RockNet: Rockfall and Earthquake Detection and Association via Multitask Learning and Transfer Learning

Wu-Yu Liao, En-Jui Lee, Chung-Ching Wang, Po Chen, Floriane Provost, Clément Hibert, Jean-Philippe Malet, Chung-Ray Chu, and Guan-Wei Lin

***Abstract*— Seismological data plays a crucial role in timely slope failure hazard assessments. However, identifying rockfall waveforms from seismic data poses challenges due to their high variability across different events and stations. To address this, we propose RockNet, a deep-learning-based multitask model capable of detecting both rockfall and earthquake events at both the single-station and local seismic network levels. RockNet consists of two submodels: the single-station model, which computes waveform masks for earthquake and rockfall signals and performs earthquake *P* and *S* phase picking simultaneously on single-station seismograms, and the association model, which determines the occurrences of local seismic events by aggregating hidden feature maps from the trained single-station model across all stations. Since the rockfall data is relatively scarce and may not be sufficient to train a deep-learning model effectively, we augment the dataset with abundant nonrockfall data and add additional tasks to promote shared interpretability and robustness. RockNet is trained and tested on a local dataset collected from the Luhu tribe in Miaoli, Taiwan, achieving macro F1-scores of 0.983 and 0.990 for the single-station model and the association model, respectively. Furthermore, we evaluate RockNet on an independent dataset collected from the Super-Sauze unstable slope region in France, and it demonstrates good generalization performance in discriminating earthquake, rockfall, and noise with a macro F1-score of 0.927. This study highlights the potential of deep learning in leveraging diverse types of inputs for seismic signal detection even with limited training data.**

***Index Terms*— Multitask learning, rockfall seismic monitoring, transfer learning.**

Manuscript received 9 January 2023; revised 2 May 2023; accepted 1 June 2023. Date of publication 8 June 2023; date of current version 23 June 2023. This work was supported in part by the National Science and Technology Council (NSTC), R.O.C., under Grant 110-2116-M-006-011, Grant 111-2116-M-006-017, and Grant 109-2116-M-006-018; and in part by the Soil and Water Conservation Bureau, Taiwan, under Grant SWCB-108-294 and Grant SWCB-109-227. The work of Wu-Yu Liao was supported by a fellowship of NSTC. The work of Po Chen was supported by the Nielson Energy Fellowship provided by the School of Energy Resources, University of Wyoming. The work of Floriane Provost was supported by a fellowship of the French Agence Nationale de la Recherche (HighLand Project) under Grant ANR-20-CE01-0006. *(Corresponding author: En-Jui Lee.)*

Wu-Yu Liao, En-Jui Lee, Chung-Ching Wang, and Guan-Wei Lin are with the Department of Earth Sciences, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: rickli92@gmail.com).

Po Chen is with the Department of Geology and Geophysics, University of Wyoming, Laramie, WY 82071 USA.

Floriane Provost is with the Institut Terre et Environnement de Strasbourg (ITES), CNRS UMR 7063—Université de Strasbourg, F-67084 Strasbourg, France.

Clément Hibert and Jean-Philippe Malet are with the Institut Terre et Environnement de Strasbourg, CNRS UMR 7063—Université de Strasbourg, 67084 Strasbourg, France, and also with the Ecole et Observatoire des Sciences de la Terre (EOST-OMIV), CNRS UAR830, Université de Strasbourg, F-67000 Strasbourg, France.

Chung-Ray Chu is with the National Science and Technology Center for Disaster Reduction, New Taipei City 23143, Taiwan.

Digital Object Identifier 10.1109/TGRS.2023.3284008

I. Introduction

ROCKFALL is one of the socioeconomic exposure risks in mountainous areas that correlates with other slope failures (e.g., landslides and debris flow). Rockfall studies usually consider meteorological factors [1], [2], [3], [4] and geomorphological evolution via instruments like airborne LiDAR and terrestrial laser scanning (TLS) that generates high-resolution topographic data [5], [6], [7], [8], [9]. Time-lapse imaging of rockfalls obtained from stereographic pairs of sequential photographs and cameras is also intuitive for rockfall identification [4], [10]. Although these methods have facilitated rockfall studies in various aspects, none are almighty and exist deficiencies. LiDAR and TLS are expensive when applied to vast areas and are ineffective during heavy rains that cause much light refraction. Photograph monitoring is susceptible to environmental visibility, such as foggy and rainy weather, and poor performance at night. Seismogram recorded by seismometers is emerging as regular monitoring data for providing timely dynamic process of mass movements. Seismic monitoring is relatively cost-effective and free from weather impacts compared to other methods. Furthermore, the seismological analysis conducted in the time and time–frequency domains could provide physical parameters like source–receiver distance, magnitude, and source location [11], [12], [13], in which the seismic energy attenuates with distances and is correlated with rockfall volume and mechanisms.

The challenge of rockfall seismic monitoring is identifying rockfall-induced signals on continuous seismic recordings. A general technique to achieve automatic rockfall waveform detection is to apply an amplitude-sensitive algorithm in the time or frequency domain to find potential events and confirm these events based on a self-designed criterion [2]. In recent decades, machine-learning algorithms have gained widespread adoption in this domain. Classical approaches, such as hidden Markov model [14], fuzzy logic [15], random forest [16], [17],

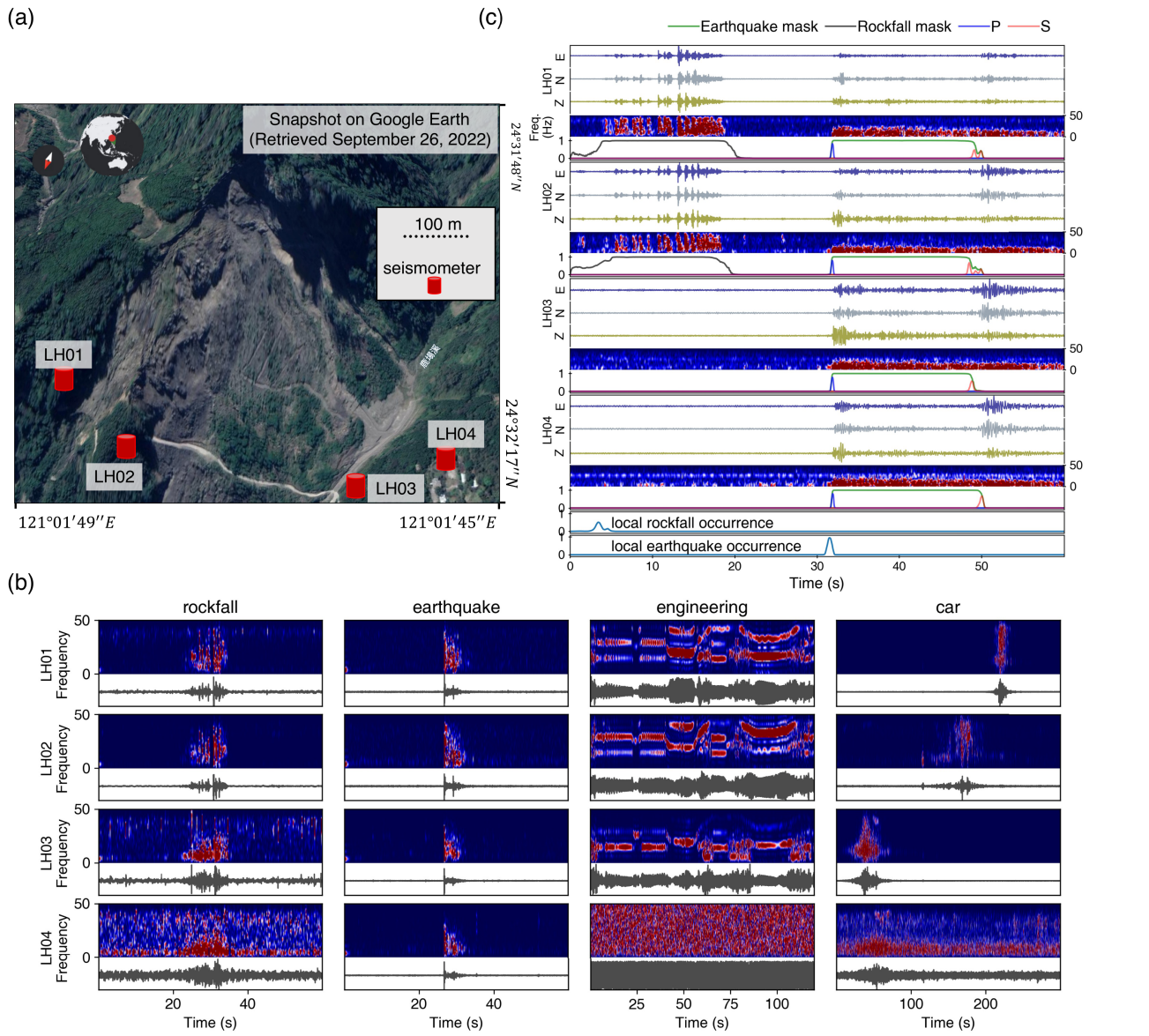


Fig. 1. (a) Aerial view of photographs displaying the distribution of stations in the Luhu unstable slope region in Taiwan. (b) Seismic waveforms and spectrograms of different source types recorded by the local network, including rockfalls, earthquakes, engineering signals, and car-induced vibrations. (c) Example of detected rockfall and earthquake events and their corresponding model output functions. The RockNet model takes the three-component recordings and spectrograms of the vertical component recording from stations as inputs. The output of the single-station detection model includes probability functions for earthquake P and S phases (blue and red lines, respectively), earthquake waveform functions (green line), and rockfall waveform functions (black lines). The probability functions for earthquake P and S phases provide information on the arrival times of P and S waves, and earthquake mask pairs the potential P and S phases of the same earthquake. The association model aggregates the encoded features of all station recordings from the single-station detection model and generates the local occurrence time functions of rockfall and earthquake events (bottom two panels).

