Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Full length article

A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

Jiaxin Jiang [a,*], David Murray [a], Vladimir Stankovic [a], Lina Stankovic [a], Clement Hibert [b], Stella Pytharouli [c], Jean-Philippe Malet [b]

[a] *Department Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK*
[b] *Institut Terre & Environnement de Strasbourg, University of Strasbourg, Strasbourg, France*
[c] *Department Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK*

ARTICLE INFO

Dataset link: https://github.com/kanata2020/Explainable-seismic-classification

Keywords:
Seismic signal analysis
Microseismic signal classification
Deep learning
Explainable artificial intelligence
Data annotation
Model training

ABSTRACT

With the increased frequency and intensity of landslides in recent years, there is growing research on timely detection of the underlying subsurface processes that contribute to these hazards. Recent advances in machine learning have introduced algorithms for classifying seismic events associated with landslides, such as earthquakes, rockfalls, and smaller quakes. However, the opaque, “black box” nature of deep learning algorithms has raised concerns of reliability and interpretability by Earth scientists and end-users, hesitant to adopt these models. Leveraging on recent recommendations on embedding humans in the Artificial Intelligence (AI) decision making process, particularly training and validation, we propose a methodology that incorporates data labelling, verification, and re-labelling through a multi-class convolutional neural network (CNN) supported by Explainable Artificial Intelligence (XAI) tools, specifically, Layer-wise Relevance Propagation (LRP). To ensure reproducibility, a catalogue of training events is provided as supplementary material. Evaluation from the French Seismologic and Geodetic Network (Résif) dataset, gathered in the Alps in France, demonstrate the effectiveness of the proposed methodology, achieving a recall/sensitivity of 97.3% for rockfalls and 68.4% for quakes.

1. Introduction

Seismic signal analysis is based on collecting, processing and performing inference on seismic signals with the goal of detecting, understanding, classifying and locating seismic events, including not only earthquakes, but also rockfalls and smaller quakes or tremors that characterise landslides and their severity. The devastating effects of landslides on humans and infrastructure have been making headlines, and more recently have been often attributed to extreme weather and/or human activities. Seismometers provide accurate recordings of mechanical waves originating from various sources, but due to their high sensitivity, distinguishing between mechanical waves originating from tectonic activities and any other signals contained in the recordings (e.g., rainfall, animals, traffic, natural noise, machinery, etc.) is not an easy task. Manually identifying events based on recordings of seismometers is a time-consuming and subjective task, prone to errors and bias. Thus, manual detection has gradually been replaced by methods that automatically detect and classify seismic events. With higher availability in seismic recordings and advances in AI, seismic signal analysis has become a very much data-driven field and has spread well beyond seismology and geoscience, as it is now of interest to much broader research communities (Mousavi and Beroza, 2022).

Deep learning has been shown to be achieve excellent detection and classification performance for a range of applications where sufficient amount of labelled data is available, including automated road extraction (Bayramoğlu and Uzar, 2023), pneumonia diagnosis from medical imaging (Gülgün and Erol, 2020), satellite image analysis (Sariturk et al., 2020; Dos, 2022), and car detection (Kaya et al., 2023). Due to the availability of many well-maintained datasets, the number of deep learning approaches used in seismology has also sky-rocketed in recent years (see Fig. 1 in Mousavi and Beroza, 2022) using enormous amounts of data to train the models. Consequently, recent literature is dominated by deep learning techniques applied to diverse tasks such as seismic event labelling using Residual Neural Network (ResNet) (Yi et al., 2021), magnitude estimation using a network that combines CNN and Recurrent Neural Networks (RNN) (Shakeel et al., 2021), event localisation using CNN architectures (Perol et al., 2018), multitask learning for classification with velocity models (Li et al., 2023b) and tackling seismic inversion problems with conditional Generative Adversarial

[*] Corresponding author.
E-mail address: jiaxin.jiang@strath.ac.uk (J. Jiang).

https://doi.org/10.1016/j.srs.2024.100189
Received 20 August 2024; Received in revised form 19 November 2024; Accepted 25 December 2024
Available online 2 January 2025

Networks (GAN) (Parasyris et al., 2023). A detailed review of deep learning architectures, specifically proposed, for event classification from seismic recordings can be found in Jiang et al. (2023).

For example, CNN-based model 'DeepQuake' (Trani et al., 2022) has demonstrated robust performance for high-magnitude earthquakes, though it has limitations with microseismic events, as demonstrated in Jiang et al. (2023). In Liao et al. (2022), RockNet, taking both 3-channel time series window and a spectrogram of the vertical channel of the window as inputs, is proposed for classifying rockfalls and earthquakes. The deep learning models achieve state-of-the-art performance in detecting and classifying seismic signals avoiding cumbersome manual feature generation, selection and extraction process, with their ability to automatically learn most discriminative features from raw recordings. However, this also means that these models are limited by the used training set, and may learn specifically spurious correlations with the prediction target (Soneson et al., 2014; Hägele et al., 2020). Furthermore, the fact that the feature engineering task is taken away from the designer, makes deep learning models opaque, and hence often referred to as "black box", which limits their use. Indeed, geoscientists are still reluctant to use them and rather rely on less complex interpretable methods based on hand-crafted features (Li et al., 2020) that ensure that relevant physical features are used for detection and classification (see, e.g., Table I in Li et al. (2020) and Table A1 in Li et al., 2023a).

Explainable artificial intelligence (XAI) (Bau et al., 2020; Holzinger, 2018), is a research direction that provides human-interpretable explanations that can potentially enhance training process, correct manual data annotation, improve models, and contribute towards building trust in AI-generated outputs (Samek et al., 2016; Montavon et al., 2022). XAI tools have been extensively used in computer vision (e.g., Lapuschkin et al., 2016) and time-series signal analysis problems (e.g., Murray et al., 2021); however, the work on explaining the output of deep learning models for seismic signal analysis, and using these explanations to improve confidence in data labelling, model training and building trust in inferred outputs, is still in its infancy.

In order to pave the way towards a regulatory framework for ensuring trust in AI, the European Commission has published seven principles of Trustworthy AI (European Commission and Directorate-General for Communications Networks, Content and Technology, 2019), which include Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination and Fairness, Societal and Environmental Well-Being and Accountability.

Depending on how the AI-based seismic analysis will be used, from understanding the subsurface processes and mechanics to hazard and disaster management, the AI systems can be seen as minimal risk to high risk, and therefore subject to strict oversight before they can be used to ensure infrastructure and human safety. Therefore, the following principles are important for seismic analysis. First, AI systems should empower decision makers when it comes to hazard assessment or infrastructure planning, allowing them to make informed decisions from the AI system outputs. The principle of Human Agency and Oversight caters for proper oversight mechanisms that need to be ensured, which can be achieved through human-on-the-loop and human-in-command approaches. Second, the principle of technical robustness and safety, in part states that AI systems need to be accurate, reliable and reproducible to ensure unintentional harm can be minimised and prevented. Accuracy refers to the ability to correct predictions based on AI models and can be implemented via rigorous evaluation and indication of likelihood of potential errors. Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions. A reliable AI system is one that works properly with a range of inputs and in a range of situations. Third, the principle of privacy and data governance enables users to trust the data gathering process and that it does not contain inaccuracies, errors or mistakes, especially with respect to labelling or cataloguing by expert geoscientists. Fourth, the principle of transparency states that the data and AI system should be transparent through traceability mechanisms in the form of documentation of datasets and processes that yielded in decision, including data gathering, data labelling and algorithms used. Furthermore, transparency also includes explainability, that is, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. This includes XAI. Fifth, transparency also states that humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations. Finally, the social and environmental well-being principle state that the AI systems should be sustainable and environmentally friendly — this can be through taking into considering the resource usage and energy consumption for training the models. Moreover, they should consider the societal impact. Monitoring, understanding, modelling and predicting landslide processes due to climate change, especially rainfall, tackle United Nations (UN) Sustainable Development Goal (SDG) 13 on Climate Action (United Nations Sustainable Development Goal 13, 2023). As explained in Vouillamoz et al. (2018), shearing and friction between the soil grains results in release of seismic energy within the landslide body. Therefore, passive seismic monitoring is a good approach to monitor and mitigate slope instabilities, as it provides high temporal resolution data in near real time that relate to the dynamics of the landslide. This means that the transition (and rapid transformation) of the landslide from slow rate sliding into a rapid slope failure may be detected and therefore mitigate associated hazards.

2. Literature review on trustworthy AI for seismic signal analysis

To ensure trust and expert's control of the decision process, machine learning-based seismic signal analysis has been performed either in a semi-automated manner (Renouard et al., 2021) using continuous expert oversight and monitoring (human-on-the-loop), using interpretable models (Li et al., 2020), or using non-interpretable models (such as Random Forests) but with numerous hand-crafted features (Provost et al., 2017) to ensure that the inference is made on signal characteristics identified by experts as important. In Li et al. (2023a) a detailed study of feature importance is presented where 119 features are constructed based on seismic signal literature and their importance tested using four different feature importance methods and different classifiers based on Support Vector Machine, Random Forest, and three graph signal processing based semi-supervised approaches. The features are experimentally ranked showing time-, frequency-, cepstrum and polarity features that are of highest importance in inference making per studied class. The results show that out of 119 constructed features only a subset contributed significantly to the decision. Note that this study was based on quantifying the importance of hand-crafted features in accurately classifying multiple event classes from continuous data, thus deep learning networks were not considered.

