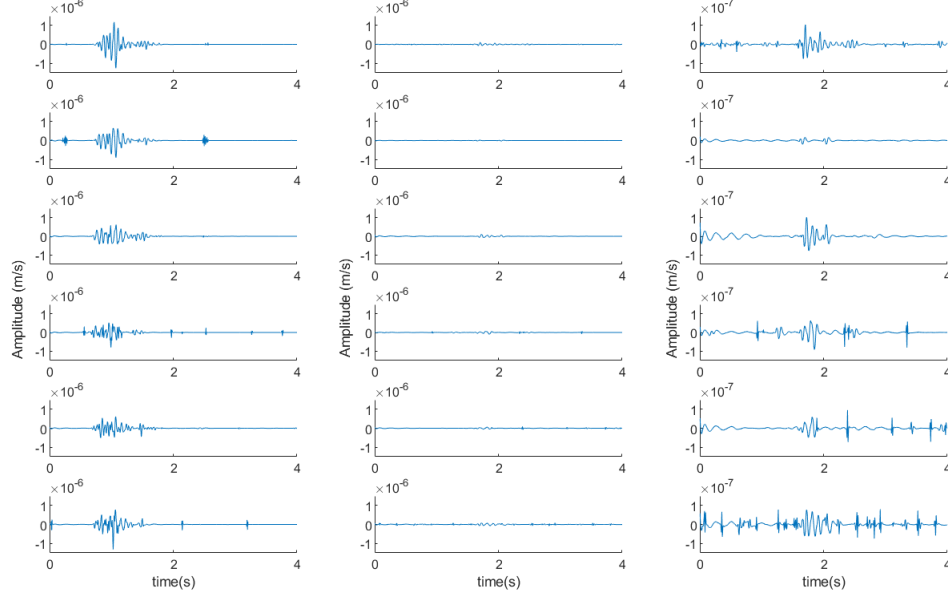


Fig. 4: Waveforms of two quake events in all six channels. The quake event which was correctly detected and classified (first column) and the quake event which was misclassified as noise (second column) with the same ordinate axis ranges from -1.5×10^{-6} m/s to 1.5×10^{-6} m/s. The misclassified quake event with the scaled zoomed (third column), its ordinate axis ranges from -1.5×10^{-7} m/s to 1.5×10^{-7} m/s.

TABLE VI: The confusion matrix for the two 'deepquake' models [25].

	'arch-time' network			'arch-spect' network		
	Earthquake	Other	Noise	Earthquake	Other	Noise
Earthquake	104	9	3	113	2	1
Other	6	162	22	1	175	14
Noise	3	23	79	1	19	85

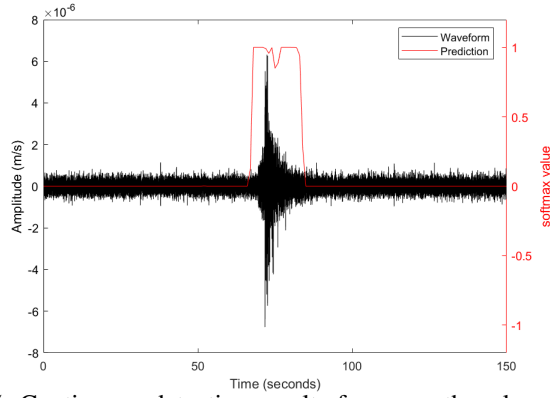


Fig. 5: Continuous detection results for an earthquake event.

Since the sampling rate of two datasets is different, we up-sampled the data from Larissa using linear interpolation to 250Hz and denoised the signal as described in Subsection III-A. The dataset from Larissa contains 86 catalogued earthquakes during the entire day of 17th March 2021, from 0:00 to 24:00. The catalogue we used can be downloaded from the Institute of Geodynamics, National Observatory of Athens [44] using as search area a circle with radius of 218km and centre at longitude 22.1777 degrees and latitude 39.6460 degrees. Our proposed models correctly classified **all** catalogued earthquakes, demonstrating good transferability.