[18], [19], [20], and support vector machines [19], [21], [22], rely on carefully selected features for effective classification. However, rockfall waveforms differ according to the source mechanisms coupled with their propagation media, numbers of blocks, volumes, and movements, causing generalization problems of the detection algorithm [11], [23], [24], [25].

In this study, we propose a deep-learning model that leverages both waveforms and spectrograms from all stations in the seismic network to automatically extract relevant features and interpret the input data at both the single-station and local seismic network levels. Deep learning, a subfield of machine learning, boasts several advantages over classical approaches [26]. These include superior feature extraction capabilities, scalability for handling large datasets, enhanced generalizability to unseen data, and the ability to learn end-to-end representations without handcrafted feature extraction or processing steps. However, deep-learning models typically require a substantial amount of training data when trained from scratch. Given the scarcity of available rockfall waveforms, which may not be sufficient for training a deep-learning model alone, we augment the model training data with additional earthquake and noneartquake data [27], [28] for model development. Furthermore, at the single-station level, our model includes tasks, such as earthquake detection

and body-wave arrival-time picking in addition to rockfall waveform detection, while at the local seismic network level, it performs two tasks: determining local occurrences of rockfall and earthquake events. By training multitask models on abundant additional nonrockfall data, our model can leverage shared interpretability across different task computations, mitigating the challenge of limited data availability for rockfall detection tasks.

Fig. 1 provides an overview of our study, where we collected and labeled four types of seismic events recorded by a local seismic network with station spacing of less than 1 km, deployed near a slope failure site. The geometry of the seismic network imposes physical constraints on seismic event confirmation due to the physics of the observed signals. For instance, earthquake waveforms would exhibit coherence in appearance and arrival time across the stations due to short spacing. In contrast, rockfall waveforms would have little time delay among all stations, and the ground motion recordings would become increasingly unclear with increasing wave propagation distance [Fig. 1(b)]. Hence, confirming rockfall and earthquake events in the local seismic network requires associating the interpretation results of all single-station recordings. Following this logic, we develop RockNet, consisting of two multitask models: a single-station detection model for interpreting the single-station recordings and an association model that performs transfer learning by aggregating the responses (i.e., feature maps) of the single-station model on each station data [Fig. 2].

II. Data

From February 26, 2019, to December 31, 2020, we deployed a local seismic network consisting of four three-component geophones produced by DiGOS, sampled at 100 Hz, in the Luhu tribe, Miaoli, Taiwan. This region has experienced frequent rockfall events since a slope failure event in April 2018, resulting in the closure of the main traffic artery due to unstable slopes [Fig. 1(a)]. We assumed that the main types of seismic events in this area were earthquakes, rockfalls, slope stabilization engineering events, and car-induced vibrations [Fig. 1(b)]. In this study, we incorporated additional datasets, including the Taiwan earthquake dataset and the STEAD dataset [27] for the single-station detection model, the INSTANCE dataset [28] for the association model, and the artificial rockfall experiment waveforms [25] for both models. To mitigate imbalanced data issues during model training, we applied data augmentation and oversampling strategies.

To compile the database from nearly two years of continuous recordings, we utilized short-term average/long-term average (STA/LTA) algorithms [29] to detect pulse-like signals and then manually labeled the events based on our field experience with encountered rockfalls. Unlike earthquake signals that exhibit highly similar waveforms and coherent time–frequency patterns across all stations, rockfall signals at all stations display high variability in the time and frequency domains. Near-field stations show high-frequency energy (greater than 10 Hz) dominating rockfall waveforms, while far-field stations exhibit relatively low-frequency energy due to significant attenuation of high-frequency portions with increasing distances. We identified engineering signals based on our field experience and determined car-induced signals by inspecting the order and speed of signal appearances across the stations [Fig. 1(b)]. For in situ earthquake waveforms, we detected and manually checked the P and S arrivals using the RED-PAN model [30]. In total, we identified 348 rockfall events recorded by more than two stations, comprising 750 sets of three-component waveforms; 1399 sets of three-component rockfall waveforms labeled with only one station; 193 car-induced events with 495 sets of three-component waveforms; 280 sets of engineering signals with 455 three-component waveforms; and 1834 earthquake events with 5324 three-component waveforms. The Luhu dataset was divided into training and testing sets in a 70%–30% ratio, with all potential rockfall event waveforms labeled with only one station excluded from the test set.

III. Methods

In this study, we propose RockNet that incorporates both multitask learning and transfer learning techniques. Multitask learning facilitates knowledge transfer among different tasks and allows for learning with limited training data, thereby improving computational efficiency compared to using separate models for each task. On the other hand, transfer learning focuses on leveraging knowledge learned from pretrained tasks to address related but distinct problems. In our approach, we transfer the representations learned by the single-station detection model to the association model, enabling the buildup of knowledge from the pretrained model to enhance the performance of the association model. By combining these two techniques, RockNet can effectively leverage both shared and transferred knowledge in a local seismic network.

A. Input of the Neural Network

We construct two sets of input vectors to the neural network for each sample: three-component seismograms $\mathbf{X}$ and vertical-component spectrogram $\mathbf{S}_{xx}^{Z}$ [Fig. 2(a)]. We denote seismogram data vector as $x^{c}(t), c \in \{E, N, Z\}, t \in \mathbb{N}$, in which c is the component index standing for east–west (E), north–south (N), and up–down (Z) component of the recording ground motion at given temporal index t. The input three-component seismogram matrix can then be defined as $\mathbf{X} = [x^{E}(t), x^{N}(t), x^{Z}(t)], t \in [t_0, t_0 + T]$, in which $t_0 = 0$ is the starting temporal index and $T = 6000$ is the duration of recordings (60-s-long recordings sampled at 100 Hz) in this study.

The vertical-component spectrogram matrix $\mathbf{S}_{xx}^{Z}$ is computed from $x^{Z}(t)$ using the squared absolute value of discrete short-time Fourier transform (STFT) with a Hann window function w in the length of $N = 20$ (i.e., 0.2 s) and a hop size $H = 10$ (i.e., 0.1 s) that determines how many samples to shift across $x^{Z}(t)$. $M = 100$ (i.e., 1 s) is the number of samples used for the Fourier transform. With regard to the given parameters,