In Trani et al. (2022), convolutional neural networks (CNNs) are used to classify isolated catalogued seismic events into noise, earthquake and other events. The authors developed a heatmap-based visualisation tool to explain model outputs via the outputs of activation functions of each filter in the convolutional layers and then overlapping the result with the raw input signal. However, this study has several weaknesses when it comes to gaining trust in model outputs. Firstly, it is not clear how explanations are formed by fusing outputs of the activation functions from different layers. Secondly, only binary classification is considered, i.e., identifying relatively well-defined earthquakes from other signals. Thirdly, the approach does not exploit advanced XAI methods, and it is not used to explain any false predictions.

In Bi et al. (2021), the authors proposed a Dual-Channel CNN Module where one channel contains raw time-domain waveforms, and the other channel contains frequency-domain information by Discrete Cosine Transform (DCT) to classify input seismic waveform into rock fracturing and noise, together with an explanation module, EUG-CAM

(Elaborate Upsampling-based Gradient-weighted Class Activation Mapping). It builds upon the principles of the gradient weighted class activation mapping (GradCAM) (Selvaraju et al., 2019), harnessing the influence of feature map values and gradients to elucidate the importance of diverse features in the last convolutional layer. Recognising the discrepancy between feature map sizes and input data dimensions, EUG-CAM uses a strategic amalgamation of transposed convolution, unpooling, and interpolation, to generate feature mappings from a coarse localisation map. This results in an explanation feature map that effectively encapsulates class activation, learning insights, and network architecture considerations. However, the model's limitation is in classifying only two classes (rock fracturing vs. noise) and its confinement to binary classification. Furthermore, the reliance on a 1-D CNN model facilitates explanations primarily within the time domain, possibly neglecting the benefits of frequency-domain insights garnered from the DCT. Additionally, the visualisation maps cannot show the adverse input signal influence (negative contribution) on classification results, hampering a comprehensive and well-rounded comprehension of the model's decision-making process.

In Jiang et al. (2023), the authors present CNN models with six channel inputs for multi-class classification of earthquakes, quakes, rockfalls and noise and use visualisation of feature maps to understand the network's internal workings. The authors examine feature maps at various convolutional layers and the second fully connected (FC) layer, gaining insights into feature extraction. Different models, including time-series, Short-time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT)-based designs, highlight the network's focus on time, frequency, and wavelet characteristics. The main observation is that early layers locate event positions and extract basic features, while deeper layers refine these features into abstract representations for classification. The second FC layer's feature distributions vary across seismic events, indicating the network's capability to distinguish three event types from noise based on learned features. In addition, Layer-wise Relevance Propagation (LRP) showed promising results in identifying the most relevant features for each class, further enhancing the interpretability of the model (Jiang et al., 2024).

2.1. Contributions

The goal of this paper is to provide comprehensive explanations to identify key features learnt by a deep neural network for multi-class classification, demonstrate that these features are in agreement with the physical properties of seismic signals and common hand-crafted features used in the literature (Li et al., 2020). The generated explanations are then used to explain instances of misclassifications and correct errors in manual labelling, jointly with a geoscientist, who verified the corrected labels of the classified events and the features associated with these events. This builds trust in the models confirming that the learnt feature representations agree with expert knowledge.

We use state-of-the-art XAI tools to explain deep learning models for detection and classification of micro-seismic signals and show how these explanations can be used to improve the designs and explain correct and wrong predictions. In particular, we use a CNN-based architecture with a frequency-domain input, for detection and classification of seismic signals into four classes: earthquake, micro-earthquake referred to as quake, rockfall and noise. These are the same classes as used in Jiang et al. (2023) and Provost et al. (2017). There are three inputs to the CNN, each comprising continuous recordings from the channels of a typical three-component seismometer, usually deployed for seismic monitoring.

Our models are trained and tested on a publicly accessible dataset Résif (FrenchLandslideObservatorySeismologicalDatacenter/RESIF, 2021) that has over 1000 labelled events, including earthquakes, quakes, rockfalls and anthropological noise. After classification, we use Layer-wise Relevance Propagation (LRP) (Bach et al., 2015) to explain the decision making process. We analyse the basis of the model for event classification and communicate the reasons for misclassification of individual events. Furthermore, if the predicted class is different to the expert label, and after inspection of the filtered signal, its STFT and LRP map, the event is sent back to the expert for re-labelling. This protocol is used to correct possible labelling mistakes in the large annotated seismic dataset.

In summary, our main contributions are:

1. ensuring data integrity by leveraging on a well-maintained ongoing seismological data portal releasing checked seismic recordings publicly, as well as cataloguing/labelling by expert geoscientists — this aspect is by nature transdisciplinary
2. traceability to enable transparency by leveraging on public datasets, where data gathering, labelling and performance with different algorithms are well documented
3. an additional catalogue of 829 labelled events for a period of 3 days, classified by the CNN, verified by an expert and labels corrected — provided as supplementary material
4. reproducibility by releasing the catalogue of 822 manually selected high quality training events as supplementary material
5. designing a multi-classifier robust to noisy continuous recordings for high performance but also indicating likelihood of potential errors
6. reliability of design by ensuring that the multiclassifier design works for a continuous input stream with noisy signals and low signal to noise ratio events
7. explainability for transparency by providing explanations of the multi-classifier outputs via XAI LRP maps
8. communication for transparency by clearly identifying the level of performance and limitations
9. tackling the UN SDG 13 by accurately detecting landslide related events that helps build trust in precursors to landslides such as rockfalls and quakes

The first three contributions are presented in Section 3, where we describe the dataset used and data pre-processing. Contributions (4)–(5) are covered in Section 4, where the proposed CNN-based architecture, the sliding-window continuous detection method, the proposed post-processing and explainability tools used are described. Section 5 demonstrates our contributions (6)–(8). The significance of this work, i.e., contribution (9) was discussed above and is demonstrated in Section 5. Finally we conclude in Section 6 with suggestions for further work.

3. Dataset

In this section we provide details about the data management, including collection, storage, release and labelling /cataloguing, describing the first three contributions of this paper.

3.1. Data gathering and context

Our raw seismometer recordings are obtained from the publicly accessible Résif Seismological Data Portal, acquired by the French Landslide Observatory OMIV (Observatoire Multi-disciplinaire des Instabilités de Versants). In particular, we focus on the Super-Sauze (SZ) slow moving landslide monitoring array, acquired by the Super-Sauze C (SZC) station of the French Landslide Observatory on the Permanent seismological records on unstable slopes which are installed at the centre of the Super-Sauze landslide deposit in Southeast France (Latitude: 44.34787°N, Longitude: 6.67805°E). The location of the SZC station is shown on the map in Fig. 1. The seismometer array consist of one central three-component sensor and three vertical one-component sensors (organised as equilateral triangle), all recording at 250 Hz sampling rate. In this project, we used data from the

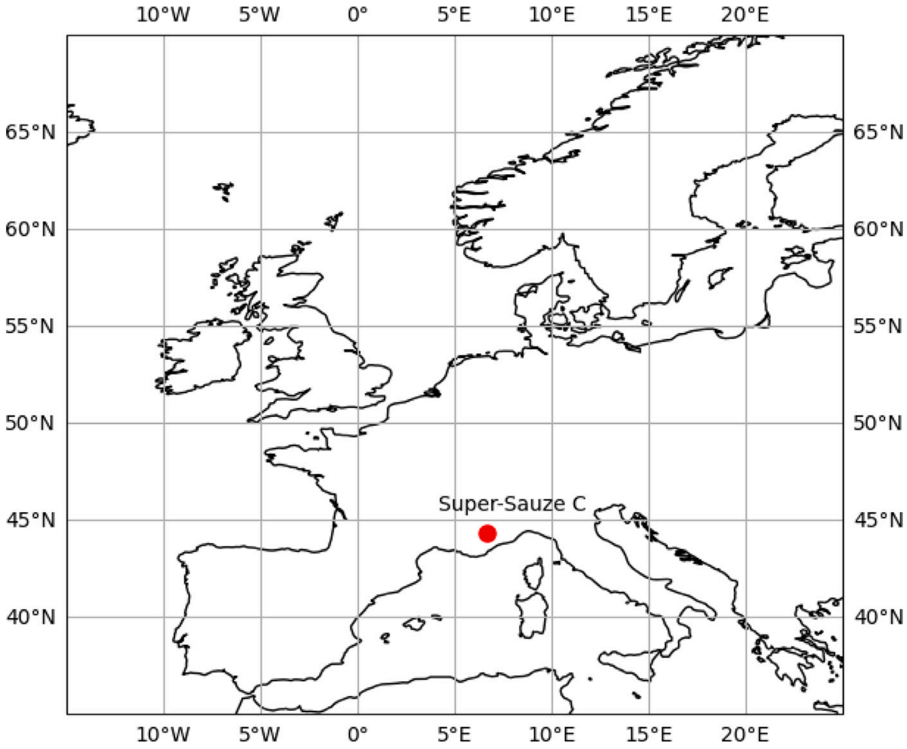


Fig. 1. Map showing the location of the Super-Sauze C (SZC) station.

three-component sensor. This choice aligns with common practices in seismic waveform classification, where a 3-channel input is standard, such as EQ-transformer (Mousavi et al., 2020) and DeepQuake (Trani et al., 2022). Additionally, it facilitates transfer learning, as many seismometers employ three-component sensors, ensuring compatibility with various seismic datasets and applications. Using 3 channels also reduces the number of false positives which can occur with arrival mismatches and reduces the computational demand. The seismometers recorded three periods: from 11 Oct. to 19 Nov. 2013; from 10 Nov. to 30 Nov. 2014; and from 9 June to 15 Aug. 2015.