Additional (to the aforementioned earthquake catalogued events) earthquake, quake and rockfall events were classified



Fig. 6: Six selected monitoring stations in Larissa, Greece. A1 configuration, based on an almost equilateral triangular array, comprises TRY1, TRY3, TRY6, and TRYN. A2 configuration, random geometry, is based on stations TRY1, TRY2, TRY3, and TRY4 that are further apart.

via our proposed approach. These findings are verified via manual detection, as well as the commercial InSite software v3.15. We randomly chose one hour within 17th March 2021, from 18:53 to 19:53 and manually searched the recordings of the same six channels to detect seismic events via visual observation. The data were filtered using a 5-100Hz band-pass filter (as per Subsection III-A). An event is valid if it was visually observed on at least 2 channels (at different stations). Note that manual detection of the events was blind,

i.e., without using the output of the proposed CNN model as guide. The start time is set as the time of the earliest arrival at any of the four stations of the array and duration is based on that station's channel. We fed the same 1 hour of continuous bandpass filtered data from Larissa to InSite software, where an amplitude threshold value of $6.6e-7m/s$ was used for detection of events.

An additional catalogue of all events detected and classified via the proposed CNN, manual detection and InSite software is produced as supplementary material for further analysis by the research community. This catalogue, which we refer to as comprehensive catalogue from now, also includes the type or class of event and its the duration. A1 and A2 refer to the events classified by the proposed CNN multi-classifier for the first (equilateral triangular) array and second array configurations, respectively. Similarly, M1 and M2 refer to events identified via manual detection on array 1 and 2, respectively, whilst I1 and I2 denote events identified by the InSite software.

The proposed CNN multi-classification model is designed for waveform pattern recognition (not P-wave picking as in [22]), therefore it does not estimate the start time and duration precisely because the model locates the event on a subset of the 6 input channels using a 10-sec window. After manual verification of traces, we observed that there are a number of occasions where the CNN model estimated multiple adjacent earthquakes as either one event or vice versa. For distant events from the station locations, the different signal phases (e.g., P wave, S wave etc.) arrive with a distinct time difference. This results in later phases, e.g., P wave reflections, being detected by the CNN model as separate earthquake events, instead of a single event. To mitigate this effect, we perform processing as follows. In the network's last FC layer, we set a bias towards classifying segments as non-events. In particular, we set as a decision threshold softmax value of 0.7 (instead of the default value of 0.5), which means that only when the softmax value greater than 0.7 is reached, the candidate window will be classified as an event. Next, as a post-processing step, we merge all events that start within 5-sec time interval, into one to prevent classifying different wave reflections into multiple events. Furthermore, for events that originated far away from the monitoring stations, the time interval between P-phase and S-phase will be large (>10 sec), and hence in this case, an earthquake event will appear as two separate events.

Referring to the proposed comprehensive catalogue, we compare events detected between time-series based on (1) CNN model with post-processing, (2) manual event detection and (3) automatic detection using the InSite software. It can be seen that the results of the proposed model and manual detection are very aligned. Indeed, only 8 earthquake events are detected manually by M1 and missed by A1, 4 of which were detected by A2. Similarly, 13 earthquakes detected by M2 are missed by A2, but 6 of these events were picked up by A1. These were more distant or more localised events, respectively. Each array configuration detected 62 earthquakes. 15 events were detected by A1 and missed by A2, or vice versa. InSite detected 32 and 38 of these earthquakes for A1 and A2, respectively, 22 of which are common for all 6

detection methods (2 arrays with manual, automatic detection and InSite). 9 (11) and 21 (8) rockfall (quake) events were detected by A1 and A2, respectively. Out of these, 7 and 4 rockfall events were detected by I1 and I2, respectively, 2 of which are common for all 6 detection methods.

