



# A Deep Active Learning Approach to the Automatic Classification of Volcano-Seismic Events

Grace F. Manley<sup>1\*</sup>, Tamsin A. Mather<sup>1</sup>, David M. Pyle<sup>1</sup>, David A. Clifton<sup>2</sup>, Mel Rodgers<sup>3</sup>, Glenn Thompson<sup>3</sup> and John Makario Londoño<sup>4</sup>

<sup>1</sup>Department of Earth Sciences, University of Oxford, Oxford, United Kingdom, <sup>2</sup>Department of Engineering Science, University of Oxford, Oxford, United Kingdom, <sup>3</sup>School of Geosciences, University of South Florida, Tampa, FL, United States, <sup>4</sup>Servicio Geológico Colombiano, Observatorio Vulcanológico y Sismológico de Manizales, Bogotá, Colombia

## OPEN ACCESS

### Edited by:

Derek Keir,  
University of Southampton,  
United Kingdom

### Reviewed by:

Tim Greenfield,  
University of Cambridge,  
United Kingdom  
Finnigan Illsley-Kemp,  
Victoria University of Wellington,  
New Zealand  
Leighton Watson,  
University of Oregon, United States

### \*Correspondence:

Grace F. Manley  
grace.manley@earth.ox.ac.uk

### Specialty section:

This article was submitted to  
Volcanology,  
a section of the journal  
Frontiers in Earth Science

**Received:** 02 November 2021

**Accepted:** 13 January 2022

**Published:** 15 February 2022

### Citation:

Manley GF, Mather TA, Pyle DM, Clifton DA, Rodgers M, Thompson G and Londoño JM (2022) A Deep Active Learning Approach to the Automatic Classification of Volcano-Seismic Events.  
*Front. Earth Sci.* 10:807926.  
doi: 10.3389/feart.2022.807926

Volcano-seismic event classification represents a fundamental component of volcanic monitoring. Recent advances in techniques for the automatic classification of volcano-seismic events using supervised deep learning models achieve high accuracy. However, these deep learning models require a large, labelled training dataset to successfully train a generalisable model. We develop an approach to volcano-seismic event classification making use of active learning, where a machine learning model actively selects the training data which it learns from. We apply a diversity-based active learning approach, which works by selecting new training points which are most dissimilar from points already in the model according to a distance-based calculation applied to the model features. We combine the active learning with an existing volcano-seismic event classifier and apply the model to data from two volcanoes: Nevado del Ruiz, Colombia and Llaima, Chile. We find that models with data selected using an active learning approach achieve better testing accuracy and AUC (Area Under the Receiver Operating Characteristic Curve) than models with data selected using random sampling. Additionally, active learning decreases the labelling burden for the Nevado del Ruiz dataset but offers no increase in performance for the Llaima dataset. To explain these results, we visualise the features from the two datasets and suggest that active learning can reduce the quantity of labelled data required for less separable data, such as the Nevado del Ruiz dataset. This study represents the first evaluation of an active learning approach in volcano-seismology.

**Keywords:** volcanology, volcano-seismic events, active learning, deep learning, CNN

## 1 INTRODUCTION

Understanding the evolution of seismic activity prior to and during eruptions is critical for understanding transitions in volcanic state (e.g., Power et al., 1994). Variations in the type of volcano seismicity can reflect the underlying source processes associated with magmatic or hydrothermal transport, and stress changes (Chouet and Matoza, 2013; McNutt and Roman 2015). Timely characterisation of the type of volcano seismicity (e.g., event types outlined in **Table 1**) is therefore imperative for assessing evolving volcanic hazards, a task typically performed by analysts in volcano observatories. Hence, the automatic detection and classification of volcano-seismic event types would be valuable for reducing the workload of analysts during periods of heightened volcanic activity (Scarpetta et al., 2005).