The description of the SZ slope deformation, together with the challenges of detecting the microseismic events is well documented in Provost et al. (2018b). Additionally, description of how the catalogue of events was generated is documented in Provost et al. (2017), where events were detected by the STA/LTA algorithm applied in the frequency domain, and manually labelled into four classes: earthquake, quake (micro-earthquake events), rockfall and natural/anthropogenic (N/A) noise. All events except noise are classed as microseismic according to Vouillamoz et al. (2018).

Rockfalls mainly occur at the main scarp of the landslide, where the rigid block falls from the steep slope (height > 100 m). The quake is likely to be triggered by material damage, surface cracks and openings within the landslide main flow. The earthquakes class includes regional seismic events in this area and the teleseisms (global large magnitude earthquakes). N/A noise events include all anthropogenic and environmental noise, due to, e.g., transportation, pedestrian walking, heavy rain, animals, strong wind, etc. It does not include noise in the form of instrumentation error.

3.2. Labelling

The SZ recordings over the data gathering duration described in the previous subsection were labelled as described in Provost et al. (2017), using STA/LTA in the frequency domain to pick events, and manual labelling of these events by an expert based on their amplitude, duration, spectrogram and location. The number of labels in this catalogue, which will be referred to as the original catalogue, for each class, is reported in Li et al. (2023a) and Jiang et al. (2023), where the events were classified on continuous recordings with classifiers using manual feature generation, and deep-learning-based classifiers with automated feature extraction, respectively. Since detection and classification were performed on the continuous data stream, the Normalised Graph Laplacian Regularisation (normGLR)-based (Li et al., 2023a) and CNN-based (Jiang et al., 2023) classifiers also reported classification of hundreds of additional non-catalogued events, with a high density of events in the period 25th to 28th Nov. 2014, which coincided with a period of high activity on the SZ slope (Provost et al., 2018a).

As reported in Li et al. (2023a), all four types of events are present in this 4-day time period, and in addition to the 120 events (65 rockfalls, 18 quakes, 23 earthquakes and 14 noise) labelled in the original catalogue, 17 quakes, 89 earthquakes and 92 rockfalls events were detected and classified by the normGLR classifier whereas an additional 260 quakes, 174 earthquakes and 32 rockfalls were detected and classified with the CNN approach of Jiang et al. (2023). These algorithms only missed 1 earthquake, 1 rockfall and 2 noise events that were present in the original catalogue.

All events detected by the normGLR classifier, the CNN classifier and an additional classifier based on Siamese networks (Murray et al., 2023) were reviewed by an expert for labelling following the methodology used to build the original catalogue, which is based on the seismic signal waveform and spectrogram features. The final outcome of the expert reviews for this 4-day period were 69 quakes, 29 earthquakes and 126 rockfalls. Note that the normGLR classifier was too sensitive, overestimating the number of earthquakes (Li et al., 2023a). The CNN-based 6-channel input multi-classifier of Jiang et al. (2023) was too sensitive for quakes and earthquakes but missed a number of rockfalls.

This exercise demonstrated the value of machine learning-based classification on continuous streaming recordings, since it is tedious for experts to manually review continuous data streams, as well as pick up the microseismic events, especially quakes and rockfalls, that are often “hidden” or “unclear” within ambient noise present in the recordings. These newly detected and expert-labelled events during the period 25th to 28th Nov. 2014, not present in the original catalogue, are released with this paper and are focus of this study.

4. Methodology

In this section, we describe our methodology. First, building on our prior work (Jiang et al., 2023), we propose an improved multi-class CNN-based classifier that utilises 3-channel inputs and a modified training strategy (see Section 4.3) to enhance precision in detecting quakes and earthquakes, as well as improve recall/sensitivity rates for rockfalls. Second, we analyse the outputs of the improved multi-classifier, as part of our human-on-the-loop contribution to verify instances of labelling error, likely to occur for large volumes of continuous streaming seismic recordings. This is carried out via the XAI-based LRP tool to visualise the features of misclassifications, which are then queried for re-evaluation by the expert.

4.1. Proposed CNN-based architecture

An STFT-based CNN model, inspired by VGGNet (Simonyan and Zisserman, 2015) and adapted from Jiang et al. (2023), is used. The influence of seismometer characteristics such as sensitivity, frequency band, and axis configuration on the reliability and effectiveness of our results was explored in Jiang et al. (2023), whereby good transferability was demonstrated with recordings from different seismometers with varying sensitivity levels and sampling rates, and geographic location. Additionally, we examined the performance impact of different seismometer configurations, comparing one-axis (single-channel) seismometers with multi-channel inputs during training. We use STFT maps as inputs for the CNN model, as these inputs were shown to provide better results on average compared to directly feeding time-series signals. Additionally, the generalisability and robustness of this architecture across different sites has been demonstrated in prior work (Jiang et al., 2023). Particularly, as evidenced by the recent trend in CNN-based architectures for analysis of seismic recordings, such networks excel in extracting hierarchical and discriminative features from complex data, making them highly effective for seismic event classification. The value of binary vs multi-class networks in terms of how multi-class models are able to achieve similar performance while requiring less models to be trained and run, and hence lower overall complexity, was demonstrated in Jiang et al. (2023). Multi-class CNN models offer enhanced feature extraction, adaptability to various data patterns that are often indistinguishable (such as local quakes and rockfalls), and improved classification performance compared to state-of-the-art DL approaches for seismic analysis, discussed in Introduction Section, that mostly focus on binary classification.

The architecture of the model is composed of convolutional layers, max pooling layers and FC layers, adapted to the input shapes and output categories, as shown in Fig. 2. Convolutional layers perform feature representation and extraction, followed by max-pooling layers that downsample the extracted feature into a feature map with smaller size.

Compared to Jiang et al. (2023), to effectively process long-duration seismic events within continuous data streams, we increase the input window of the CNN model to 15 s (from 10 s). We also reduce convolution kernels and neural nodes in each layer, achieving a balance between model complexity and performance. Moreover, recognising the prevalence of waveforms captured by three-component sensors, the input to the network is 3-channel input data, in contrast to 6-channel used in Jiang et al. (2023), which significantly expands the model’s applicability across a wider range of scenarios.

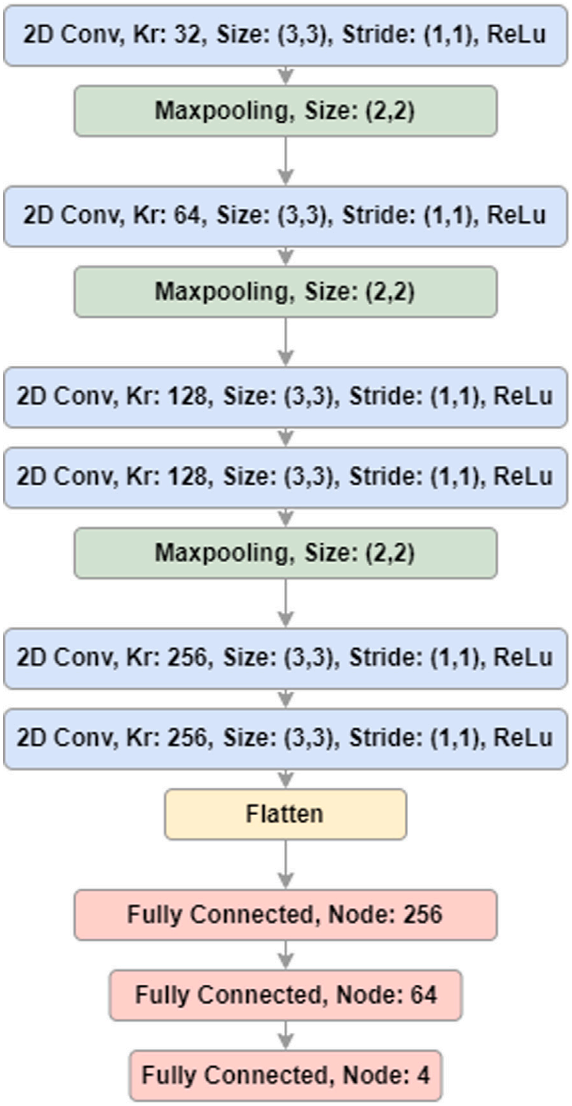

2D Conv, Kr: 32, Size: (3,3), Stride: (1,1), ReLu
Maxpooling, Size: (2,2)
2D Conv, Kr: 64, Size: (3,3), Stride: (1,1), ReLu
Maxpooling, Size: (2,2)
2D Conv, Kr: 128, Size: (3,3), Stride: (1,1), ReLu
2D Conv, Kr: 128, Size: (3,3), Stride: (1,1), ReLu
Maxpooling, Size: (2,2)
2D Conv, Kr: 256, Size: (3,3), Stride: (1,1), ReLu
2D Conv, Kr: 256, Size: (3,3), Stride: (1,1), ReLu
Flatten
Fully Connected, Node: 256
Fully Connected, Node: 64
Fully Connected, Node: 4

Fig. 2. STFT-based CNN for seismic classification. Kr denotes the number of kernels, and ‘Flatten’ function transforms the input data into a 1D array.