Tables VII and VIII summarise the 3 types of events detected by the proposed approach (A1 and A2), that were manually corroborated (referred to as TP), missed by the proposed approach (FN) and not confirmed manually (FP). We observe that no quake and rockfall events detected by the proposed CNN were missed. Fewer earthquakes were missed by A1 than A2, since the A2 configuration includes stations that are spread further apart. Most earthquake events are detected by both A1 and A2 configurations, with some events of distant origin picked by A2 only, and localised events by A1 only. Details of the 3 distinct events and their time of occurrence, that were detected manually, via InSite Software and proposed CNN are provided in the comprehensive catalogue³.

TABLE VII: Larissa results from proposed CNN on one hour data using readings from TYR 1, 3, 6, and N stations (A1): Comparison between automatic and manual detection.

Class	No. of manually corroborated events (TP)	No. of events only observed by manual detection (FN)	No. of events only observed by CNN (FP)
earthquake	40	13	8
quake	10	0	3
rockfall	7	0	2

TABLE VIII: Larissa results from proposed CNN on one hour data using readings from TYR 1, 2, 3 and 4 stations (A2): Comparison between automatic and manual detection.

Class	No. of manually corroborated events (TP)	No. of events only observed by manual detection (FN)	No. of events only observed by CNN (FP)
earthquake	16	19	11
quake	2	0	2
rockfall	19	0	29

G. Explainability of internal workings of proposed architecture via feature maps vizualization

We visualize the feature maps at the output of different convolutional layers in our proposed network designs and visualize the features at the output of the second FC layer (which is the input to the output layer). The feature maps are the result of applying the filters to the input of convolution layers. Visualizing feature maps can be used to explain which input features are extracted in convolutional layers and analyse the influence of time, frequency, wavelet domain representation on the interior of the network. A visualization example is shown in Figure 7. The same earthquake event forms an input to all three proposed CNNs. The figure shows the first 9 feature maps output by the first, the second, and the fourth convolutional layer after max-pooling, for each of the three proposed models. For time series, we can see the input signal being transformed into many earthquake-like signals by the first convolution layer. Then in the second convolution

³<https://doi.org/10.15129/589f7af3-26b3-4a93-b042-fbc8100fc977>.