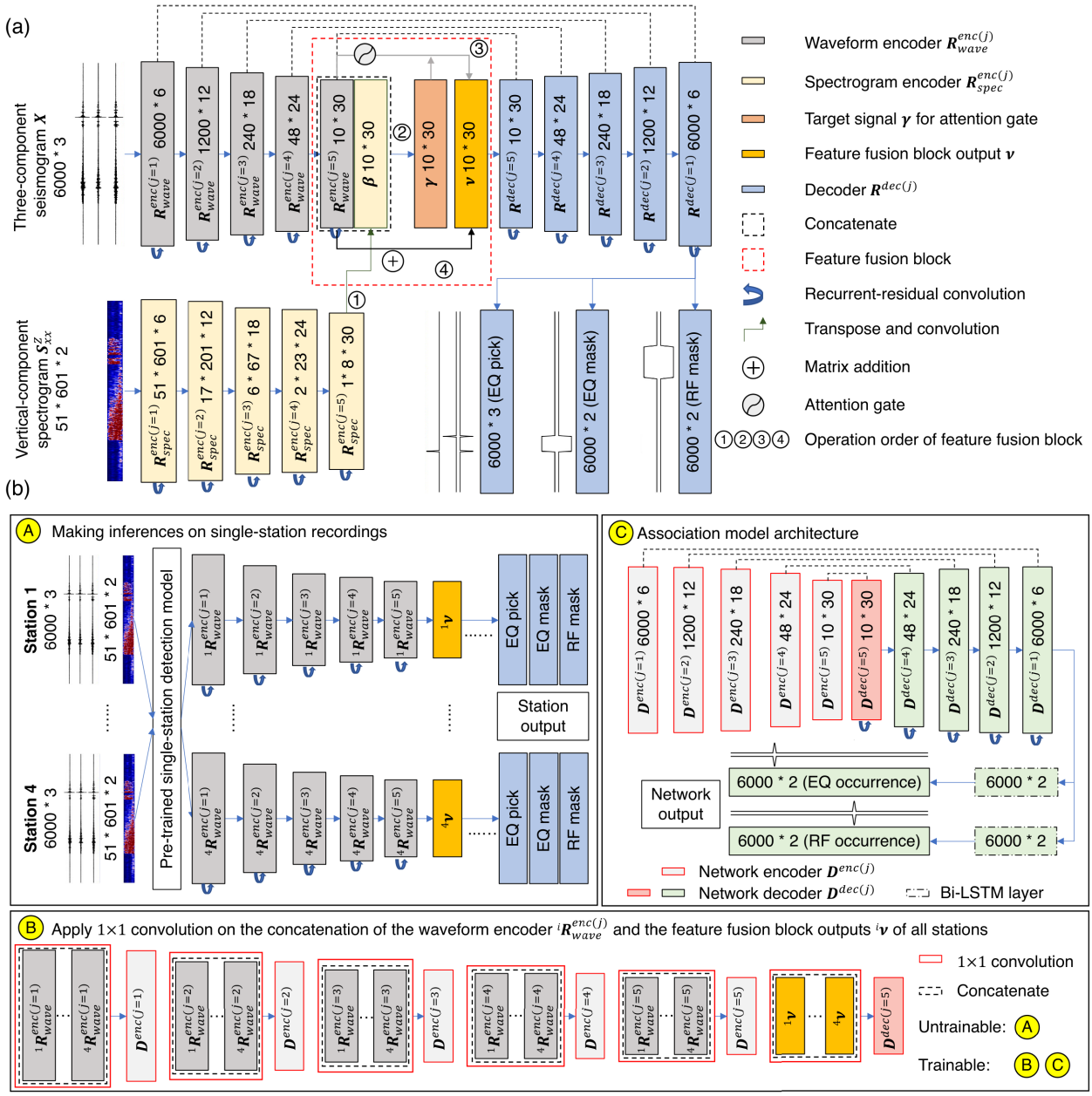


Fig. 2. Model architecture. (a) Inputs of the single-station detection model consist of three-component seismograms $\mathbf{X}$ and vertical-component spectrograms $\mathbf{S}^{Z}_{xx}$. These inputs are processed by separate encoders, namely the waveform encoder $\mathbf{R}^{enc(j)}_{wave}$ and the spectrogram encoder $\mathbf{R}^{enc(j)}_{spec}$. The waveform encoder and the spectrogram encoder fuse at the feature fusion block, which also connects to the model decoder $\mathbf{R}^{dec(j)}$. The outputs of the single-station detection model include earthquake phase-time functions, earthquake waveform masking functions, and rockfall waveform masking functions. (b) Association model concatenates and merges the waveform encoder tensors from all station recordings, which are produced by the trained single-station detection model. The merged encoder is then skip-connected to a new trainable decoder $\mathbf{D}^{dec(j)}$ to generate network outputs, specifically the local occurrence time functions of earthquake and rockfall events.

the spectrogram is computed by

$$\mathbf{S}^{Z}_{xx}(k,m) = \left|\mathbf{F}^{Z}(k,m)\right|^2 = \left|\sum_{n=0}^{N-1} x^{Z}[n+mH]w[n]e^{-i\frac{2\pi}{M}nk}\right|^2. \tag{1}$$

The discrete STFT involves dividing the signal into overlapping segments $x[n+mH]$, applying a window function, and computing the Fourier transform of each windowed segment. The resulting $\mathbf{S}^{Z}_{xx}(k,m)$ is the real-valued element of the spectrogram matrix, and $\mathbf{F}^{Z}(k,m)$ is the complex-valued element of the STFT matrix at time frame $m \in \mathbb{N}$ and frequency bin $k \in \mathbb{N}$. Before computing STFT, we pad zeros to both ends of $x^{Z}(t)$ to keep the maximal time frame equal to $L/H + 1 = 601$. The number of frequency bins is $M/2 + 1 = 51$. Finally, the shape of $\mathbf{S}^{Z}_{xx}$ is (51, 601, 2), in which the last axis refers to the real and imaginary values. To apply our method to continuous data in practical applications, we utilize a sliding window approach on the seismograms of all stations. This involves computing matrix $\mathbf{X}$ and $\mathbf{S}^{Z}_{xx}$ per slide as inputs to RockNet, and then performing postprocessing on the output functions accordingly. For

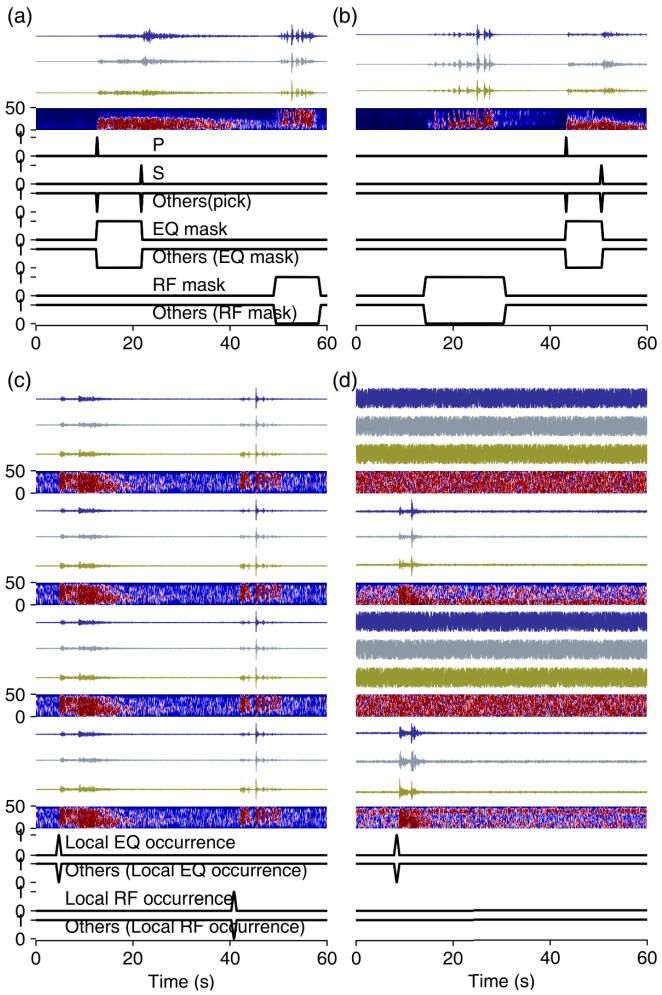


Fig. 3. Templates of the model input waveforms and corresponding target functions for (a) and (b) single-station detection model and (c) and (d) association model. RF and EQ are abbreviations of rockfall and earthquake events, respectively. We transform the time labels (i.e., the P and S arrival times and the local occurrences of earthquake and rockfall waveforms) into truncated Gaussian functions. (a) and (b) Marching mosaic waveforms are generated by randomly superimposing the earthquake and rockfall waveforms and shifting them back and forth. (c) Earthquake-rockfall mosaic waveforms generated for the single-station detection model are replicated four times as the augmentation data for the association model. (d) Earthquake event from the Luhu dataset with a random station order. We replace the waveform data with random values for the stations with missing data or ground-truth labels.

instance, we set a model input length of 6000 samples and a sliding interval of 1500 samples. By keeping track of the output rockfall occurrence function, one can identify any detected peaks that consistently maintain a high prediction value over four consecutive predictions (i.e., 6000/1500 = 4 times). This allows for effective monitoring of rockfall events on continuous data.

B. Single-Station Detection Model

An earthquake signal typically includes P and S phases [Fig. 1(c)]. In comparison, rockfall signals are induced by rock impacts with various sizes on the weight-bearing medium with rolling, free-falling, or bouncing movements. Thus, a rockfall waveform can appear as successive or discrete pulse-like signals with broad energy frequency ranges and is usually more easily identified in the time–frequency domain [Fig. 1(c)]. We take three-component seismograms $\mathbf{X}$ and vertical-component spectrogram $\mathbf{S}_{xx}^{Z}$ as the single-station detection model inputs [Fig. 2(a)].