**TABLE 1** | Description of the volcano-seismic event types contained in the datasets used for this study.

Event type	Typical frequency range	Selected potential source mechanisms	Dataset with these events included	References for potential source mechanisms
VT	5–15 Hz	Brittle rock failure	Nevado del Ruiz, Llaima	Roman and Cashman (2006), Roman and Power, (2011), Lahr et al. (1994)
LP	0.5–5 Hz	Resonance of fluid-filled crack or conduit; magma-hydrothermal interactions; magmatic degassing; brittle melt fracture	Nevado del Ruiz, Llaima	Chouet, (1996), Chouet et al. (1994), Lahr et al. (1994), Neuberg et al. (2000), Thomas and Neuberg (2012), Bean et al. (2014)
Hybrid	High-frequency onset (>5 Hz) with low-frequency decay (0.5–5 Hz)	Brittle triggering event followed by conduit resonance	Nevado del Ruiz	Lahr et al. (1994), Neuberg et al. (2000), Neuberg et al. (2006), Harrington and Brodsky, (2007), Rodgers et al. (2016)
Tremor	1–12 Hz	Extended magma movement, explosions	Llaima	Julian, (1994), Baptie et al. (2002), Neuberg et al. (2000), Hellweg, (2000), Jellinek and Bercovici, (2011)
Tectonic	Over 5 Hz	Non-volcanic brittle failure events e.g., local or regional seismic events	Llaima	Marzocchi et al. (1993), La Femina et al. (2004), Canário et al., (2020b)

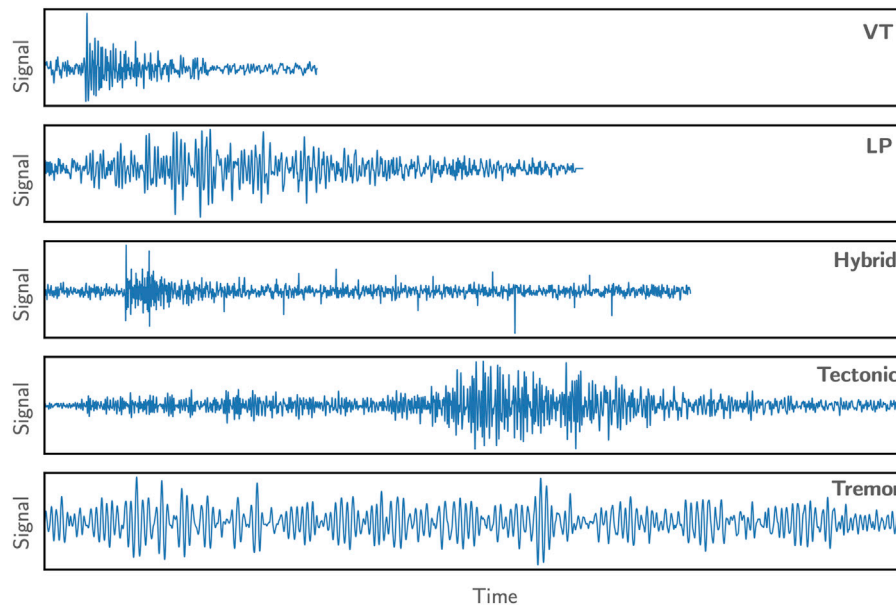
Machine learning is a powerful tool in the processing and study of volcano-seismic events and there have been advances in their automatic detection and classification using supervised machine learning approaches (e.g., Langer et al., 2006; Apolloni, 2009; Malfante et al., 2018). Recent contributions to the literature have explored the use of convolutional neural networks (CNNs) for the purposes of volcano-seismic event classification (e.g., Canário J. et al., 2020; Lara et al., 2021). Rather than acting on a set of user-designed features, CNNs operate by learning a series of filters to extract useful features from the data and output a classification. Neural networks are formulated from an input layer, a number of hidden layers which transform the data, and an output layer from which predictions are made. The resulting set of features is proposed to be more robust of intra-class variability than a set of manually chosen or hand-crafted features (Chan et al., 2015).

Supervised machine learning approaches require labelled data for training and larger amounts of data are required for deep learning models to acquire an accurate representation of the input data (Hinton et al., 2012; Krizhevsky et al., 2012). Hence, the labelling time required to curate a data volume large enough to establish an automatic classification system can be significant depending on the difficulty of the labelling task. Active learning is the process by which a machine learning model chooses the optimum data for training to achieve higher accuracy using a smaller amount of training data (Settles, 2009). Labels for these chosen training points are then assigned by an expert. Active learning is useful for contexts in which obtaining labels for unlabelled samples is difficult or time-consuming. For this reason, active learning has previously been applied in healthcare engineering with the motivation to develop tools for automated classification of health-related data, whilst minimising the time burden placed upon healthcare professionals to label large quantities of examples for training (e.g., Smailagic et al., 2018; Kiyasseh et al., 2020). This motivation

is shared in the volcano monitoring setting, where the manual classification of events requires expert knowledge. During periods of heightened volcanic activity, the time demands upon observatory staff will increase, thus minimising the time spent on manual classification for establishing an automatic classifier is advantageous (Scarpetta et al., 2005).