4.2. Sliding window-based detection

Raw signals recorded by 3-channel (North, East and vertical direction) seismic recorders are used. Since the classes of interest are 5-60 Hz bandwidth, we first use a band-pass filter to remove low frequency noise (denoising) as in Jiang et al. (2023). To allow prediction on a continuous stream of signals, a sliding window method is used to segment the continuous stream into smaller windows as in Saad and Chen (2020, 2021). The window size and overlap are selected based on the temporal resolution required for the events of interest. A window size of 3750 samples (i.e., 15 s) is used. The overlap between consecutive windows is set to 93% of window size (3500 samples (14 s)), which corresponds to a shift by 1 sec, allowing the CNN model to capture the temporal dynamics of the events of interest. For each window, the CNN model is used to predict the probabilities of each class being present.

Furthermore, since the peak amplitude of signals belonging to different classes is large, to improve the learning ability of the models, we perform normalisation of the filtered recordings. In particular, in order to enable the model to focus on classifying the input signals and facilitate the subsequent explanation of the classification results, we normalise each 15-s window by subtracting mean and dividing by the maximum of the absolute value of each input window.

For the STFT map input, in order to get good time and frequency resolution, ‘Boxcar’ window with length of 128 samples with 70% overlap is used. We perform STFT on denoised and normalised time series input window. Thus, the input shape for the STFT-based model is 65 × 95 × 3 samples.

4.3. Training and testing

The inputs to the model for both training and testing comprise STFT maps generated from the raw recordings as discussed in the previous subsection. Our prior work in Jiang et al. (2023) demonstrate that CNN models tend to be overly sensitive. To address this, we refine the sensitivity of our CNN by only using the high-quality events to train the model. Specifically, we visually inspected and chose events from the original catalogue to ensure that the set used for training comprised only high-quality events based on signal clarity and high-SNR (Signal-to-Noise Ratio) for earthquake, quake and rockfall classes. All noise events originate from the original catalogue. In addition to the manually selected events, we utilise the labelled events from the 25th November 2014 (one day) to train the model further. These additional data allows us to augment the training set with events that are not included in the high-quality subset of the original catalogue and help to improve precision and recall.

The list of all the high-quality events from the original catalogue as well as the events from the 25th November 2014 used for training can be found as supplementary material for the purposes of reproducibility, as the second principle of Trustworthy AI. During testing phase, we use STFT maps from 26th to 28th Nov. 2014, which are not included in the training set. These labelled events are released with this paper as supplementary material.

4.4. Post-processing

While the sliding window technique enables continuous detection, it can introduce certain challenges. One of the main issues is that it may break the continuity of the event waveform, leading to potential inconsistencies or artefacts in the classification results. This occurs because the sliding window segments are treated independently, without considering the temporal context or smooth transitions between adjacent windows. To address this problem, post-processing techniques are often employed to refine and enhance the detection output by taking into account the temporal relationships between adjacent windows.

The proposed post-processing system is based on threshold filtering, median filtering, and Gaussian kernel filtering of the softmax output of the CNN. In addition, a peak selection method is applied to resolve cases where two classes of events have very similar detection results. (1) Threshold filtering: the softmax output of the CNN is filtered with a threshold value (set to 0.5), and all values below this threshold are set to zero. This is done to remove low-probability detections. (2) Median filtering: After the threshold filtering step, the probability distribution may contain isolated spikes. To remove these isolated spikes, we apply a median filter to each class separately. In addition to removing isolated spikes, the median filter can also merge spikes that are very close together, resulting in smoother and more continuous probability distributions. We set the size of the median filter to 5. (3) Gaussian kernel filtering: a Gaussian kernel filter is applied to the median filtered output to smooth the probability distribution. Gaussian kernel is defined with a sum of 1 and a length of 15. Its standard deviation is 5. (4) Peak selection: after using Gaussian kernel filtering, we select the highest peak (i.e., the longest duration) as the final output. This peak selection method allows us to choose the class of the event with the longest duration, as it indicates a higher confidence level in the classification result.

4.5. Explainability-informed re-labelling

Unlike classifiers such as RF, SVM and (norm)GLR-based classifiers that take hand-crafted features as inputs and where feature importance was studied in detail in Li et al. (2023a), the CNN multi-classifier is essentially a “black box” since we do not know what features were deemed important. We therefore utilise LRP to understand feature importance for the deep-learning CNN multi-classifier.

LRP (Bach et al., 2015) is a state-of-the-art XAI method, that shows the contribution of each sample in the input data to the classification results and can be implemented in the pre-trained model (Chan et al., 2023). In this paper, LRP is used to help identify which parts of the seismic signal are most important in making the final classification decision. This helps understanding which features of the seismic signal are most relevant for seismic detection, and identify any potential biases in the model. In addition, LRP can provide interpretable and detailed explanations of the model’s decision-making process, which can be useful for communicating the model’s results to human experts.

The LRP method starts from the output of the model, sets the output value before activation function as relevance, and gradually back propagates relevance, iteratively, layer by layer, to the input nodes. In the backpropagation, relevance follows the conservation law, that is, a neuron’s relevance equal to the sum of relevance as it flows out towards all other neurons. Various propagation rules have been proposed, such as LRP-γ, LRP-ϵ and LRP-0 rule (Montavon et al., 2022). In this paper, we used LRP-ϵ rule which is suitable for convolutional layers and max pooling layers (Montavon et al., 2017), and is defined as:

$$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_{0,j} a_j w_{jk}} R_k, \tag{1}$$

where R_j represents an LRP relevance score assigned to neuron j, a_j denotes an input activation, w_{jk} is the weight connecting neuron j to neuron k in the layer above, $\sum_{0,j}$ denotes that we sum over all neurons j in the lower layer plus a bias term w_{0k} with $a_0 = 1$. ϵ is a regularisation term, i.e., a small value that prevents the denominator from being 0.

We generate LRP maps for all events whose CNN-based predicted class does not correspond to the event class label as provided by the expert via the procedure described in Section 3.2 (i.e., misclassification). Then, we ask the same expert to review the recording, this time together with the LRP feature importance map, to ensure trust in the labels. The “corrected” labels (those that the expert agrees were originally wrongly labelled) are then marked and released as part of the supplementary material together with their STFT and LRP maps. The whole process is shown in Fig. 3.

5. Results

In this section, we first demonstrate our Contribution (5 & 6), by reporting the performance of the proposed models on the test dataset using standard classification performance measures as in Jiang et al. (2023). Then, we present our explainability results as per Contribution (7) and discuss main reasons behind misclassification (Contribution (8)).

5.1. Analysis of classifier output

Our models are implemented in Keras framework. Since the activation function of the output layer is softmax, we use categorical cross entropy as loss function. The used optimiser is Adam with an initial learning rate of 0.0007. Adaptive learning rate adjustment is implemented, which reduces the learning rate by a factor of 0.9 when loss improvements plateau for 5 epochs. Training is performed over 100 epochs with a batch size of 128. For the second training session, utilising the data from November 25, the model is trained over a total of 50 epochs. To prevent the risk of overfitting due to additional training,

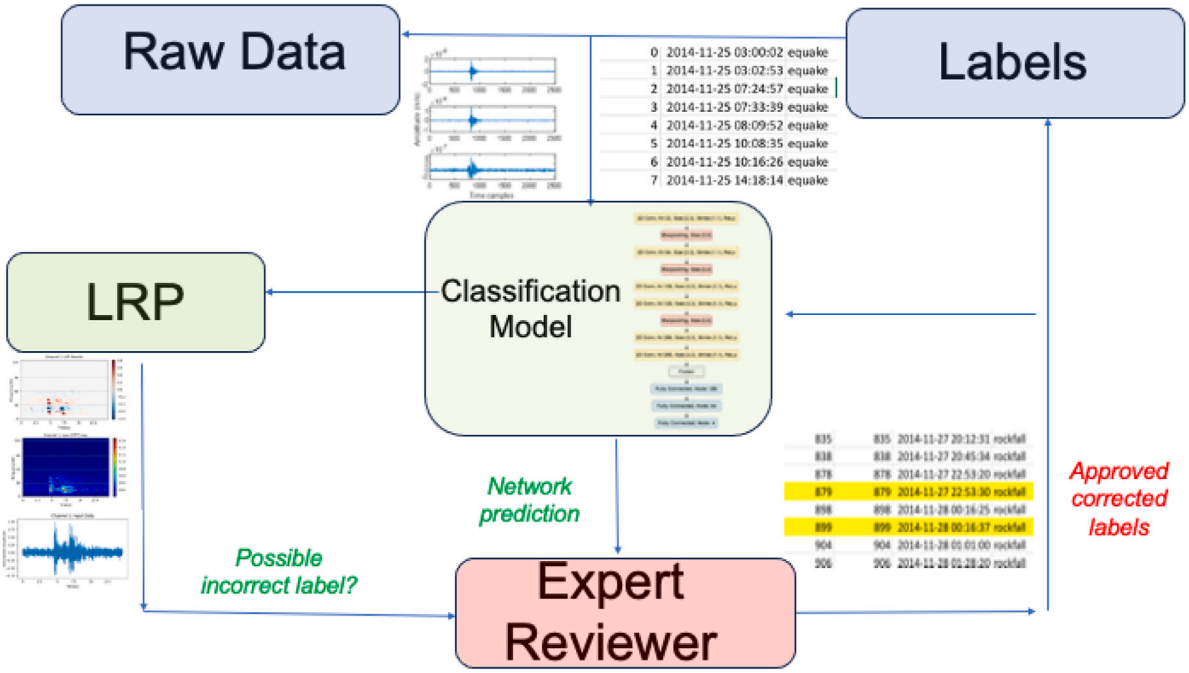

Raw Data
Labels
0 2014-11-25 03:00:02 equake
1 2014-11-25 03:02:53 equake
2 2014-11-25 07:24:57 equake
3 2014-11-25 07:33:39 equake
4 2014-11-25 08:09:52 equake
5 2014-11-25 10:08:35 equake
6 2014-11-25 10:16:26 equake
7 2014-11-25 14:18:14 equake
LRP
Classification Model
835 835 2014-11-27 20:12:31 rockfall
838 838 2014-11-27 20:45:34 rockfall
878 878 2014-11-27 22:53:20 rockfall
879 879 2014-11-27 22:53:30 rockfall
898 898 2014-11-28 00:16:25 rockfall
899 899 2014-11-28 00:16:37 rockfall
904 904 2014-11-28 01:01:00 rockfall
906 906 2014-11-28 01:28:20 rockfall
Approved corrected labels
Network prediction
Possible incorrect label?
Expert Reviewer

Fig. 3. Flowchart of the proposed human-on-the-loop process.