The single-station detection model consists of: 1) two encoders processing two kinds of input data independently (i.e., three-component waveforms $\mathbf{X}$ for the waveform encoder and the vertical-component spectrogram $\mathbf{S}_{xx}^{Z}$ for the spectrogram encoder); 2) a feature fusion block merging the two encoder features; and 3) a decoder decoding the high-dimensional features from the feature fusion block and generate output sequences aligned with the input seismograms along time axis [Fig. 2(a)]. The two encoders separately map the Z score standardized input data to a high-dimensional feature space with a series of recurrent-residual convolution (RRC) blocks [30], [31], [32]. Let $\mathbf{R}_{\text{wave}}^{\text{enc}(j)}$ and $\mathbf{R}_{\text{spec}}^{\text{enc}(j)}$ be the jth level RRC layer of waveform encoder and spectrogram encoder, $\mathbf{R}_{\text{wave}}^{\text{enc}(j=5)} \in \mathbb{R}^{(10*30)}$ and $\mathbf{R}_{\text{spec}}^{\text{enc}(j=5)} \in \mathbb{R}^{(1*8*30)}$ are the last layer of both encoders in this study, which are also the inputs for the feature fusion block.

Let conv$(\cdot)$ denote the general convolution operation in our network and $[\cdot]$ being the concatenation operation. We define $\boldsymbol{\alpha} = \mathbf{R}_{\text{wave}}^{\text{enc}(j=5)}$, $\boldsymbol{\beta} = \text{conv}(\mathbf{R}_{\text{spec}}^{\text{enc}(j=5)})$ that expands and transforms the shape of $\mathbf{R}_{\text{spec}}^{\text{enc}(j=5)}$ to be aligned with that of $\mathbf{R}_{\text{wave}}^{\text{enc}(j=5)}$ [Fig. 2(a①)], and then $\boldsymbol{\gamma} = \text{conv}([\alpha, \beta]) \in \mathbb{R}^{(10*30)}$ is available [Fig. 2(a②)]. The products of feature fusion block can be simplified as $\boldsymbol{\nu} = \boldsymbol{\alpha} + \text{AG}(\text{target} : \boldsymbol{\gamma}; \text{gate} : \boldsymbol{\alpha})$ [Fig. 2(a④)], where the attention gate AG(target : $\boldsymbol{\gamma}$; gate : $\boldsymbol{\alpha}$) [32], [33] computes the attention score using target signal $\boldsymbol{\gamma}$ and gating signal $\boldsymbol{\alpha}$ from a coarser level [Fig. 2(a③)]

$$\text{AG}(\boldsymbol{\gamma}; \boldsymbol{\alpha}) = \boldsymbol{\gamma} \cdot \sigma\{\text{conv}(\boldsymbol{\rho}(\text{conv}(\boldsymbol{\gamma}) + \text{conv}(\boldsymbol{\alpha})))\} \quad (2)$$

in which $\rho(\cdot)$ and $\sigma(\cdot)$ separately represents the ReLU and Sigmoid activation function. The decoder $\mathbf{R}^{\text{dec}(j)}$ upsamples its first input feature map $\boldsymbol{\nu}$ with a series of RRC blocks that skip-connected with the feature maps of waveform encoder at the same depth j (i.e., in the same shape of feature map). The three output vectors associated with the assigned tasks of earthquake P and S phase arrival time picking, earthquake waveform masking, and rockfall waveform masking are then branched from $\mathbf{R}^{\text{dec}(j)=1}$ with regular convolution layers. Here, the P and S phase-time functions label the P and S arrivals along $\mathbf{X}$, and the earthquake waveform mask wraps the P–S pair of the same waveform. We considered the earthquake waveform mask as an additional constraint to facilitate the process of earthquake location [34], [35].

C. Association Model

The encoders of the single-station detection model learn to extract informative features from the input data to accomplish the assigned tasks. The core idea of the association model is to assemble the encoder representations of the trained single-station detection model for all station recordings and determine the seismic event occurrence times, which are highly related to the tasks of the single-station model. The association model is also constructed using a U-Net architecture in which the

encoder and the first layer of the decoder are based on the pretrained single-station detection model [Fig. 2(b)].

Specifically, the feature maps for transfer learning are from the waveform encoder ${}^{i}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}, j \in [1, 5], i \in \mathbb{N}$ and the feature fusion block ${}^{i}\boldsymbol{\nu}$, where i is the station index [Fig. 2(bⒶ)]. The jth level layer of the association model encoder $\mathbf{D}^{\text{enc}(j)}, j \in [1, 5]$ is obtained by performing a 1×1 convolution on the concatenation of ${}^{i}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}$ (i.e., $[{}^{1}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}; {}^{2}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}; {}^{3}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}; {}^{4}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}]$ in this study [Fig. 2(bⒷ)]). The first layer of the association model decoder at depth $j = 5$, $\mathbf{D}^{\text{dec}(j=5)}$, is obtained by performing a 1×1 convolution on the concatenation of ${}^{i}\boldsymbol{\nu}$ (i.e., $[{}^{1}\boldsymbol{\nu}; {}^{2}\boldsymbol{\nu}; {}^{3}\boldsymbol{\nu}; {}^{4}\boldsymbol{\nu}]$) [Fig. 2(bⒷ)]. Here, the pretrained weights from the feature maps ${}^{i}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}$ and ${}^{i}\boldsymbol{\nu}$ are not trainable during the training process of the association model. In this way, we can assure the knowledge transfer from the pretrained single-station detection model. The rest of four RRC layers (i.e., $\mathbf{D}^{\text{dec}(j=4)}$, $\mathbf{D}^{\text{dec}(j=3)}$, $\mathbf{D}^{\text{dec}(j=2)}$, and $\mathbf{D}^{\text{dec}(j=1)}$) of the decoder are trained from scratch, while skip-connected with the encoder for learning to interpret the assemblages of ${}^{i}\mathbf{R}_{\text{wave}}^{\text{enc}(j)}$ and ${}^{i}\boldsymbol{\nu}$ [Fig. 2(bⒸ)]. Finally, the output layers of determining the occurrence of earthquake and rockfall are branched from $\mathbf{D}^{\text{dec}(j=1)}$ with a preceding Bi-LSTM layer [36], respectively.

Generally, the single-station detection model is independent but included in the association model, and the prediction results of the two models do not interfere with each other. Rather than training a single model that can produce the same outputs, we build RockNet with two related but independent models. The reason for doing so is to explore the methods of transfer learning. Also, many more computing resources are needed to train a new model that optimizes both the single-station detection and association tasks from scratch.

D. Target Function Prototypes and Model Optimization

RockNet optimizes two types of target functions: mask functions that wrap the target waveforms and truncated Gaussian functions that are transformed from the ground-truth time labels of the P and S phase arrivals of earthquakes and local seismic event occurrences with standard deviations of 0.2, 0.3, and 0.5 s [Fig. 3]. The standard deviation values of 0.2 and 0.3 are obtained with model calibration experiments [32]. The time index of P arrival, S arrival, and event occurrences can be retrieved by the peak values of the truncated Gaussian functions. Overall, the two models of RockNet are built under a multitask learning framework with relevant tasks. The single-station model optimizes three tasks: seismic phase arrival time picking; earthquake waveform masking; and rockfall waveform masking [Fig. 2(a)]. The association model optimizes two functions for estimating the local occurrences of earthquake and rockfall events, in which the truncated Gaussian functions are centered at the sample 0.5 s before the first label of all stations [Fig. 2(b)]. For every task, we add a function of the "others" class to meet the softmax normalization criterion of the model output layer, which squeezes the model outputs to [0, 1] and sums them up to 1 for every sample along the time axis. During model training, cross-entropy is applied as a loss function $\mathbf{H}$ to estimate the differences between the current predictions $\mathbf{p}$ (i.e., the softmax-normalized outputs) and the target functions $\mathbf{q}$ for model optimization purposes; the optimizer is Adam [37] with a learning rate of 1e-4 in this study

$$\mathbf{H}(\mathbf{p}, \mathbf{q}) = -\sum_{x}^{6000} \sum_{1}^{N} \mathbf{p}_N(x) \log \mathbf{q}_N(x). \tag{3}$$

In (3), x refers to the number of samples along the time axis (a total of 6000 samples in this study) and N refers to the number of target function vectors (three for the earthquake phase arrival time picking functions and two for the other functions). The total loss L of training epoch I is defined as the weighted sum of the cross-entropy values computed for each task $\mathbf{k}$

$$L(I) = \sum_{\mathbf{k}}^{N} \lambda_{\mathbf{k}}(I) \mathbf{H}(\mathbf{p}(I), \mathbf{q}(I))_{\mathbf{k}} \tag{4}$$

where the task weightings λ are estimated using the dynamic weight averaging (DWA) strategy [30], [38].