We use an approach to active learning which has been primarily applied to choose an optimum training dataset from a large pool of unlabelled data. This problem is important for volcano-seismology, where campaign-based monitoring may mean that a glut of data is received at once. For volcanoes which have had a long period of repose prior to eruption, the seismic network may be sparse or non-existent (Moran et al., 2008). During times of volcanic crisis, new seismic stations may be installed within a seismic network, thus there may be no backlog of previous data to train an automatic classifier upon. Alternatively, digitisation of legacy data may also generate a large volume of unlabelled data for which there is a time-burden for digitising and verifying the associated labels (Thompson et al., 2020).

Previous machine learning techniques in volcano seismology have used transfer learning to make the most of limited labelled training data. Transfer learning is the process of using knowledge (e.g., model structure or weights) from a previous classification task to obtain a model for a new classification task in which limited or no training data is available (Pan and Yang, 2009). Transfer learning schemes are generally more successful where the size of the training dataset is small in comparison to the previous classification task. Lapins et al. (2021) apply transfer learning techniques for detecting phase arrivals at Nabro Volcano, Eritrea and show that a model with feature extraction layers based on an established seismic phase picking model can be used to label seismic phases where only a limited training set is available. Bueno et al. (2020) incorporate transfer learning and uncertainty-based active learning to create a



**FIGURE 1** | Representative examples of waveforms of each class from the datasets used in this study, where VT represents a volcano-tectonic event and LP represents a long-period event. The VT, LP, tectonic and tremor examples are from the Llaima dataset, and the hybrid example is from the Nevado del Ruiz dataset. Each waveform has an amplitude normalised to between 0 and 1 and the range of the x-axis in each panel is 64s (which corresponds to the length of the input waveform to the model).

program for volcano-seismic analysis and classification via a model pre-trained on events from both Mount St Helens and Bezymianny volcanoes. However, no analysis of the potential for active learning to reduce the size of the training set in the field of volcano-seismic event classification has yet been done.

In this paper, we present an active learning approach (after Sener and Savarese, 2017) to select optimal training data. We then combine this active learning approach with an existing volcano-seismic event classification network SeismicNet (Canário J. et al., 2020). We apply this method to two datasets of detected volcano-seismic events which are fully labelled with analyst-reviewed classifications (these event types are outlined in **Table 1** and reasons for selecting these datasets are discussed in **Section 2.1**):

- (1) 5614 detected seismic events from the 2012 unrest and eruption of Nevado del Ruiz Volcano, Colombia. These events are labelled as VT, LP and Hybrid events by analysts from the Servicio Geológico Colombiano (SGC).
- (2) 3592 detected seismic events from Llaima Volcano, Chile over the 2010–2016 non-eruptive period (Canário JP. et al., 2020). These events are labelled as VT, LP, Tremor and Tectonic events by analysts from the Observatorio Volcanológico de los Andes del Sur (OVDAS). Canário et al., 2020a used this dataset to train and test neural network architectures for automatic event classification.

We evaluate the success of the active learning approach in selecting the training dataset against an equivalent model trained with randomly selected training points. We withhold the labels from the training data until points are selected, where we add the

points and associated labels to the training dataset. After training the model, we test the model performance on unseen data.

## 2 METHODS

### 2.1 Data

For this experiment we use two seismic datasets which are annotated with analyst-reviewed labels. These labels are withheld from the model during the active learning experiment (outlined in **Section 2.3**) until a point is selected to be included in the training dataset (through either active or random sampling), and the process of labelling the waveforms is simulated by providing the true labels to the model when they are selected.