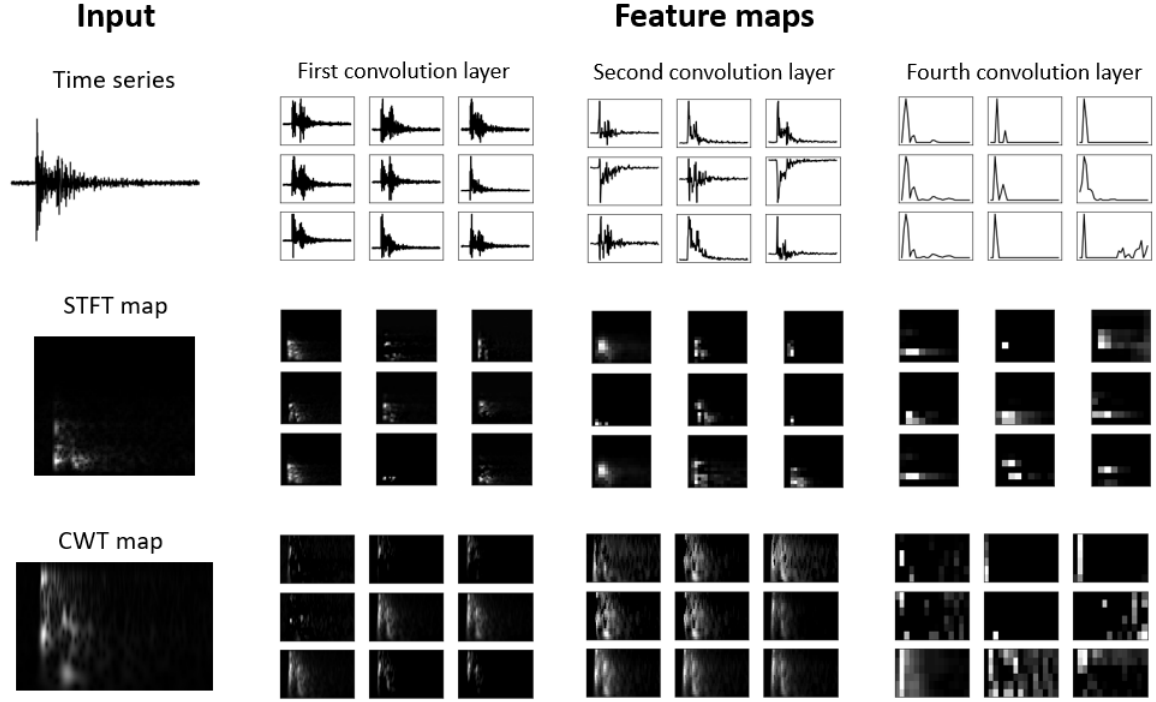


Fig. 7: Feature maps of the first, second, and fourth convolutional layer of the three CNNs with seismic events as input.