Table 1
Confusion Matrix - Proposed CNN-based network with post-processing against expert labels (the numbers in brackets indicate recall/sensitivity rates).

		Model			
		Quake	Earthquake	Rockfall	Noise
Expert	Quake	**26** (56.5%)	2	9	9
	Earthquake	0	**15** (83.3%)	1	2
	Rockfall	2	0	**72** (97.2%)	0
	Noise	110	13	58	**538** (75.1%)

early stopping is implemented; that is, if the training accuracy did not exhibit significant improvement within 5 consecutive epochs, the training process is terminated early.

In the 3-day testing period (26th–28th Nov.), the expert labelled 46 quakes, 18 earthquakes, 74 rockfalls and 719 noise events. The confusion matrix in Table 1 compares the output of the proposed CNN-based network, with post-processing (Section 4.4), to the expert labels. As is common practice for seismic signal classification on continuous data (Provost et al., 2017), the confusion matrix also includes recall/sensitivity values in brackets. Recall is the ratio of true positives to the sum of true positives and false negatives. In Section 3.2, it is demonstrated that during the 4-day period from November 25th to 28th, there are 6 additional earthquakes not labelled in the original catalogue (Provost et al., 2017). The model discussed in Jiang et al. (2023) detected a much larger number, specifically 174 additional, earthquakes. This comparison shows the significant improvement in the precision of earthquake classification achieved by our model. Additionally, our model achieved high recall (sensitivity) for rockfall events. As expected, quake and noise events can be confused with the other 3 classes, due to heterogeneity of the noise signal and very low signal amplitude of quake signals. Next, we leverage on LRP to explain the origin of misclassifications.

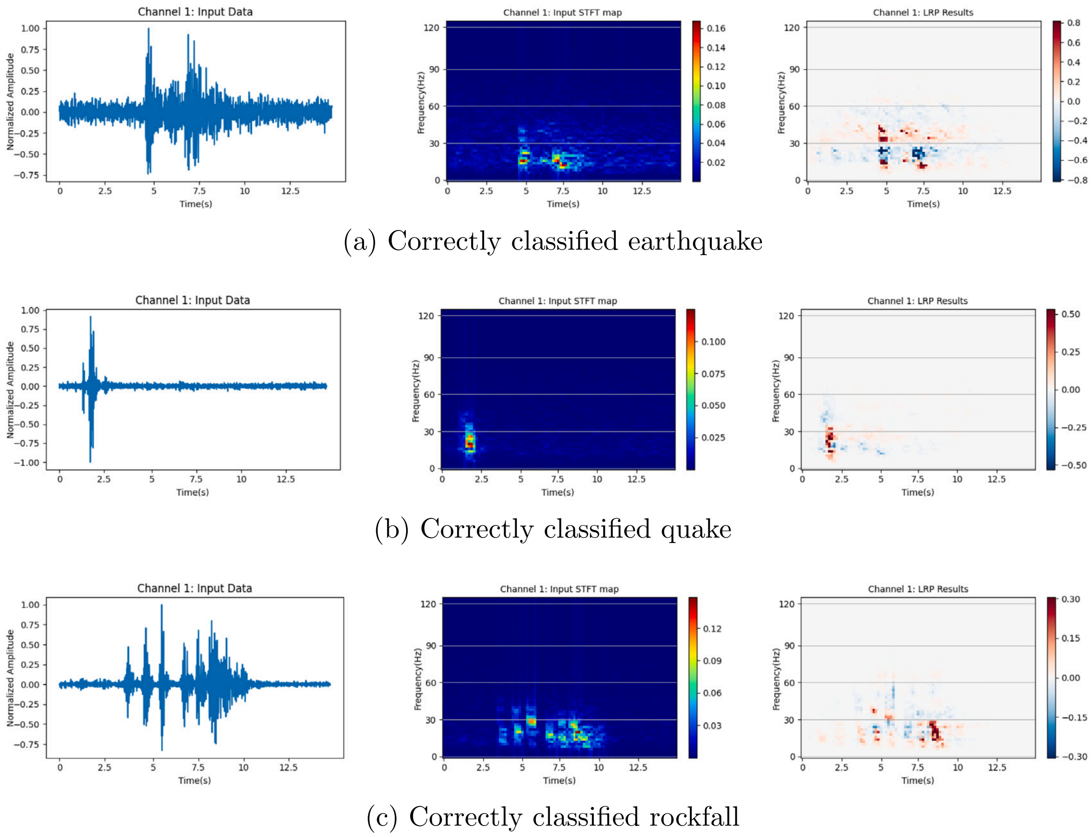
5.2. Explainability

The used package for embedding LRP into our models is iNNvestigate (Alber et al., 2019) which supports Keras framework in Python 3. Default parameters of the LRP-ϵ rule are used.

Fig. 4(a) shows an example of a correctly classified earthquake event. Positive and negative values of the LRP relevance represent positive and negative contributions to the classification results, of the corresponding STFT, respectively. The distribution of LRP relevance is focused on the high frequencies (about 40 to 50 Hz) when the P-wave is picked as well as the low frequencies (around 15 to 20 Hz) of the P-wave and, after roughly 5sec, the low frequencies of the S-wave with intermediate noise shown in light blue correctly identified as not contributing (negative contribution). This example shows that the model learnt, and uses as basis for its predictions, that the P-waves of earthquake events tend to have both high and low frequencies (around 50 Hz and 20 Hz, respectively) and that high energy content of S-Waves follows in time.

Fig. 4(b) shows an example of a correctly classified quake event. Quake events are of shorter duration than earthquakes, have lower amplitudes, and energy focused in low frequencies. LRP relevance is concentrated in the single peak (positive and negative) of the event waveform, suggesting that the normalised maximum amplitude is the key distinguishing feature. In the frequency domain, the LRP map clearly shows the importance of the peak that has energy mainly focused below 30 Hz while there is also a small positive contribution between 30 to 40 Hz.

Fig. 4(c) shows an example of a correctly classified rockfall event. While the relevance score of quake events is concentrated on a single peak, relevance of rockfall events is concentrated on multiple peaks, which also shows an important property of rockfall events – multiple significant peaks. Looking at the LRP map, relevance has multiple focused points corresponding to multiple short waves – a characteristic of rockfalls. In addition, although both rockfall and quake events have



(a) Correctly classified earthquake

(b) Correctly classified quake

(c) Correctly classified rockfall

Fig. 4. Correctly classified examples of earthquake, quake and rockfall: The first column shows the time-series signal, middle column the STFT, and the right column is the LRP relevance heatmap.

a frequency band between 10 to 30 Hz, LRP relevance is mostly concentrated at frequencies greater than 20 Hz for rockfalls and below 20 Hz for quakes.

Similar visualisation maps are produced for other correctly classified events. In summary, the model searches: (a) for P-wave and S-wave peaks and their corresponding frequency contributions to predict an earthquake; (b) a short wave with a single peak below 20 Hz to decide quake; (c) multiple significant frequency components around 25 Hz to decide that the target signal is rockfall. This is in accordance to the characteristics of the three signal classes (Provost et al., 2017; Li et al., 2020; Jiang et al., 2023). Next, we will analyse misclassified events to explain why they occur and how they can be avoided.