**Table 1** summarises the frequency characteristics and potential source processes of the event types which are contained in the two datasets used for this study and representative examples of these event types are plotted in **Figure 1**. Volcano-tectonic (VT) events are characterised by impulsive onsets and dominant frequencies above 5 Hz (Lahr et al., 1994). They are thought to be brittle failure events that can be triggered by mechanisms such as stress change from magma intrusion (Roman and Cashman, 2006), or an increase in pore fluid pressure due to volatile transport (Huppert and Sparks, 2016). Long Period (LP) events are characterised by an emergent onset followed by a set of decaying harmonic oscillations, and are typically concentrated within the frequency range of 0.5–5 Hz (Lahr et al., 1994). The source mechanism of LP events can be attributed to resonance of a fluid-filled crack or conduit (Chouet

1996; Neuberg 2000). These signals are also termed Low Frequency (LF) events and can reflect a broader set of potential source mechanisms including slow-rupture earthquakes in unconsolidated edifice (Bean et al., 2014), magma-hydrothermal interaction (Jolly et al., 2017) and magmatic degassing (Rodgers et al., 2015a). Throughout this manuscript we use the term LP as this is the term assigned within the labelling schemes of the datasets used here. Hybrid events (HB) contain a mixture of both VT and LP signals, typically with a high frequency impulsive onset similar to a VT event and a decay which is prolonged similar to an LP event (Lahr et al., 1994). Potential source mechanisms for hybrid events include a brittle-failure triggering event which induces conduit resonance (e.g., Neuberg et al., 2006). Harrington and Brodsky (2007) suggest that hybrids can be attributed to brittle failure events with low rupture velocity which have undergone complex path effects. Rodgers et al. (2016) identify families of repeating events that transition from hybrids into LPs, suggesting similar source processes with subtle changes over time in the triggering process. Tremor events (TR) are a continuous low-frequency signal, usually associated with fluid migration. Suggested source mechanisms for tremor include sustained triggering of LP events in magma conduits (Powell and Neuberg 2003), restricted or choked flow of fluids in the sub-surface (Chouet 1996; Hellweg 2000; Lesage et al., 2006) or magma degassing or fissure eruptions at the surface (Battaglia and Aki 2003; Nadeau et al., 2011). In the Llaima dataset, the tectonic event type (TC) includes events with a non-volcanic origin, such as regional or local earthquakes (Canário JP. et al., 2020). Source mechanisms of tectonic events are indistinguishable from VTs, i.e., brittle shear failure due to stress changes, but these events can be distinguished from VTs as being located outside of the volcanic network. Cataloguing tectonic events is important for volcano monitoring as interactions between volcanoes and regional tectonic activity have been documented in many places (Marzocchi et al., 1993; Linde and Sacks, 1998; La Femina et al., 2004), and distal VTs may in some cases be considered tectonic events and excluded from volcano-seismic catalogues (Roman et al., 2008).

For this study, we trial the methods on two datasets from Nevado del Ruiz, Colombia and Llaima Volcano, Chile. We choose these datasets for trialling the active learning approach for two reasons. Firstly, the datasets each contain a different number of classes: the Llaima dataset contains 4 event types (VT, LP, tremor and tectonic events) and the Nevado del Ruiz dataset contains 3 event types (VT, LP and hybrid events). Hence, we can test whether the active learning approach is more beneficial where a greater number of event types are being classified. Secondly, the datasets are from different types of volcanic activity: the Llaima dataset is from 2010–2016 which is a non-eruptive period, and the Nevado del Ruiz dataset is from 2012, which covers precursory unrest and eruption from 22nd February 2012 onwards. This distinction is important for trialling an active learning approach on seismic data from both non-eruptive activity and heightened unrest as the spectral contents of seismic events can vary during unrest and eruption (Rodgers et al., 2016). For example, Buurman and West (2010) observe an