E. Additional Training Dataset and Data Augmentation

The challenge of limited sample availability has long been a concern in training rockfall detection algorithms. However, this study addresses this issue by incorporating a substantial dataset of hundreds of thousands of earthquake and nonearthquake samples for model training. Within the multitask learning framework, the inclusion of a large non-rockfall dataset aids in enhancing the model's interpretability of seismograms and enables the development of pattern recognition capabilities for rockfall waveforms by leveraging information from different tasks.

To further enhance the model's generalizability, two augmentation approaches are employed: the marching mosaic waveform augmentation (MMWA) and the earthquake early warning augmentation (EEWA). The MMWA technique involves superimposing multiple earthquake samples on each other and applying shifts to the semisynthetic waveforms in both forward and backward directions. The EEWA approach, on the other hand, employs time-clipped earthquake waveforms where only P-waves are visible within the receptive field of the model, as described in Liao et al. [30]. The earthquake and nonearthquake samples used in this study were collected and generated from the Taiwan earthquake dataset and the STEAD dataset for the single-station model, while the INSTANCE dataset was utilized for developing the association model. Each sample consists of four station recordings with a random station order. The training and validation data for both models were split with an 80%–20% ratio.

When dealing with a large number of additional earthquake training samples, the issue of imbalanced learning arises due to severely skewed class distributions. To address this challenge, we employ the MMWA strategy, which involves superimposing and randomly shifting the rockfall and earthquake waveforms [Fig. 3] to augment the datasets. This augmentation technique is applied to the training sets of

TABLE I
MODEL PERFORMANCE ACHIEVED ON THE LUHU TESTING DATASET*

	Single-station detection model						Association model					
Description	Compare the effect of input data						Compare the effect of training sample size					
Model	Waveform + Vertical-component spectrogram			Waveform			Luhu dataset + INSTANCE			Luhu dataset		
Training samples	393K			393K			440K			47K		
Class	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others
Threshold	0.5 (optimal)			0.5 (optimal)			0.3 (optimal)			0.1 (optimal)		
Test size	1581	218	980	1581	218	980	551	100	346	551	100	346
Precision	0.990	0.981	0.991	0.989	0.995	0.967	0.993	0.989	0.997	0.988	0.989	0.969
Recall	1.000	0.959	0.980	0.999	0.853	0.982	1.000	0.970	0.991	1.000	0.890	0.977
F1-score	0.995	0.970	0.985	0.994	0.919	0.974	0.996	0.980	0.994	0.994	0.937	0.973
Macro F1-score	0.983			0.962			0.990			0.968		

*EQ, RF, and Others refer to earthquakes, rockfalls, and other signals that typically refers to car-induced vibration and engineering signals, respectively.

both the single-station model and the association model. For the association model, which takes inputs from four stations, we replicate the single-station mosaic waveforms four times as the inputs for the association model. To prevent the model from memorizing specific orders of recorded stations and waveforms, we randomly set the station order of the input tensor. It should be noted that not all events in the Luhu dataset have four station recordings. Some samples may have missing data or insufficiently distinct target waveforms for labeling. To address this, we complement or substitute these samples with random values in the association model to ensure robust training and prevent bias toward specific events.

IV. RESULTS

A. Benchmark Test on the Luhu Dataset

For the benchmark test on the Luhu dataset, we computed the evaluation metrics for three classes: rockfall, earthquake, and others, in which the last class involves car-induced vibrations and engineering noises. Table I lists the best performance of all models on the Luhu dataset, in which the optimal positive detection threshold is determined by $\text{argmax}(\text{F1}(\tau)), \tau \in [0.1, 0.2, \ldots, 0.8]$, with τ being the threshold value and $\text{F1}(\tau)$ being the macro F1-score under τ. Considering the single-station detection model, we compare the performances achieved by the model with and without the vertical-component spectrogram serving as the model input. With the additional time–frequency information, the model improves its rockfall recall rate by 10.6%, indicating fewer false-negative detections (i.e., misidentifying rockfalls as earthquakes and others). The performance improvement conforms to our experiences in distinguishing between different seismic sources, in which time–frequency representations help when time-series features are not highly typical and vice versa. For the association model, we explore the training data quantity factor under the transfer learning framework. The association model that trains with only the Luhu dataset achieves a macro-F1-score of 0.968. With the additional training dataset of INSTANCE, the total dataset gets more imbalanced. However, the association model performs better with a macro-F1-score of 0.990, in which the rockfall recall rate improves by 8%. We infer that its advances are derived from the positive transfer of knowledge from the well-trained single-station detection model. Fig. 4(a) shows prediction examples in which the rockfall waveform amplitude, signal-to-noise ratio (SNR), and high-frequency energy attenuate with increasing distance. In Fig. 4(b), the rockfall waveform is unclear at stations LH03 and LH04, and the P wave of the earthquake interferes with the rockfall waveform, while other

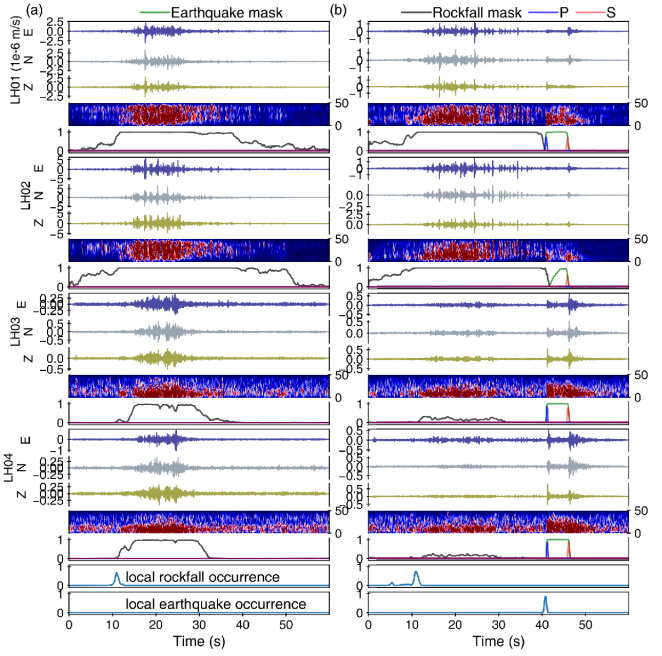


Fig. 4. Detection examples of rockfall events that occur between stations LH01 and LH02. These examples are not included in the training data. (a) and (b) Overall rockfall waveform amplitudes, SNRs, and high-frequency energy attenuation processes from stations LH01 and LH02 to stations LH03 and LH04 are conspicuous. (b) P wave of the earthquake strongly interferes with the rockfall waveform, and the single-station model does not detect the arrival of the phase at station LH02. The rockfall waveform amplitude and SNR discrepancies are more evident than the earthquake waveforms with same-scale amplitudes over all stations. Nevertheless, the association model still identifies the earthquake and rockfall occurrences with high prediction values.

TABLE II
ASSOCIATION MODEL PERFORMANCE ON THE SUPER-SAUZE DATASET*

Model	Provost et al., (2017) [18]			RockNet			RockNet		
Class	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others
Note	Randomly selected			Randomly selected			All samples		
Threshold	-			0.3 (optimal)			0.3 (optimal)		
Test size	70	70	70	70	70	70	391	406	351
Precision	0.959	0.949	0.894	0.971	1.000	0.932	0.967	0.943	0.870
Recall	0.913	0.940	0.949	0.971	0.943	0.986	0.969	0.902	0.912
F1-score	0.935	0.945	0.921	0.971	0.971	0.958	0.968	0.922	0.890
Macro F1-score	0.934			0.967			0.927		

*In the Super-Sauze dataset we do not consider the "quake" events. Class "Others" typically refers to anthropogenic/environmental noises in the original paper. Ó

TABLE III
ASSOCIATION MODEL PERFORMANCE ON THE SUPER-SAUZE DATASET COMPARED TO MODELS OF JIANG ET AL. [39]*

Model	RockNet			Time-series [39]			STFT [39]			CWT [39]		
Class	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others
Note	Test data of [39]											
Test size	116	120	105	116	120	105	116	120	105	116	120	105
Precision	0.982	0.936	0.807	0.991	0.892	0.924	0.974	0.925	0.867	0.991	0.908	0.886
Recall	0.957	0.850	0.914	0.991	0.939	0.874	0.958	0.902	0.910	0.983	0.908	0.894
F1-score	0.969	0.891	0.857	0.991	0.915	0.898	0.966	0.914	0.888	0.987	0.908	0.890
Macro F1-score	0.906			0.935			0.922			0.929		

*Time-series, STFT, CWT refers to the models of input data in the form of time-series waveform, short-time Fourier transform, and continuous wavelet transform.