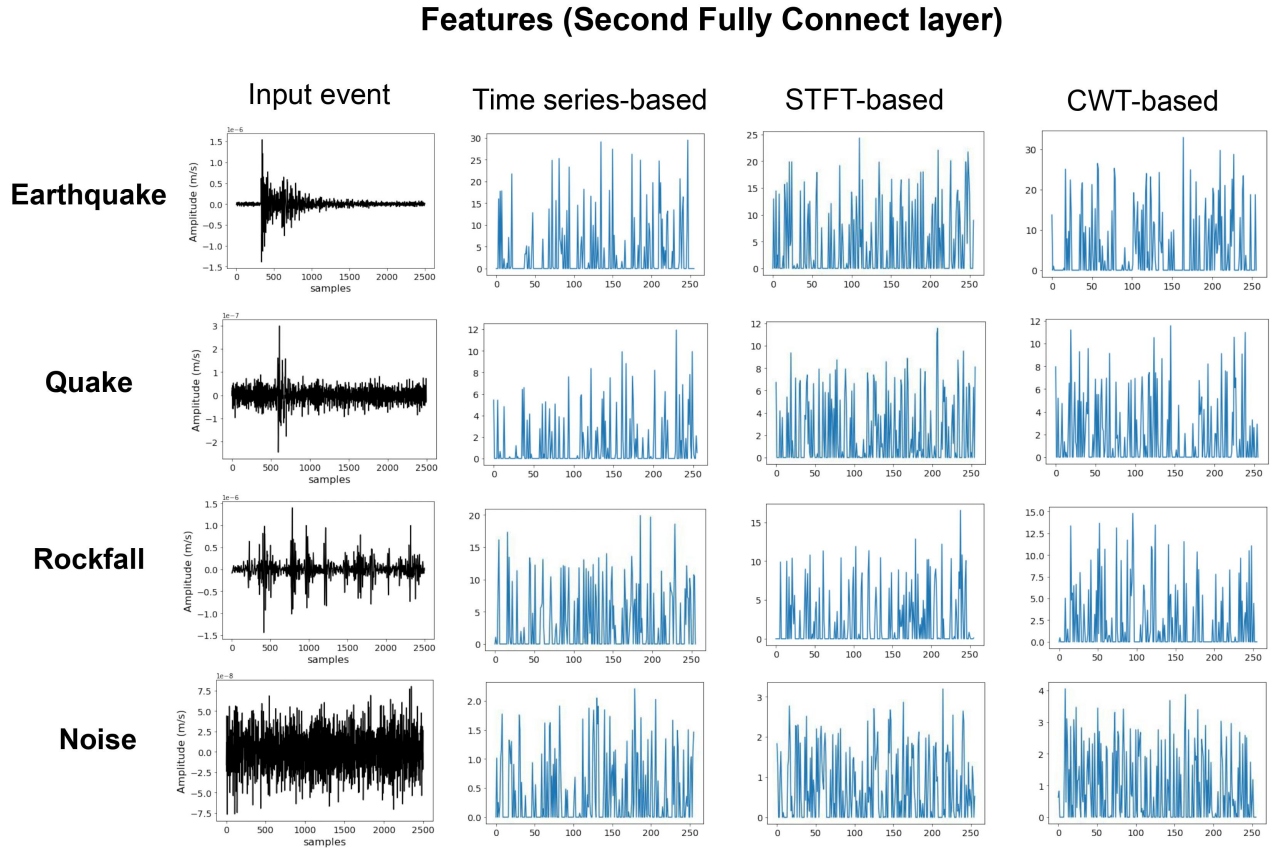


Fig. 8: Feature maps of the second fully connected layer of the three CNNs for four classes.

layer, the feature map extracts more detailed features from the input signals. The key features in earthquake signals (i.e., P-waves and S-waves) are preserved and enhanced in the fourth convolutional layer, where we can see clearly extracted peaks of events to be classified. In the STFT network, the first convolutional layer highlights the frequency band and the time step where the event occurred. From the second convolutional layer, it can be seen that the STFT-based model has learned the frequency features of the event, as the feature maps highlight the high-frequency and low-frequency range of the event. The fourth convolutional layer transforms these feature maps of different highlighted frequencies into more abstract representations. The CWT-based model works similarly to the STFT-based model. In the first convolutional layer, the feature maps highlight where, in time and scale, the event occurs. Compared with the feature maps of the first layer, the feature maps of the second layer have larger highlighted areas of the event, as the model is extracting more detailed wavelet features. Then, the fourth convolutional layers transform the features into abstract representations. In summary, through the visualization of feature maps, we explain the internal process of CNN in classifying seismic signals. The first convolutional layer is often used to find the position of the event in the input window. Then, CNN starts to extract more detailed features of the event. With the deepening of the convolutional layers, the extracted features are gradually transformed into abstract representations for subsequent classification. After the feature maps output by the convolutional layers are flattened, these one-dimensional features are classified by the FC layers. There are 256 nodes in the second FC, so there are 256 features. Figure 8 shows the features extracted by the second FC layer (which is the input of the last layer) for the four classes by three different models. For earthquake, the max value of the feature tends to be large (greater than 20), while quake and rockfall events have relatively small max features (between 10 and 20), and the max features of noise are the smallest (less than 10). As deep learning is a black-box algorithm, it is difficult to determine what attributes of the event these 252 features represent. However, we can see that different events have different feature distributions, and seismic events (earthquakes, quakes and rockfalls) have larger feature values than noise. This means that the CNN can extract different features for the seismic event waveforms, but does not extract many features from noise signals.

H. Complexity Analysis

Table IX shows the execution time needed to process one input window of 10 sec with 6 channels (equivalent to 15000 samples), including denoising, transform and each CNN model testing. CNN models were designed and tested using Python 3 and the Keras framework. The denoising and transform steps are also programmed in Python 3. All experiments were performed on an i5-10310U CPU. The time series-based model has the lowest complexity requiring 24ms to output result for a 10-sec window, followed by the STFT-based (37ms), and CWT-based model (808ms). This is expected since the time series- and STFT-based models have 5,684,036 trainable

parameters, while the CWT-based model has 6,035,012 parameters. Note that the time series-based and STFT-based models require much less than the 1sec window time shift applied on continuous data traces, and hence these two models can process the data in real time.

TABLE IX: The execution time for each model in ms.