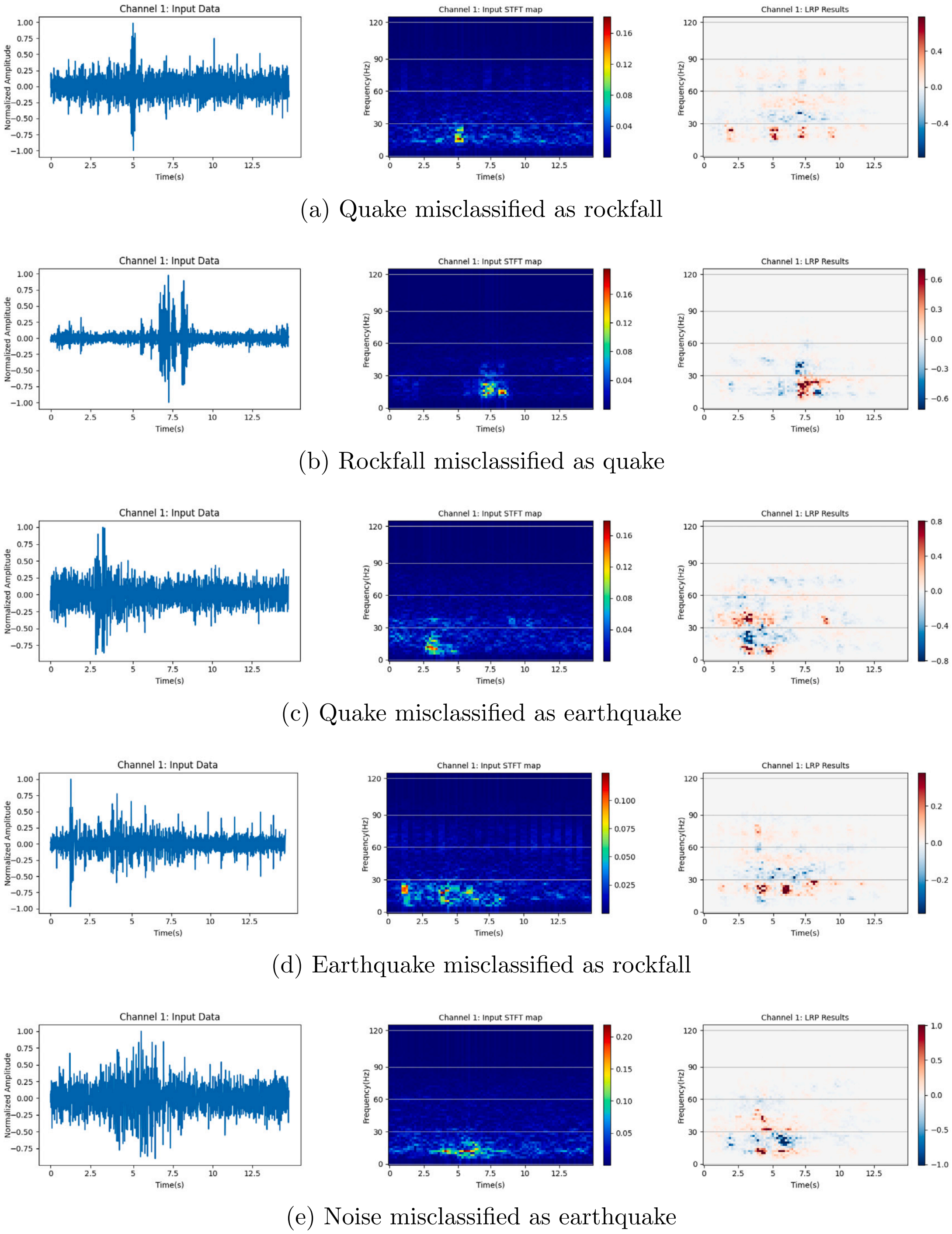
5.3. Explaining origin of misclassification

In this section, we show how LRP can be used for model diagnosis. The confusion matrix presented in Table 1 shows that the quake signals are sometimes misclassified as rockfalls. Interestingly, however, rockfall signals are rarely misclassified as quakes (only 2 misclassified events). To investigate this further, Fig. 5(a) shows an example of a quake event misclassified as rockfall. In the LRP map, relevance distribution is very scattered. That is, the LRP relevance is not focused on the quake event's peak, but instead picked up several consecutive peaks, where the positive relevance is correctly concentrated at 5 s. This indicates that the model correctly recognised a quake event's peak appearing around 5 s, but there was a high energy signal in nearby frequency bands, influencing the final prediction. On the other hand, there are many positive relevancies at different times that correspond to frequencies between 20 Hz to 30 Hz, which is akin to the learnt rockfall 'behaviour'. Thus, the main reason of misclassification between quake and rockfall is that the signal-to-noise ratio of the quake event was very low, with a noise signal appearing immediately after, mimicking multiple peaks of rockfall events.

In Fig. 5(b), we show an instance in which a rockfall event is misclassified as a quake. The rockfall event displays multiple peaks; however, these peaks, aside from the principal one, are of low magnitude and the event has a very short time span. Analysis of the LRP representation illustrates a concentration of positive effects (depicted in red) at the primary peak of the event. Conversely, numerous negative contributions (depicted in blue) are observed at the secondary peaks, suggesting that the presence of these multiple peaks is not taken into account due to their limited magnitudes; hence, the model finally classifies this event as a quake.

In Fig. 5(c), we present an instance of a quake misclassified as an earthquake. This misclassification is evident in the LRP map, where both high-frequency and low-frequency components simultaneously exhibit positive contributions around the 3-s period. Thus, the model interprets this segment as a P-wave. Furthermore, at approximately 5 s into the waveform, a positive contribution appears in the low-frequency range. Although the primary peak of this event occurs around 3 s, the spectrogram reveals that the low-frequency component persists for an extended duration. Moreover, the event is influenced by higher-frequency noise (exceeding 30 Hz), and this high-frequency noise coincides with the primary waveform peak around the 3 s. Consequently, this led the model to mistakenly identify it as a P-wave, with the prolonged low-frequency component being mistakenly identify as a S-wave. These observations align with seismic features of earthquakes, thereby causing the model's misclassification as an earthquake event.

In Fig. 5(d), we encounter an instance where an earthquake is mistakenly classified as a rockfall. The LRP map highlights multiple spectral peaks, which is a feature of rockfall events. However, this event may have resulted from an earthquake occurring amidst background noise, exhibiting a distinctive multi-peak pattern. Thus, despite the



(a) Quake misclassified as rockfall

(b) Rockfall misclassified as quake

(c) Quake misclassified as earthquake

(d) Earthquake misclassified as rockfall

(e) Noise misclassified as earthquake

Fig. 5. Misclassified examples.

presence of a P-wave at approximately 1 s and an S-wave at roughly 4 s, complex background noise caused misclassification.

In Fig. 5(e), the misclassification of noise as an earthquake is shown. The noise signal exhibits prominent peaks around 4 s and 5.5 s. Examination of the LRP map reveals the model's recognition of low-frequency and high-frequency components (15–20 Hz) around the 4-s mark, along with low-frequency signals at 5.5 s (15 Hz). This aligns with the characteristic features of P-waves and S-waves in earthquake signals, resulting in the model's misclassification as an earthquake. The result might have been different if time-series signals were inputted to the network instead of the STFT maps as can be seen from the left time-series plot that shows high level of noise throughout the signal.

We can see from these examples that most misclassifications are due to high level of background noise. The next example highlights another origin of error related to the filtering process. Fig. 6 displays an unfiltered earthquake waveform with a frequency below 3 Hz, characteristic of low-frequency earthquakes that are rarely associated with active landslides (Masuda et al., 2020). Since our focus is on detecting local seismic events related to landslides, we apply a bandpass filter in the 5–60 Hz range (see Section 4.2), which excludes these low-frequency earthquakes. Consequently, this filter removed the low-frequency event's waveform, leaving only background noise as input to the CNN. As illustrated in Fig. 7, the LRP map indicates that the model failed to extract meaningful features from the filtered input, resulting in

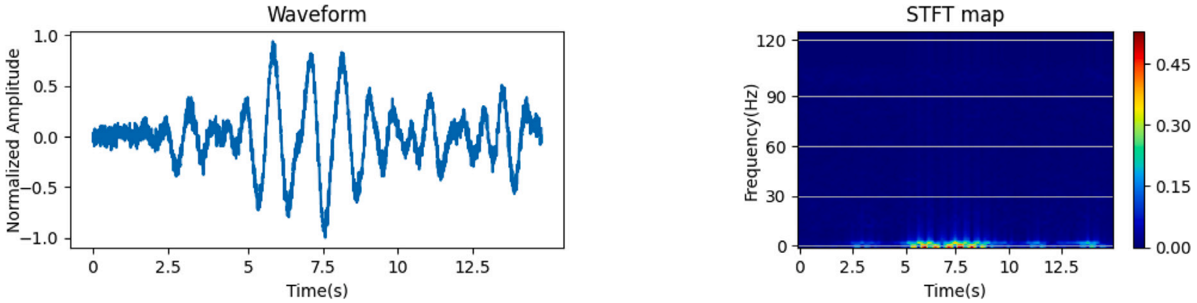


Fig. 6. Waveform (left) and STFT map (right) of the unfiltered low-frequency earthquake.

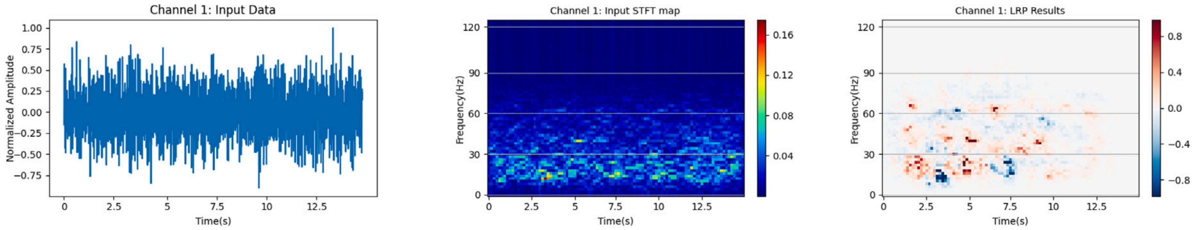


Fig. 7. Waveform (left), STFT map (middle) and the LRP map (right) of the filtered low-frequency earthquake.

the earthquake being misclassified as noise. This misclassification can be attributed to the rarity and uniqueness of low-frequency earthquakes on landslides, as our filter inadvertently eliminated their distinctive waveforms, confounding the CNN's classification process.