TABLE IV
ASSOCIATION MODEL PERFORMANCE ON THE LUHU DATASET COMPARED TO MODELS OF JIANG ET AL. [39]

Model	RockNet			Time-series [39]			STFT [39]			CWT [39]		
Class	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others	EQ	RF	Others
Note	Test data of the Luhu dataset in this study											
Test size	551	100	346	551	100	346	551	100	346	551	100	346
Precision	0.993	0.989	0.997	0.900	0.120	0.676	0.570	0.680	0.538	0.835	0.490	0.653
Recall	1.000	0.970	0.991	0.747	0.571	0.750	0.885	0.231	0.536	0.781	0.430	0.769
F1-score	0.996	0.980	0.994	0.817	0.198	0.711	0.696	0.344	0.537	0.807	0.458	0.706
Macro F1-score	0.990			0.575			0.525			0.657		

station recordings can help to confirm the coherency of the rockfall and earthquake occurrences.

B. Examination on the Super-Sauze Unstable Slope Dataset

We evaluated the generalization of our trained model using an independent dataset collected from the Super-Sauze unstable slope [18]. The labeled dataset contains anthropogenic/environmental noises (351 events), earthquakes (391 events), rockfalls (406 events), and "quake" signals (235 events) that we excluded due to potential ambiguity in label meaning across different datasets. We considered the anthropogenic/environmental noises as "other" events relative to rockfall and earthquake events. The dataset was collected from three time periods (11 October–19 November 2013, 10–30 November 2014, and 9 June–15 August 2015) using six one-component short-period seismometers (Noemax and Sercel L4C) and two three-component broadband seismometers (RefTek 130S-01). The seismometers were arranged in two equilateral triangular arrays, with a three-component seismometer placed at the center surrounded by three vertical one-component seismometers on the east and west sides of the Super-Sauze unstable slope. To meet the input criterion of four three-component seismograms for our model, we treated the three vertical one-component seismograms as a three-channel vector. Our association model achieved a macro-F1-score of 0.927 on all samples, with F1 scores of 0.968, 0.922, and 0.890 for earthquakes, rockfalls, and others (anthropogenic/environmental noises), respectively (Table II).

C. Compared With Other Developed Machine-Learning Models

To compare RockNet with the random forest model from [18], we excluded the labeled "quake" samples and derived the confusion matrix components (true-positive, true-negative, false-positive, and false-negative) for the earthquake, rockfall, and anthropogenic/environmental noise classes using the data from [18, Table 2]. We then computed the evaluation metrics for earthquake, rockfall, and "others" class, which includes anthropogenic/environmental noise and all samples classified as "quake" events. Provost et al. [18] tested on

70 samples for each class, with the rest used for training, and the results were averaged over 100 runs. We also computed the evaluation metrics using randomly selected 70 samples for each class, 100 times, and found that the average results were similar to those reported in [18], despite not using any Super-Sauze unstable slope data in our model training process. These results demonstrate that our model generalizes well to independent datasets, which may be due to the similar physical mechanisms underlying events in different regions, as reflected in the seismic recordings [40].

In a recent study by Jiang et al. [39], three deep-learning models were trained on the Super-Sauze dataset using six-channel input data with a duration of 10 s and a sampling rate of 250 Hz, provided in three forms: raw time-series seismograms, STFT spectrograms, and continuous wavelet transform (CWT) spectrograms. The models classify anthropogenic/environmental noises, earthquakes, rockfalls, and "quake" events using one-hot encoded outputs. We conducted a benchmark test on the same test data used by Jiang et al. [39], excluding the "quake" events, and report the results in Table III. While the models developed by Jiang et al. [39] generally outperform our model on the Super-Sauze dataset, our model achieved a noteworthy macro-F1-score, approximately 3% lower than theirs. Furthermore, we compared RockNet and the models developed by Jiang et al. [39] on the Luhu dataset in our study, presenting the results in Table IV. We randomly selected data from two stations with manual labels, resulting in six input channels for the models developed by [39]. The comparison shows a significant performance gap between the models developed by [39] on the Super-Sauze and Luhu datasets. Our cross-validation of the models on both datasets highlights the superior generalizability and robustness of RockNet.

V. Discussion

Accurately discriminating between rockfalls, earthquakes, and other sources using seismic recordings is a challenging and crucial task. Unlike earthquakes, which typically exhibit body waves, waveforms of rockfalls vary greatly depending on the type, amount of rocks, and dynamics of the event. To distinguish between seismic recordings of rockfalls and other sources, previous studies have used characteristics such as waveform and spectral features. For example, Provost et al. [18] utilized 71 attributes on the waveform and spectral features to build a random forest model. Other studies have used features on spectrograms of waveforms to develop deep-learning models (e.g., [41], [42]), or have used waveforms and different spectrograms individually for different deep-learning models [39]. However, these previous studies either relied on expert-selected features or used seismic waveforms and spectrograms individually. In contrast, RockNet uses both seismic waveforms and spectrograms jointly to distinguish different seismic sources, achieving better performance than using only one of them (Table I). Moreover, RockNet leverages the feature maps from the single-station model and incorporates them into the association model [Fig. 2], which estimates the probabilities of rockfall occurrences based on the seismic recordings obtained from the local seismic network. Additionally, RockNet addresses the challenge of incomplete waveforms by using MMWA, which applies random time shifts in training data [30]. By using multiple features in a multitask learning framework, RockNet achieves high accuracy in discriminating between different rockfalls, earthquakes, and other sources and has potential applications in real-time monitoring and hazard assessment.

In our comparisons of the RockNet with other deep-learning models (Tables III and IV), the RockNet achieved a macro-F1-score of over 0.9 on the Super-Sauze dataset, while other models [39] achieved a macro-F1-score of lower than 0.7 on the Luhu dataset. The superior generalizability of the RockNet on an independent dataset could be attributed to its joint use of both waveforms and spectrograms for detecting rockfall waveforms, whereas other models only used one of these features independently. However, we also recognize that other factors can affect model performance, such as the quality and quantity of training data used. In this case, the length of input data used in training may be a critical factor influencing model performance. For instance, while the RockNet takes 60 s of seismic waveforms and spectrograms as inputs, models developed by [39] only takes 10 s of seismic waveforms or spectrograms as inputs. However, it should be noted that some rockfall events may take longer than 10 s to complete, as illustrated in Fig. 1(b) and (c). It is important to note that besides the model architecture, there are other factors that can significantly affect the performance of a deep learning model.

While the RockNet model has demonstrated promising performance in detecting and characterizing rockfall events, there are several limitations to consider. One major limitation is the scarcity of seismic recordings of rockfalls, which can hinder the model's ability to generalize to other sites and conditions. Due to the small energy released by rockfall events, only the seismometers located near the site can record them. Moreover, the complex dynamic processes and different types of rockfalls (e.g., rockfalls, debris falls, or mixed) can result in various features of seismic recordings, which can further challenge the development of a robust and generalized model. Compared to earthquakes, there are considerably fewer available seismic recordings of rockfall events. This shortage of data may limit the ability of the RockNet and other deep-learning models to achieve high performance on independent datasets. For instance, while models developed for earthquake waveform detection or seismic phase picking often exhibit similar performances on unseen datasets from other study areas, the RockNet and the models developed by [39] show varying degrees of performance degradation on independent datasets (Tables III and IV). To overcome this limitation, it is essential to collect more diverse seismic recordings of rockfalls to develop models that can generalize well to different sites and conditions.

To develop a more reliable machine-learning model for rockfall event detection, efforts can be made in both data collection and advanced machine-learning techniques. The availability of a large dataset that contains high-quality seismic recordings of rockfalls will benefit both model training and evaluation of model performance. Unsupervised

clustering methods have recently shown promise in grouping seismic recordings according to their waveform features without relying on manual labeling. This approach could be useful in providing a more objective and efficient identification of recordings of rockfalls, reducing the need for manual identifications and potentially improving the quality of datasets [43]. Additionally, advanced machine-learning techniques can be adopted to further improve the accuracy of the model. For example, physics-informed neural networks (PINNs) [44] can be used to incorporate physical equations into the learning process as rigid restrictions, potentially improving the accuracy and interpretability of the models. Advanced deep ensemble learning approaches [45] can be used to ensemble information from multiple stations, similar to our association model, for more reliable detection of rockfall recordings. With ongoing developments in deep learning and the increasing availability of data and resources, further optimization and advancements in seismic event discrimination are expected in the future.