10 seconds input Fs=250Hz, 6 channels 15000 samples	Approach	Time (ms)
Denoising	band-pass filter	2
Transform	STFT CWT	7 625
CNN models	Time series-based STFT-based CWT-based	22 28 181

V. CONCLUSIONS

The paper proposes microseismic classification on continuous recordings (no additional detection step needed) via a CNN, exploiting the inherent feature engineering ability of deep learning. Three CNN models were developed for three types of data inputs: temporal waveform, STFT and CWT maps. These proposed models were trained on the labelled Résif dataset, in order to detect/classify three types of events, namely earthquakes, quakes and rockfalls. During testing on an unseen portion of the Résif dataset, the time-series-, STFT- and CWT-based models all had similar performance for the three microseismic events and anthropogenic noise classes. Additionally, the time series-based model was observed to be the fastest during complexity analysis, demonstrating near real-time performance. The ability of our proposed pre-trained (on the Résif dataset) model to classify events from continuous recordings in a geologically distinct site was demonstrated via transferability to a 24hr dataset from the region of Larissa, Greece. All 86 catalogued earthquakes made available to us during that time period were correctly detected and classified despite major differences in monitoring layout used at the two sites and geological terrain. Using a less favourable deployment geometry, results were still within a satisfactory range. Additional (not catalogued previously) earthquake events, as well as rockfalls and quakes, were detected and classified by the proposed approach. These were corroborated via manual detection and detection via the commercial InSite software. The resulting catalogue of additional events is made available to the research community as supplementary material for further analysis. The implication of this study is that the time-consuming effort in manually labelling/cataloguing events is now automated on large unseen datasets, and weaker-signal events, such as quakes that are not always visible to the naked eye during manual observation, can be labelled automatically.

ACKNOWLEDGMENTS

We thank Dr. F. Provost for providing a catalogue of labelled events used in this paper, and Profs Papadimitriou and Karakostas, and Dr Kostoglou from the Aristotle University of Thessaloniki for providing access to and interpreting the dataset from Larissa. This work was supported in part by the EPSRC Prosperity Partnership research and innovation programme under grant agreement EP/S005560/1. For the

purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

REFERENCES

- [1] J. Gomberg, P. Bodin, W. Savage, and M. E. Jackson, "Landslide faults and tectonic faults, analogs?: The slumgullion earthflow, colorado," *Geology*, vol. 23, no. 1, pp. 41–44, 1995.
- [2] A. Tonnellier, A. Helmstetter, J.-P. Malet, J. Schmittbuhl, A. Corsini, and M. Joswig, "Seismic monitoring of soft-rock landslides: the super-sauze and valoria case studies," *Geophysical J. Int.*, vol. 193, no. 3, pp. 1515–1536, 2013.
- [3] M. Walter, J. Gomberg, W. Schulz, P. Bodin, and M. Joswig, "Slidequake generation versus viscous creep at softrock-landslides: synopsis of three different scenarios at slumgullion landslide, heumoes slope, and super-sauze mudslide," *J. Env. & Eng. Geophy.*, vol. 18, pp. 269–280, 2013.
- [4] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, "Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms," *J. Volcanology & Geoth. research*, vol. 180, no. 1, pp. 1–8, 2009.
- [5] C. Hibert, A. Mangueney, G. Grandjean, C. Baillard, D. Rivet, N. M. Shapiro, C. Satriano, A. Maggi, P. Boissier, V. Ferrazzini, et al., "Automated identification, location, and volume estimation of rockfalls at piton de la fournaise volcano," *J. Geophy. Research: Earth Surface*, vol. 119, no. 5, pp. 1082–1105, 2014.
- [6] A. E. Ruano, G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, and P. M. Ferreira, "Seismic detection using support vector machines," *Neurocomputing*, vol. 135, pp. 273–283, 2014.
- [7] M. Wenner, C. Hibert, A. van Herwijnen, L. Meier, and F. Walter, "Near-real-time automated classification of seismic signals of slope failures with continuous random forests," *Natural Hazards and Earth System Sciences*, vol. 21, pp. 339–361, 2021.
- [8] F. Provost, C. Hibert, and J.-P. Malet, "Automatic classification of endogenous landslide seismicity using the random forest supervised classifier," *Geophy. Research Let.*, vol. 44, no. 1, pp. 113–120, 2017.
- [9] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, "Automated platform for microseismic signal analysis: Denoising, detection, and classification in slope stability studies," *IEEE Trans. Geoscience & Remote Sensing*, 2020.
- [10] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances*, vol. 4, no. 2, p. e1700578, 2018.
- [11] R. Tibi, L. Linville, C. Young, and R. Brogan, "Classification of local seismic events in the utah region: A comparison of amplitude ratio methods with a spectrogram-based machine learning approach," *Bulletin Seism. Society America*, vol. 109, no. 6, pp. 2532–2544, 2019.
- [12] Y. Chen, G. Zhang, M. Bai, S. Zu, Z. Guan, and M. Zhang, "Automatic waveform classification and arrival picking based on convolutional neural network," *Earth & Space Science*, vol. 6, pp. 1244–1261, 2019.
- [13] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection," *Scientific reports*, vol. 9, no. 1, pp. 1–14, 2019.
- [14] W. Zhu and G. C. Beroza, "Phasenet: a deep-neural-network-based seismic arrival-time picking method," *Geophysical J. Int.*, vol. 216, no. 1, pp. 261–273, 2019.
- [15] B. Ku, J. Min, J.-K. Ahn, J. Lee, and H. Ko, "Earthquake event classification using multitasking deep learning," *IEEE Geoscience & Remote Sensing Let.*, 2020.
- [16] FrenchLandslideObservatorySeismologicalDatacenter/RESIF, "Observatoire multi-disciplinaire des instabilités de versants (omiv)." Accessed: 2021. [online]. Available: <https://seismology.resif.fr/>. DOI: 10.15778/RESIF.MT.
- [17] EDIA. Accessed: 2021. [online]. Available: <http://www.orfeus-eu.org/data/eida/>.
- [18] X. Liao, J. Cao, J. Hu, J. You, X. Jiang, and Z. Liu, "First arrival time identification using transfer learning with continuous wavelet transform feature images," *IEEE Geoscience & Remote Sensing Let.*, vol. 17, pp. 2002–2006, 2019.
- [19] G. Zhang, C. Lin, and Y. Chen, "Convolutional neural networks for microseismic waveform classification and arrival picking," *Geophysics*, vol. 85, no. 4, pp. WA227–WA240, 2020.
- [20] R. M. Dokht, H. Kao, R. Visser, and B. Smith, "Seismic event and phase detection using time–frequency representation and convolutional neural networks," *Seismological Research Let.*, vol. 90, pp. 481–490, 2019.
- [21] L. Linville, K. Pankow, and T. Draelos, "Deep learning models augment analyst decisions for event discrimination," *Geophy. Research Let.*, vol. 46, no. 7, pp. 3643–3651, 2019.