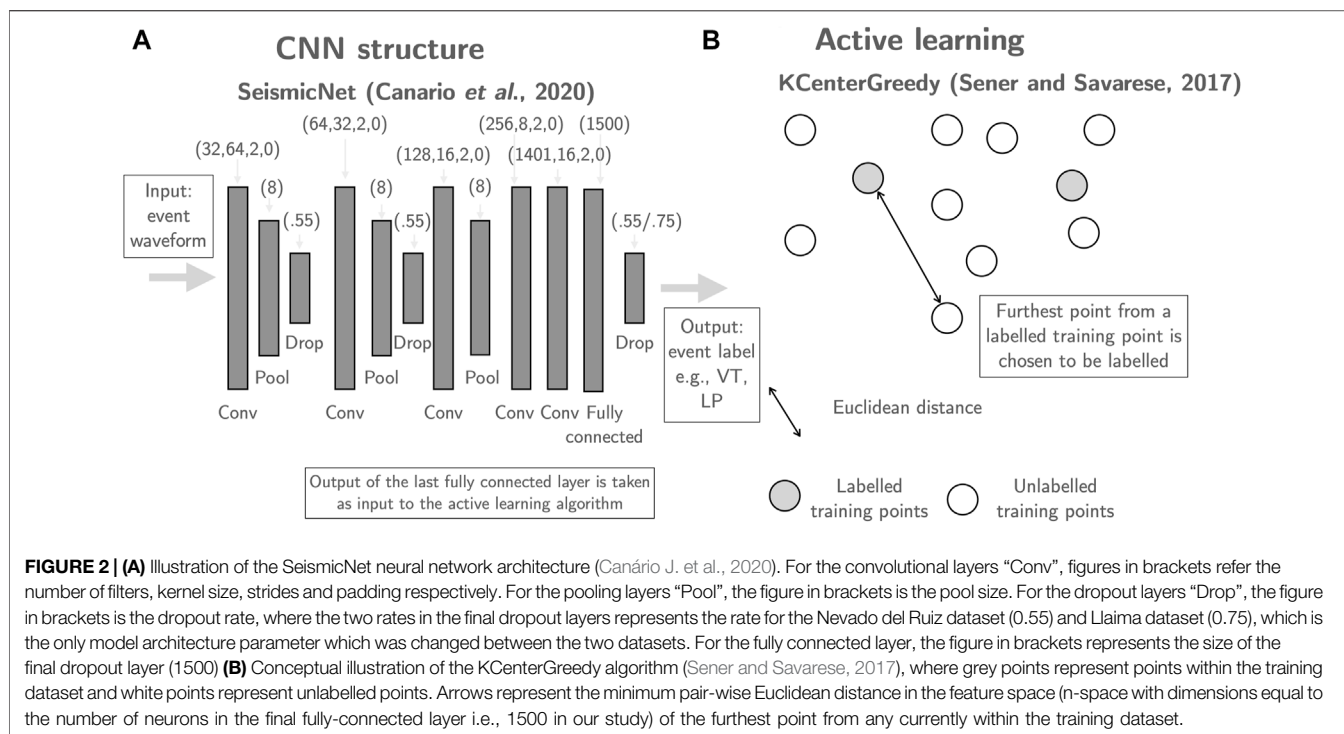
increase in events dominated by low frequency energy prior to explosions during the 2006 eruption of Augustine Volcano. This distinction allows us to test whether the active learning approach performs better on events with greater variability.

### 2.1.1 Nevado del Ruiz Dataset

We use data from the 2012 eruptive period of Nevado del Ruiz Volcano, Colombia, an ice-covered stratovolcano located within the Andes (Londono, 2016). The 2012 eruption of Nevado del Ruiz represented the first emission of ash since the 1985 eruption, which caused over 25,000 fatalities when lahars inundated surrounding communities, including the city of Armero (Lowe et al., 1986; Naranjo et al., 1986). The data are from the vertical component of the seismic station TOLZ. The dataset includes 5614 waveforms of three types: 2363 VT events, 2867 LP events and 384 hybrid events. These events have been classified by analysts at the Servicio Geológico Colombiano (SGC). The waveforms have been filtered using a band pass filter to include frequencies between 0.5 and 20 Hz. Each input waveform is 64 s long, and shorter waveforms are padded with 0 values to reach this length. Seismic data from between 2005–2006 at Nevado del Ruiz have previously been used to train and test an automatic classifier using Hidden Markov Models, which achieved a classification accuracy of approximately 88% (Cárdenas-Peña et al., 2013). This dataset was significantly smaller than the dataset used here and involved an extensive data pre-processing stage to extract frequency-based features and select the features which would provide the greatest information over time. This approach contrasts with our method, in which very little pre-processing is required and no prior knowledge about waveform properties from the whole study period is used to optimise the approach. Long-term data from Nevado del Ruiz from 2007–2014 were automatically classified using a random forest classifier (Rodgers et al., 2015b). This study used the full spectrogram as input vector and using ~7% of analyst-labelled events for training achieved a classification accuracy of approximately 77%. Hand-crafted features derived from this catalogue of detected events have previously been used to classify eruptive and non-eruptive activity during the 2012 eruption of Nevado del Ruiz (Manley et al., 2020).