VI. Conclusion

In this study, we presented RockNet, a deep-learning-based model for detecting rockfall and earthquake events in local seismic networks. By jointly utilizing the waveforms and spectrograms of seismic recordings and associating information from individual stations, RockNet achieves high accuracy in event detection. Additionally, we addressed the challenge of imbalanced learning through the use of multitask learning and transfer learning strategies, as well as a data augmentation technique that superimposes and shifts waveforms. Our model exhibits strong performance on both testing and independent datasets, indicating its potential for use in onsite earthquake and rockfall monitoring. However, the limited availability of seismic recordings of rockfalls currently poses a challenge to developing even more reliable models. To further improve the accuracy and interpretability of these models, future work could explore advanced machine-learning techniques. Additionally, ongoing advancements in deep learning and the increasing availability of data and resources will likely lead to further progress in seismic event discrimination. We anticipate that the development of more reliable models for detecting rockfall and earthquake events in local seismic networks will contribute to improved onsite monitoring and disaster management in the future.

VII. Open Research

The labeled data [46] are available in Dryad (https://doi.org/10.5061/dryad.tx95x6b2f). The code and model are open sources at GitHub (https://github.com/tso1257771/RockNet) and Zenodo (https://doi.org/10.5281/zenodo.7458571, [47]).

Acknowledgment

This work was supported by the National Center for High-Performance Computing (NCHC), Taiwan, for providing some computational and storage resources. The continuous seismic data of the Luhu unstable slope were provided by the National Science and Technology Center for Disaster Reduction, Taiwan. The continuous seismic data of the Super-Sauze unstable slope were provided by the Observatoire Multidisciplinaire des Instabilités de Versant [48]. The authors also thank Dr. Wei-An Chao and Dr. Jui-Ming Chang for providing artificial rockfall experiments waveforms for toy tests during model development.

References

[1] M. Krautblatter and M. Moser, "A nonlinear model coupling rockfall and rainfall intensity based
newline on a four year measurement in a high Alpine rock wall (Reintal, German Alps)," *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 4, pp. 1425–1432, Aug. 2009. [Online]. Available: https://nhess.copernicus.org/articles/9/1425/2009/

[2] A. Helmstetter and S. Garambois, "Seismic monitoring of Séchilienne rockslide (French Alps): Analysis of seismic signals and their correlation with rainfalls," *J. Geophys. Res.*, vol. 115, Aug. 2010, doi: 10.1029/2009jf001532.

[3] A. Delonca, Y. Gunzburger, and T. Verdel, "Statistical correlation between meteorological and rockfall databases," *Natural Hazards Earth Syst. Sci.*, vol. 14, no. 8, pp. 1953–1964, Aug. 2014. [Online]. Available: https://nhess.copernicus.org/articles/14/1953/2014/

[4] J. D'Amato, D. Hantz, A. Guerin, M. Jaboyedoff, L. Baillet, and A. Mariscal, "Influence of meteorological factors on rockfall occurrence in a middle mountain limestone cliff," *Natural Hazards Earth Syst. Sci.*, vol. 16, no. 3, pp. 719–735, Mar. 2016. [Online]. Available: https://nhess.copernicus.org/articles/16/719/2016/

[5] M. Lato, J. Hutchinson, M. Diederichs, D. Ball, and R. Harrap, "Engineering monitoring of rockfall hazards along transportation corridors: Using mobile terrestrial LiDAR," *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 3, pp. 935–946, Jun. 2009. [Online]. Available: https://nhess.copernicus.org/articles/9/935/2009/

[6] H. Lan, C. D. Martin, C. Zhou, and C. H. Lim, "Rockfall hazard analysis using LiDAR and spatial modeling," *Geomorphology*, vol. 118, nos. 1–2, pp. 213–223, May 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169555X10000127

[7] M. J. Lato, M. S. Diederichs, D. J. Hutchinson, and R. Harrap, "Evaluating roadside rockmasses for rockfall hazards using LiDAR data: Optimizing data collection and processing protocols," *Natural Hazards*, vol. 60, no. 3, pp. 831–864, Jun. 2011, doi: 10.1007/s11069-011-9872-y.

[8] M. J. Royán, A. Abellán, M. Jaboyedoff, J. M. Vilaplana, and J. Calvet, "Spatio-temporal analysis of rockfall pre-failure deformation using terrestrial LiDAR," *Landslides*, vol. 11, no. 4, pp. 697–709, Nov. 2013, doi: 10.1007/s10346-013-0442-0.

[9] A. M. Fanos, B. Pradhan, S. Mansor, Z. M. Yusoff, and A. F. B. Abdullah, "A hybrid model using machine learning methods and GIS for potential rockfall source identification from airborne laser scanning data," *Landslides*, vol. 15, no. 9, pp. 1833–1850, Apr. 2018, doi: 10.1007/s10346-018-0990-4.

[10] N. Matsuoka, "A multi-method monitoring of timing, magnitude and origin of rockfall activity in the Japanese Alps," *Geomorphology*, vol. 336, pp. 65–76, Jul. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169555X19301230

[11] C. Hibert, A. Mangeney, G. Grandjean, and N. M. Shapiro, "Slope instabilities in Dolomieu crater, Réunion Island: From seismic signals to rockfall characteristics," *J. Geophys. Res.*, vol. 116, no. 4, 2011. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JF002038

[12] F. Dammeier, J. R. Moore, F. Haslinger, and S. Loew, "Characterization of Alpine rockslides using statistical analysis of seismic signals," *J. Geophys. Res.*, vol. 116, Nov. 2011, doi: 10.1029/2011jf002037.

[13] V. L. Zimmer and N. Sitar, "Detection and location of rock falls using seismic and infrasound sensors," *Eng. Geol.*, vol. 193, pp. 49–60, Jan. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013795215001234

[14] C. Hammer, M. Ohrnberger, and D. Fäh, "Classifying seismic waveforms from scratch: A case study in the Alpine environment," *Geophys. J. Int.*, vol. 192, no. 1, pp. 425–439, Jan. 2013, doi: 10.1093/gji/ggs036.

[15] C. Hibert et al., "Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise Volcano," *J. Geophys. Res., Earth Surf.*, vol. 119, no. 5, pp. 1082–1105, May 2014, doi: 10.1002/2013jf002970.

[16] A. Maggi, V. Ferrazzini, C. Hibert, F. Beauducel, P. Boissier, and A. Amemoutou, "Implementation of a multistation approach for automated event classification at Piton de la Fournaise Volcano," *Seismological Res. Lett.*, vol. 88, no. 3, pp. 878–891, Mar. 2017, doi: 10.1785/0220160189.

[17] C. Hibert, F. Provost, J.-P. Malet, A. Maggi, A. Stumpf, and V. Ferrazzini, "Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise Volcano using a random forest algorithm," *J. Volcanol. Geothermal Res.*, vol. 340, pp. 130–142, Jun. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377027316303948

[18] F. Provost, C. Hibert, and J.-P. Malet, "Automatic classification of endogenous landslide seismicity using the random forest supervised classifier," *Geophys. Res. Lett.*, vol. 44, no. 1, pp. 113–120, Jan. 2017, doi: 10.1002/2016gl070709.

[19] M. Malfante, M. D. Mura, J. I. Mars, J.-P. Métaxian, O. Macedo, and A. Inza, "Automatic classification of volcano seismic signatures," *J. Geophys. Res., Solid Earth*, vol. 123, no. 12, pp. 10645–10658, Dec. 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JB015470

[20] 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 Earth Syst. Sci.*, vol. 21, no. 1, pp. 339–361, Jan. 2021. [Online]. Available: https://nhess.copernicus.org/articles/21/339/2021/

[21] M. Curilem et al., "Pattern recognition applied to seismic signals of Llaima Volcano (Chile): An evaluation of station-dependent classifiers," *J. Volcanol. Geothermal Res.*, vol. 315, pp. 15–27, Apr. 2016, doi: 10.1016/j.jvolgeores.2016.02.006.

[22] M. Malfante, M. D. Mura, J. Metaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018, doi: 10.1109/msp.2017.2779166.

[23] P. Bottelin et al., "Seismic and mechanical studies of the artificially triggered rockfall at Mount Néron (French Alps, December 2011)," *Natural Hazards Earth Syst. Sci.*, vol. 14, no. 12, pp. 3175–3193, Dec. 2014, doi: 10.5194/nhess-14-3175-2014.

[24] L. Feng, V. Pazzi, E. Intrieri, T. Gracchi, and G. Gigli, "Rockfall seismic features analysis based on in situ tests: Frequency, amplitude, and duration," *J. Mountain Sci.*, vol. 16, no. 5, pp. 955–970, May 2019, doi: 10.1007/s11629-018-5286-6.