5.4. Re-labelling results

Fig. 8 shows three examples of misclassifications, which could be due to human error during expert labelling. The example shown in Fig. 8(a), is an event classified by the model as noise, though the domain experts labelled it as a quake. In the STFT representation of the signal, no obvious peak corresponding to the event was discernible. Moreover, the LRP map exhibits a disordered distribution of relevance. Collectively, these findings lead to the argument that the event in question is more likely to be anthropogenic noise rather than a quake. Fig. 8(b) illustrates a similar situation where the event is mistakenly labelled as an earthquake. There are no clear P-waves at both low and high frequencies, and there are no S-waves with high energy content. For this earthquake event, we also examined the unfiltered raw signal, and it still did not exhibit any earthquake waveform characteristics. Fig. 8(c) shows an example that was classified as a rockfall by the CNN model, while the expert labelled it as a seismic quake. It can be concluded from the LRP map that the model focused on multiple peaks in the event, with a frequency distribution centred around 30 Hz, characteristics that align with typical rockfall patterns. In contrast, quakes tend to exhibit a single dominant peak, a feature that was notably absent in the input STFT map, where multiple peaks were discernible. Consequently, based on these distinctive patterns and spectral features, it becomes evident that the event in question is more accurately classified as a rockfall.

Here we list all corrections made to the expert catalogue, following above explainability and queries. Specifically, 7 quakes were relabelled as noise as per example Fig. 8(a), 1 earthquake was relabelled as noise (shown in Fig. 8(b)), and 1 quake as rockfall (Fig. 8(c)). In addition, some noise events were labelled by the expert though these events occurred very close to earthquake, quake and rockfall events, which potentially caused confusion. Hence, we removed all noise events that occurred in close proximity (within 30s) to the earthquake, quake and rockfall events — this way 38 noise events were removed.

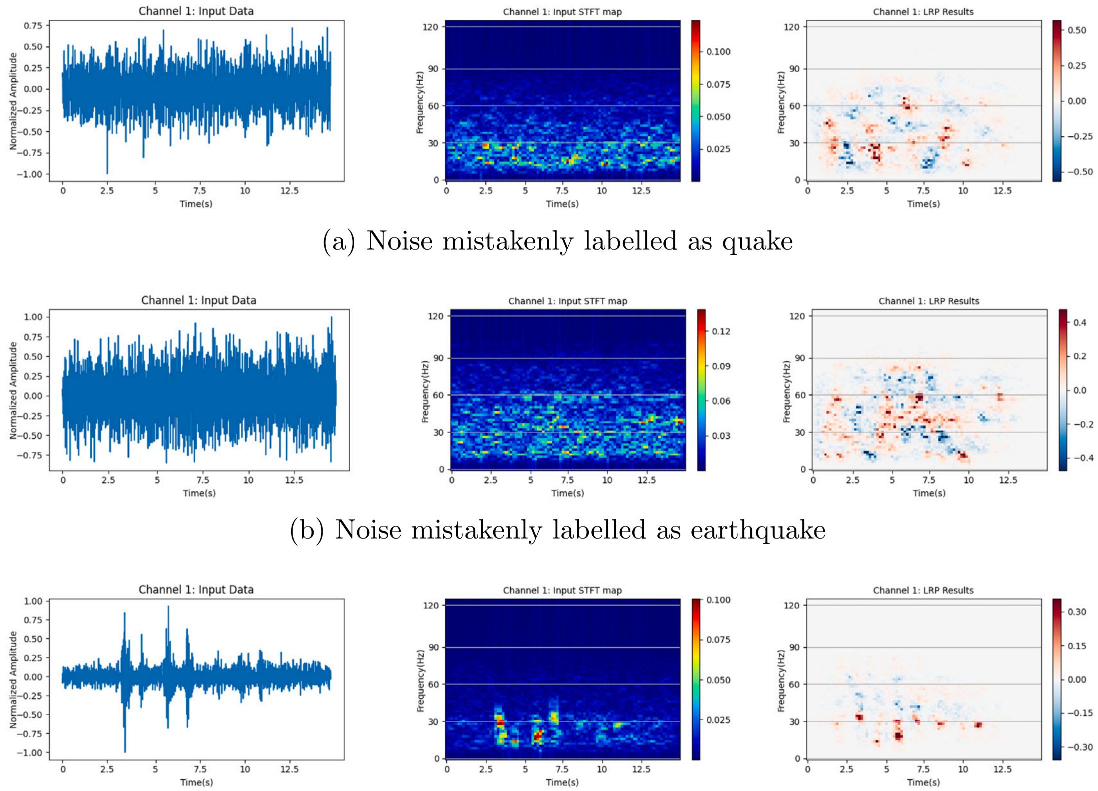
Thus, after this relabelling there are 38 quakes, 17 earthquakes, 75 rockfalls and 689 anthropogenic noise events in total. The verified catalogue of events is provided as supplementary material to this paper, as a contribution to address the second and third principles of Trustworthy AI, related to reproducibility and data access. Specifically, the 260 verified events on the 25th Nov. 2015 are listed in the Training events supplementary material, identified by the date. The 819 verified events on 26th to 28th Nov. 2014 are listed in the Additional 3-day catalogue supplementary material. In order for other researchers to enable benchmarking, Table 2 and Table 3 show the confusion matrix and classification performance after the re-labelling, respectively. Although the F1-score for quake events is low, we have a high recall but precision is low because of 8 instances of false positives for rockfall. There are relatively few instances of quake and earthquake, which explains why the F1-score is not the best indicator of performance and the confusion matrix provides a more explainable and trustworthy measure of performance.

6. Conclusions and future work

The paper discusses the significance of the 7 principles of Trustworthy AI, including human oversight, technical robustness, data governance and transparency to the challenging problem of micro-seismic signal analysis. To this effect, we propose a human-on-the-loop microseismic multi-class classification method together with LRP to shed light on feature importance in order to in turn verify any possible human labelling error.

We demonstrate that the generated LRP maps assist human experts in manual event classification. LRP clearly identifies properties of the signals extracted by the network when making decisions. Based on this, we concluded, for example, that the main reason why quake events are often misclassified as rockfall is due to appearance of a noise signal at multiple higher frequencies that mimics rockfalls. Due to human error, experts may occasionally mislabel events in the catalogue due to the similarity of event characteristics, complexity of seismic data and large volume of data that needs to be processed. However, the availability of LRP maps as a visual aid can offer a valuable tool to verify and refine the expert's classifications. This collaborative synergy between automated and manual classification can enhance the accuracy of microseismic catalogues, contributing to a better understanding of geological processes.

Besides assisting with event labelling, another application of the LRP maps is improving the model's performance. Indeed, by observing



(a) Noise mistakenly labelled as quake

(b) Noise mistakenly labelled as earthquake

(c) Rockfall mistakenly labelled as quake

Fig. 8. Three examples of events with labels corrected.

Table 2
The confusion matrix after label correction. The numbers in the brackets show the recall values.

		Model			
		Quake	Earthquake	Rockfall	Noise
Expert	Quake	**26** (68.4%)	2	8	2
	Earthquake	0	**15** (88.2%)	1	1
	Rockfall	2	0	**73** (97.3%)	0
	Noise	95	11	37	**546** (79.2%)

Table 3
The classification performance after label correction.

	Precision	Recall	F1-score
Quake	0.21	0.68	0.32
Earthquake	0.54	0.88	0.67
Rockfall	0.61	0.97	0.75
Noise	0.99	0.79	0.88

the insights gained through XAI tools, we discern specific features of input events that are prone to misclassification by the CNN, which is instrumental in enhancing the robustness and generalisability of the model that can be achieved by adding more events in the training set that closely resemble the challenging input patterns identified through XAI. For example, when we discover that certain event features consistently lead to misclassifications, we collect and add more events with similar attributes into the training dataset. This targeted data augmentation approach has the potential to improve the model's ability to distinguish between challenging seismic events, thereby increasing model's robustness and classification performance.

Since LRP assigns relevance scores to highlight the most influential features for each classification, it is important to determine if these relevance patterns remain stable across various geographic areas and seismometer characteristics, such as sensitivity, sampling rate, and axis configurations. This evaluation will help ascertain the reliability of LRP explanations across diverse equipment types and environments. In future work, we plan to test our system in various geographic regions and with different seismometer configurations to assess the consistency and robustness of LRP interpretability, enhancing the broader applicability and trustworthiness of our approach.

Given the potential variability in expert interpretations, it is important to explore how different experts' insights may affect labelling. Future studies could employ a multi-expert assessment framework that incorporates confidence levels, based on the methodologies proposed by Sobot et al. (2024), to better understand this variability and further enhance the reliability of the classification process.

Since classification of quakes remains challenging, the current model could be adapted to classify a broader range of events, including low frequency events and types of anthropogenic noise, by expanding the training set and retraining the model, with LRP providing the explanations. To maximise accuracy and trust in AI-driven seismic signal analysis, integrating human expertise with AI models is important. Developing interactive explainability tools that facilitate iterative feedback from geoscientists could lead to continuous improvements in model performance and foster greater confidence in AI-generated outputs.