### 2.1.2 Llaima Dataset

We use the dataset presented by Canário et al. (2020b) which encompasses 4 event types (VT, LP, Tremor and Tectonic) from 2010–2016 at Llaima Volcano, Chile. The data are from the vertical component of the seismic station LAV, with labels from Observatorio Volcanológico de los Andes del Sur (OVDAS). The total dataset size is 3592 waveforms, which comprises 304 VT events, 1310 LP events, 490 tremor events and 1488 tectonic events (e.g., regional events). Therefore, the Llaima dataset is approximately 36% smaller than the Nevado del Ruiz dataset. As described in Canário et al. (2020b) the waveforms have been filtered using a band-pass filter to include frequencies between 1 and 10 Hz and subsequently normalised by the maximum value. Each input waveform is 64 s long, and shorter waveforms are padded with 0 values to reach this length. This dataset was previously used for studying



**FIGURE 2 | (A)** Illustration of the SeismicNet neural network architecture (Canário J. et al., 2020). For the convolutional layers “Conv”, figures in brackets refer the number of filters, kernel size, strides and padding respectively. For the pooling layers “Pool”, the figure in brackets is the pool size. For the dropout layers “Drop”, the figure in brackets is the dropout rate, where the two rates in the final dropout layers represents the rate for the Nevado del Ruiz dataset (0.55) and Llaima dataset (0.75), which is the only model architecture parameter which was changed between the two datasets. For the fully connected layer, the figure in brackets represents the size of the final dropout layer (1500) **(B)** Conceptual illustration of the KCenterGreedy algorithm (Sener and Savarese, 2017), where grey points represent points within the training dataset and white points represent unlabelled points. Arrows represent the minimum pair-wise Euclidean distance in the feature space (n-space with dimensions equal to the number of neurons in the final fully-connected layer i.e., 1500 in our study) of the furthest point from any currently within the training dataset.

the performance of a range of neural network architectures by Canário et al., 2020a and achieved validation accuracies of over 90% using the SeismicNet, though the models were not tested on unseen data in their study.