[25] C. Hibert et al., "Machine learning prediction of the mass and the velocity of controlled single-block rockfalls from the seismic waves they generate," *EGUsphere*, vol. 2022, pp. 1–25, Jan. 2022. [Online]. Available: https://egusphere.copernicus.org/preprints/egusphere-2022-522/

[26] A. Ng. (2017). *Machine Learning Yearning*. [Online]. Available: http://www.mlyearning.org/(96)

[27] 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.

[28] A. Michelini, S. Cianetti, S. Gaviano, C. Giunchi, D. Jozinovic, and V. Lauciani. (2021). *Instance the Italian Seismic Dataset for Machine Learning*. [Online]. Available: http://www.pi.ingv.it/instance/

[29] R. Allen, "Automatic phase pickers: Their present use and future prospects," *Bull. Seismological Soc. Amer.*, vol. 72, no. 6B, pp. 225–242, Dec. 1982.

[30] W. Liao, E. Lee, D. Chen, P. Chen, D. Mu, and Y. Wu, "RED-PAN: Real-time earthquake detection and phase-picking with multitask attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 2900111.

[31] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[32] W.-Y. Liao, E.-J. Lee, D. Mu, P. Chen, and R.-J. Rau, "ARRU phase picker: Attention recurrent-residual U-Net for picking SeismicP- andS-phase arrivals," *Seismological Res. Lett.*, vol. 92, no. 4, pp. 2410–2428, Mar. 2021, doi: 10.1785/0220200382.

[33] J. Schlemper et al., "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[34] E.-J. Lee, W.-Y. Liao, D. Mu, W. Wang, and P. Chen, "GPU-accelerated automatic microseismic monitoring algorithm (GAMMA) and its application to the 2019 Ridgecrest earthquake sequence," *Seismological Res. Lett.*, vol. 91, no. 4, pp. 2062–2074, Jul. 2020.

[35] W. Liao, E. Lee, D. Mu, and P. Chen, "Toward fully autonomous seismic networks: Backprojecting deep learning-based phase time functions for earthquake monitoring on continuous recordings," *Seismological Res. Lett.*, vol. 93, no. 3, pp. 1880–1894, 2022, doi: 10.1785/0220210274.

[36] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[38] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1871–1880.

[39] J. Jiang, V. Stankovic, L. Stankovic, E. Parastatidis, and S. Pytharouli, "Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5906414, doi: 10.1109/tgrs.2023.3262412.

[40] F. Provost et al., "Towards a standard typology of endogenous landslide seismic sources," *Earth Surf. Dyn.*, vol. 6, no. 4, pp. 1059–1088, Nov. 2018. [Online]. Available: https://esurf.copernicus.org/articles/6/1059/2018/

[41] D. Lee, E. Aune, N. Langet, and J. Eidsvik, "Ensemble and self-supervised learning for improved classification of seismic signals from the Åknes rockslope," *Math. Geosci.*, vol. 55, no. 3, pp. 377–400, Nov. 2022, doi: 10.1007/s11004-022-10037-7.

[42] N. Langet and F. M. J. Silverberg, "Automated classification of seismic signals recorded on the Åknes rock slope, Western Norway, using a convolutional neural network," *Earth Surf. Dyn.*, vol. 11, no. 1, pp. 89–115, Feb. 2023, doi: 10.5194/esurf-11-89-2023.

[43] C.-C. Wang et al., "Cluster analysis of slope hazard seismic recordings based upon unsupervised deep embedded clustering," *Seismological Res. Lett.*, Apr. 2023, doi: 10.1785/0220230011.

[44] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045.

[45] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105151, doi: 10.1016/j.engappai.2022.105151.

[46] W. Liao, E.-J. Lee, C.-C. Wang, and P. Chen, "RockNet: Rockfall and earthquake detection and association via multitask learning and transfer learning," *IEEE Trans. Geosci. Remote Sens.*, early access, Jun. 8, 2023, doi: 10.1109/TGRS.2023.3284008.

[47] W. Liao, "Rocknet: Rockfall and earthquake detection and association via multitask learning and transfer learning," Dept. Earth Sci., Nat. Cheng Kung Univ., Tainan, Taiwan, Tech. Rep., Dec. 2022, doi: 10.5281/zenodo.7458571.

[48] Observatoire Multidisciplinaire Des Instabilites De Versant. (2012). *OMIV—Time Serie of Seismic Waveforms Recorded on Landslides and Catalogue of Endogenous Seismic Events*. [Online]. Available: https://dataosu.obs-besancon.fr/FR-18008901306731-2018-12-03-02_OMIV-Time-serie-of-seismic-waveforms-recorded.html

Wu-Yu Liao received the bachelor's and master's degrees in earth sciences from National Cheng Kung University, Tainan, Taiwan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Department of Earth Sciences.

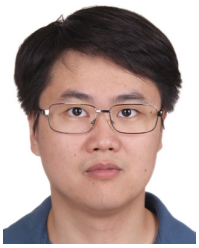
His research interests are deep-learning applications in seismology, especially, combining deep-learning- and physics-based algorithms and the real-world applications of deep-learning products.

En-Jui Lee received the master's degree in earth sciences from the State University of New York (SUNY) at Binghamton, Binghamton, NY, USA, in 2008, and the Ph.D. degree in geophysics from the University of Wyoming, Laramie, WY, USA, in 2013.

He is currently a Professor with the Department of Earth Sciences, National Cheng Kung University, Tainan, Taiwan. His research interests focus on earthquake monitoring, seismic hazard analysis, full-wave seismic tomography, earthquake source inversions, and applications of machine learning in seismology.

Chung-Ching Wang received the bachelor's and master's degrees in earth sciences from National Cheng Kung University, Tainan, Taiwan, in 2017 and 2022, respectively.

He is currently working as a Junior System Engineer with Innolux Corporation, Miaoli City, Taiwan.



Po Chen received the bachelor's degree in geophysics from Peking University, Beijing, China, in 2000, and the Ph.D. degree in geological sciences from the University of Southern California, Los Angeles, CA, USA, in 2005.

He was a Post-Doctoral Research Associate with the Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA, from 2006 to 2008. He joined the Department of Geology and Geophysics, University of Wyoming, Laramie, WY, USA, as an Assistant Professor in 2008 and was promoted to Associate Professor in 2014.

Floriane Provost received the Ph.D. degree in geophysics from the University of Strasbourg, Strasbourg, France, in 2018.

She worked as a Post-Doctoral Researcher with European Space Agency (ESA/ESRIN), Frascati, Italy, and the University Laval, Quebec, QC, Canada. She is currently a Post-Doctoral Researcher with the French National Centre for Scientific Research (CNRS), Paris, France. Her research interests include landslide monitoring using remote sensing, seismology, machine learning, and big data analysis.

Clément Hibert received the Engineering degree in geophysics and the Ph.D. degree in seismology from the Institut de Physique du Globe de Paris, Paris, France, in 2012.

He is currently an Assistant Professor with the University of Strasbourg, Strasbourg, France. He is managing seismological data for the French Landslides Observatory (OMIV), Strasbourg. His research focuses on environmental seismology and how we can mobilize machine-learning methods to explore large seismological datasets, in particular, to study landslides.

Jean-Philippe Malet received the Ph.D. degree in engineering geology and the Habilitation degree in geophysics from the University of Strasbourg, Strasbourg, France, in 2003 and 2014, respectively.

He is currently the Director of Research with the School and Observatory of Earth Sciences (EOST), French National Centre for Scientific Research (CNRS), Paris, France. He is responsible for the French Landslide Observatory (OMIV). His research interests focus on landslide and other geohazards monitoring, modeling, forecasting and assessment, using various techniques in geomorphology, geophysics, remote sensing, machine learning, and big data analysis.

Chung-Ray Chu received the master's degree in undersea technology from National Sun Yat-sen University, Kaohsiung, Taiwan, in 2003, and the Ph.D. degree in hydraulic and ocean engineering from National Cheng Kung University, Tainan, Taiwan, in 2013.

He is currently working as an Assistant Researcher with the Slopeland and Hydrology Division, National Science and Technology Center for Disaster Reduction (NCDR), New Taipei, Taiwan. His research interests focus on the application of various remote-sensing techniques in landslide and geo-hazards monitoring.

Guan-Wei Lin received the master's and Ph.D. degrees in geosciences from National Taiwan University, Taipei, Taiwan, in 2005 and 2010, respectively.

He is currently an Associate Professor with the Department of Earth Sciences, National Cheng Kung University, Tainan, Taiwan. His research focuses on engineering geology, surface geological processes, geotechnical engineering, and slope hazards.