CRediT authorship contribution statement

Jiaxin Jiang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **David Murray:** Writing – review & editing, Validation, Formal analysis, Data curation. **Vladimir Stankovic:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Lina Stankovic:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Clement Hibert:** Validation, Investigation, Data curation. **Stella Pytharouli:** Project administration, Funding acquisition. **Jean-Philippe Malet:** Resources, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the EPSRC Prosperity Partnership research and innovation programme, UK under grant agreement EP/S005560/1, and in part by the EPSRC New Horizons research programme, UK EP/X01777X/1. The contextual data interpretation and labelling work by experts on the SZ dataset was supported by RSE Saltire International Collaboration Awards.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.srs.2024.100189.

Data availability

The code is provided in a GitHub repository at https://github.com/kanata2020/Explainable-seismic-classification. This includes the models and data used for classification, as well as algorithms for explainable visualisation.

References

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K.T., Montavon, G., Samek, W., Müller, K.-R., Dähne, S., Kindermans, P.-J., 2019. iNNvestigate neural networks! J. Mach. Learn. Res. 20 (93), 1–8.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10 (7), 1–46.

Bau, D., Zhu, J.-Y., Strobelt, H., Lapedriza, A., Zhou, B., Torralba, A., 2020. Understanding the role of individual units in a deep neural network. Proc. Natl. Acad. Sci. 117 (48), 30071–30078.

Bayramoğlu, Z., Uzar, M., 2023. Performance analysis of rule-based classification and deep learning method for automatic road extraction. Int. J. Eng. Geosci. 8 (1), 83–97.

Bi, X., Zhang, C., He, Y., Zhao, X., Sun, Y., Ma, Y., 2021. Explainable time–frequency convolutional neural network for microseismic waveform classification. Inform. Sci. 546, 883–896.

Chan, A., Schneider, M., Körner, M., 2023. XAI for early crop classification. In: IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2657–2660.

Dos, M.E., 2022. Determination of city change in satellite images with deep learning structures. Adv. Remote. Sens. 2 (1), 16–22.

European Commission and Directorate-General for Communications Networks, Content and Technology, 2019. Ethics Guidelines for Trustworthy AI. Publications Office.

FrenchLandslideObservatorySeismologicalDatacenter/RESIF, 2021. Observatoire Multidisciplinaire des Instabilites de Versants (OMIV). http://dx.doi.org/10.15778/RESIF.MT, [online]. Available: https://seismology.resif.fr/. (Accessed: 2021).

Gülgün, O.D., Erol, H., 2020. Classification performance comparisons of deep learning models in pneumonia diagnosis using chest X-ray images. Turkish J. Eng. 4 (3), 129–141.

Hägele, M., Seegerer, P., Lapuschkin, S., Bockmayr, M., Samek, W., Klauschen, F., Müller, K.-R., Binder, A., 2020. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. Sci. Rep. 10 (1), 1–12.

Holzinger, A., 2018. From machine learning to explainable AI. In: 2018 World Symposium on Digital Intelligence for Systems and Machines. DISA, IEEE, pp. 55–66.

Jiang, J., Stankovic, V., Stankovic, L., Murray, D., Pytharouli, S., 2024. Explainable AI for transparent seismic signal classification. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 8801–8805.

Jiang, J., Stankovic, V., Stankovic, L., Parastatidis, E., Pytharouli, S., 2023. Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 61, 1–14.

Kaya, Y., Şenol, H.İ., Yiğit, A.Y., Yakar, M., 2023. Car detection from very high-resolution UAV images using deep learning algorithms. Photogramm. Eng. Remote Sens. 89 (2), 117–123.

Lapuschkin, S., Binder, A., Montavon, G., Müller, K.-R., Samek, W., 2016. Analyzing classifiers: Fisher vectors and deep neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, pp. 2912–2920.

Li, J., Stankovic, L., Pytharouli, S., Stankovic, V., 2020. Automated platform for microseismic signal analysis: Denoising, detection, and classification in slope stability studies. IEEE Trans. Geosci. Remote Sens. 59 (9), 7996–8006.

Li, J., Stankovic, L., Stankovic, V., Pytharouli, S., Yang, C., Shi, Q., 2023a. Graph-based feature weight optimisation and classification of continuous seismic sensor array recordings. Sensors 23 (1).

Li, J., Ye, M., Stankovic, L., Stankovic, V., Pytharouli, S., 2023b. Domain knowledge informed multitask learning for landslide-induced seismic classification. IEEE Geosci. Remote. Sens. Lett. 20, 1–5.

Liao, W.-Y., Lee, E.-J., Wang, C.-C., Chen, P., Provost, F., Hiber, C., Malet, J.-P., Chu, C.-R., Lin, G.-W., 2022. RockNet: Rockfall and earthquake detection and association via multitask learning and transfe learning. Authorea.

Masuda, K., Ide, S., Ohta, K., Matsuzawa, T., 2020. Bridging the gap between low-frequency and very-low-frequency earthquakes. Earth Planets Space 72 (1), 1–9.

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.-R., 2022. Layer-wise relevance propagation: An overview. In: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer-Verlag, Berlin, Heidelberg, pp. 193–209.

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., Müller, K.-R., 2017. Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognit. 65, 211–222.

Mousavi, S.M., Beroza, G.C., 2022. Deep-learning seismology. Science 377 (6607), eabm4470.

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y., Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nature Commun. 11 (1), 1–12.

Murray, D., Stankovic, L., Pytharouli, S., Stankovic, V., 2023. Semi-supervised seismic event detection using siamese networks. EGU General Assembly 2023, April 2023.

Murray, D., Stankovic, L., Stankovic, V., 2021. Transparent AI: Explainability of deep learning based load disaggregation. In: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. BuildSys '21, Association for Computing Machinery, New York, NY, USA, pp. 268–271.

Parasyris, A., Stankovic, L., Stankovic, V., 2023. Synthetic data generation for deep learning-based inversion for velocity model building. Remote Sens. 15 (11), 2901.

Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4 (2), e1700578.

Provost, F., Hibert, C., Malet, J.-P., 2017. Automatic classification of endogenous landslide seismicity using the random forest supervised classifier. Geophys. Res. Lett. 44 (1), 113–120.

Provost, F., Malet, J.-P., Gance, J., Helmstetter, A., Doubre, C., 2018a. Automatic approach for increasing the location accuracy of slow-moving landslide endogenous seismicity: the APOLoc method. Geophys. J. Int. 215 (2), 1455–1473.

Provost, F., Malet, J.-P., Hibert, C., Helmstetter, A., Radiguet, M., Amitrano, D., Langet, N., Larose, E., Abancó, C., Hürlimann, M., Lebourg, T., Levy, C., Le Roy, G., Ulrich, P., Vidal, M., Vial, B., 2018b. Towards a standard typology of endogenous landslide seismic sources. Earth Surf. Dyn. 6 (4), 1059–1088.

Renouard, A., Maggi, A., Grunberg, M., Doubre, C., Hibert, C., 2021. Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning. Seism. Soc. Am. 92 (6), 3725–3742.

Saad, O.M., Chen, Y., 2020. Earthquake detection and P-wave arrival time picking using capsule neural network. IEEE Trans. Geosci. Remote Sens. 59 (7), 6234–6243.

Saad, O.M., Chen, Y., 2021. CapsPhase: Capsule neural network for seismic phase classification and picking. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Samek, W., Binder, A., Montavon, G., Lapuschkin, S., Müller, K.-R., 2016. Evaluating the visualization of what a deep neural network has learned. IEEE Trans. Neural Netw. Learn. Syst. 28 (11), 2660–2673.

Sariturk, B., Bayram, B., Duran, Z., Seker, D.Z., 2020. Feature extraction from satellite images using segnet and Fully Convolutional Networks (FCN). Int. J. Eng. Geosci. 5 (3), 138–143.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2019. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. 128 (2), 336–359.

Shakeel, M., Itoyama, K., Nishida, K., Nakadai, K., 2021. EMC: Earthquake magnitudes classification on seismic signals via convolutional recurrent networks. In: 2021 IEEE/SICE International Symposium on System Integration. SII, IEEE, pp. 388–393.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Sobot, T., Stankovic, V., Stankovic, L., 2024. Human in the loop active learning for time-series electrical measurement data. Eng. Appl. Artif. Intell. 133, 108589.

Soneson, C., Gerster, S., Delorenzi, M., 2014. Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. PLoS One 9 (6), e100335.

Trani, L., Pagani, G.A., Zanetti, J.P.P., Chapeland, C., Evers, L., 2022. DeepQuake—An application of CNN for seismo-acoustic event classification in The Netherlands. Comput. Geosci. 159, 104980.

United Nations Sustainable Development Goal 13, 2023. [online]. Available: https://sdgs.un.org/goals. (Accessed: 2023).

Vouillamoz, N., Rothmund, S., Joswig, M., 2018. Characterizing the complexity of microseismic signals at slow-moving clay-rich debris slides: the Super-Sauze (south-eastern France) and Pechgraben (Upper Austria) case studies. Earth Surf. Dyn. 6 (2), 525–550.

Yi, D., Yiran, S., Ismet, C., Xun, L., Guangyao, S., 2021. Classification of clustered microseismic events in a coal mine using machine learning. J. Rock Mech. Geotech. Eng. 13 (6), 1256–1273.