- [22] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Commun.*, vol. 11, pp. 1–12, 2020.
- [23] O. M. Saad, Y. Chen, A. Savvaidis, S. Fomel, and Y. Chen, "Real-time earthquake detection and magnitude estimation using vision transformer," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 5, p. e2021JB023657, 2022.
- [24] O. M. Saad, G. Huang, Y. Chen, A. Savvaidis, S. Fomel, N. Pham, and Y. Chen, "Scalodeep: A highly generalized deep learning framework for real-time earthquake detection," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 4, p. e2020JB021473, 2021.
- [25] L. Trani, G. A. Pagani, J. P. P. Zanetti, C. Chapeland, and L. Evers, "Deepquake—an application of cnn for seismo-acoustic event classification in the netherlands," *Computers & Geosciences*, vol. 159, p. 104980, 2022.
- [26] M. Shakeel, K. Itoyama, K. Nishida, and K. Nakadai, "Emc: Earthquake magnitudes classification on seismic signals via convolutional recurrent networks," in *IEEE/SICE Int. Symp. Sys. Integrat.*, pp. 388–393, 2021.
- [27] L. Liu, W. Song, C. Zeng, and X. Yang, "Microseismic event detection and classification based on convolutional neural network," *J. Applied Geoph.*, p. 104380, 2021.
- [28] L. Linville, R. C. Brogan, C. Young, and K. A. Aur, "Global-and local-scale high-resolution event catalogs for algorithm testing," *Seism. Research Let.*, vol. 90, no. 5, pp. 1987–1993, 2019.
- [29] U. M. Park, "Usgs northern california network," *International Federation of Digital Seismograph Networks, Dataset/Seismic Network*, doi, vol. 10, 1967.
- [30] S. Horton, "Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central arkansas with potential for damaging earthquake," *Seismological Research Letters*, vol. 83, no. 2, pp. 250–260, 2012.
- [31] NECIS. Accessed: 2021. [online]. Available: <http://necis.kma.go.kr/>.
- [32] IRIS. Accessed: 2021. [online]. Available: <https://www.iris.edu/>.
- [33] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "Stanford earthquake dataset (stead): A global data set of seismic signals for ai," *IEEE Access*, vol. 7, pp. 179464–179476, 2019.
- [34] A. Savvaidis, B. Young, G.-c. D. Huang, and A. Lomax, "Texnet: A statewide seismological network in texas," *Seismological Research Letters*, vol. 90, no. 4, pp. 1702–1715, 2019.
- [35] Y. Duan, Y. Shen, I. Canbulat, and G. Si, "Classification of clustered microseismic events in a coal mine using machine learning," *J. Rock Mechanics and Geotech. Eng.*, 2021.
- [36] KNMI, "Netherlands seismic and acoustic network," *Royal Netherlands Meteorological Institute (KNMI), Other/Seismic Network*. (10.21944/E970fd34-23b9-3411-B366-E4f72877d2c5), 1993.
- [37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [38] J.-P. Malet, D. Laigle, A. Remaître, and O. Maquaire, "Triggering conditions and mobility of debris flows associated to complex earthflows," *Geomorphology*, vol. 66, no. 1–4, pp. 215–235, 2005.
- [39] A. Helmstetter and S. Garambois, "Seismic monitoring of séchilienne rockslide (french alps): Analysis of seismic signals and their correlation with rainfalls," *J. Geoph. Res.: Earth Surface*, vol. 115, no. F3, 2010.
- [40] M. Walter, C. Arnhardt, and M. Joswig, "Seismic monitoring of rock-falls, subsurface fracture processes, and superficial fissure development at the super-sauze, french alps, mudslide," *Eng. Geology*, 2011.
- [41] Z. E. Ross, M.-A. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learningshort note," *Bulletin of the Seismological Society of America*, vol. 108, no. 5A, pp. 2894–2901, 2018.
- [42] O. M. Saad and Y. Chen, "Earthquake detection and p-wave arrival time picking using capsule neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6234–6243, 2020.
- [43] O. M. Saad and Y. Chen, "Capsphase: Capsule neural network for seismic phase classification and picking," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.
- [44] G.I. Accessed: 2021. [online]. Available: <http://www.gein.noa.gr/en/>.
- [45] V. Karakostas, C. Papazachos, E. Papadimitriou, M. Fomelis, A. Kiratzi, C. Pikridas, A. Kostoglou, C. Kkallas, N. Chatzis, S. Bitharis, et al., "The march 2021 tyrnos, central greece, doublet (mw6.3 and mw6.0): Aftershock relocation, faulting details, coseismic slip and deformation," *Bulletin of the Geological Society of Greece*, vol. 58, pp. 131–178, 2021.