## 2.2 Machine Learning Classification

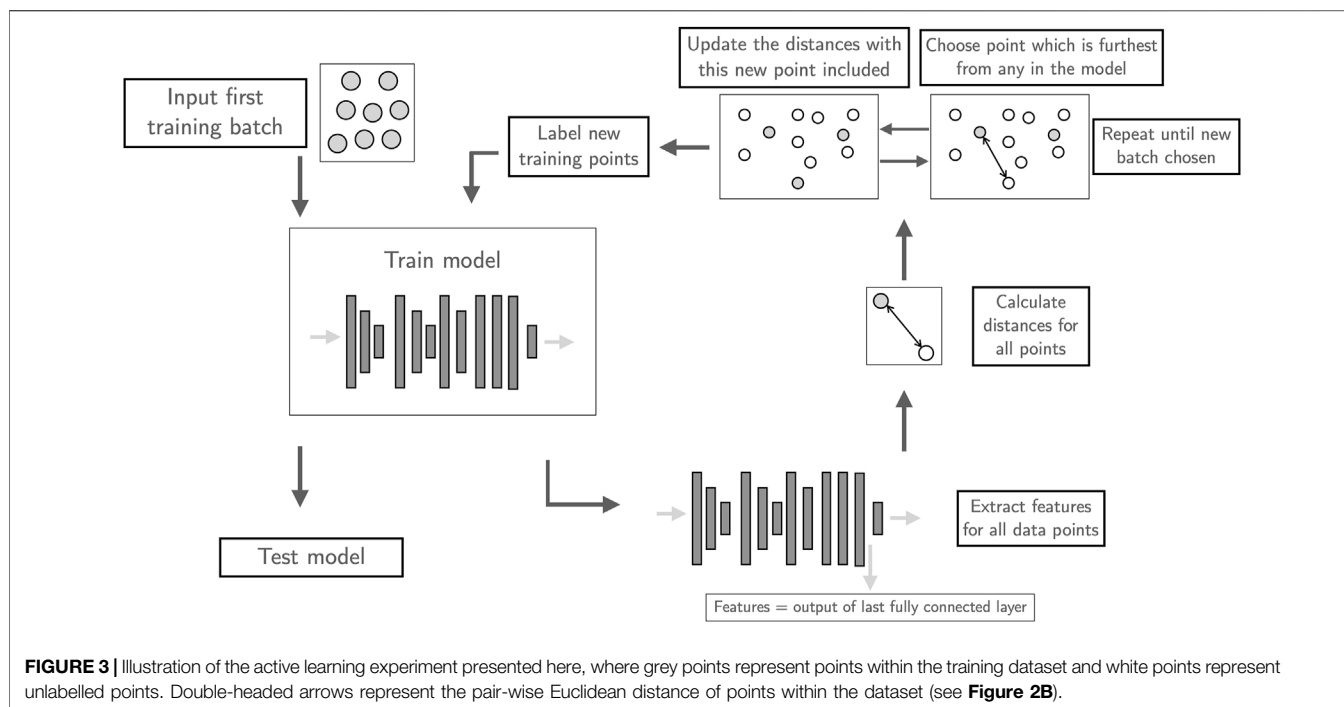
We combine the active learning selection of training data with an existing volcano-seismic event classifier. We use an existing CNN-based volcano-seismic event classifier, known as SeismicNet, proposed by Canário et al. (2020a), illustrated in Figure 2A. The SeismicNet classifier is based on a similar architecture designed for classifying waveforms from audio, known as SoundNet (Aytar et al., 2016). The architecture takes the raw waveforms as input rather than derived information such as a spectrogram representation or wavelet transform. We choose this classifier as it achieved comparable results to other CNN-based approaches (e.g., Curilem et al., 2018; Lara et al., 2021) but required significantly fewer pre-processing steps, which is concordant with the aim of this study to develop a method for the prompt labelling of a pool of unlabelled events.

This classifier is a Convolutional Neural Network (CNN), a neural network which contains a number of layers which execute convolutions on the input data and downsample (or downsize) the signal. These layers act as the “feature extraction” process for the CNN. Convolutional layers apply convolution over the data with a given size (known as the filter size). These convolutions are used in combination with pooling layers, which reduce the size of the data, and dropout layers which are randomly set to 0 with a given probability, and are used to increase the generalisability of a model and prevent overfitting to the training data (Srivastava et al., 2014). Finally, fully connected layers are used to compute

the final classification output of the model. We use the categorical cross entropy loss function and the stochastic gradient descent optimiser to fit the model. The models were implemented in Keras v2.4.0 with a Tensorflow backend. The network architecture is outlined in Figure 2A, with the parameters shown for each layer. Before running the experiments, we tested the model architecture, varying parameters such as dropout and hyperparameters such as the model batch size and number of epochs for training, and chose the parameters which achieved the greatest validation accuracy in training. We choose a model batch size of 16 and number of training epochs of 150 for the Nevado del Ruiz dataset and 200 for the Llaima dataset.

## 2.3 Active Learning

Diversity-based active learning works on the principle that a model will achieve better performance if more variation across the training set is sampled by the model during the training period. Because the extent of the variability in the training set has been sampled during the model training period, in theory the model will perform well on unseen testing data under the assumption that the training data captures the full diversity of the dataset. In contrast to diversity-based active learning, uncertainty-based active learning schemes involve choosing examples to be labelled based on the uncertainty of the model in the given classification, thereby prioritising examples which lie close to the boundary between classes (Jain and Kapoor, 2009). However, for deep learning classifiers, the probability output of the model is not representative of the model confidence (Gal and Ghahramani, 2016) and therefore the classifier must be adapted to a Bayesian framework to adequately quantify the uncertainty of



the classification. We therefore select a diversity-based method of active learning such that the method can be combined with previous classifiers.

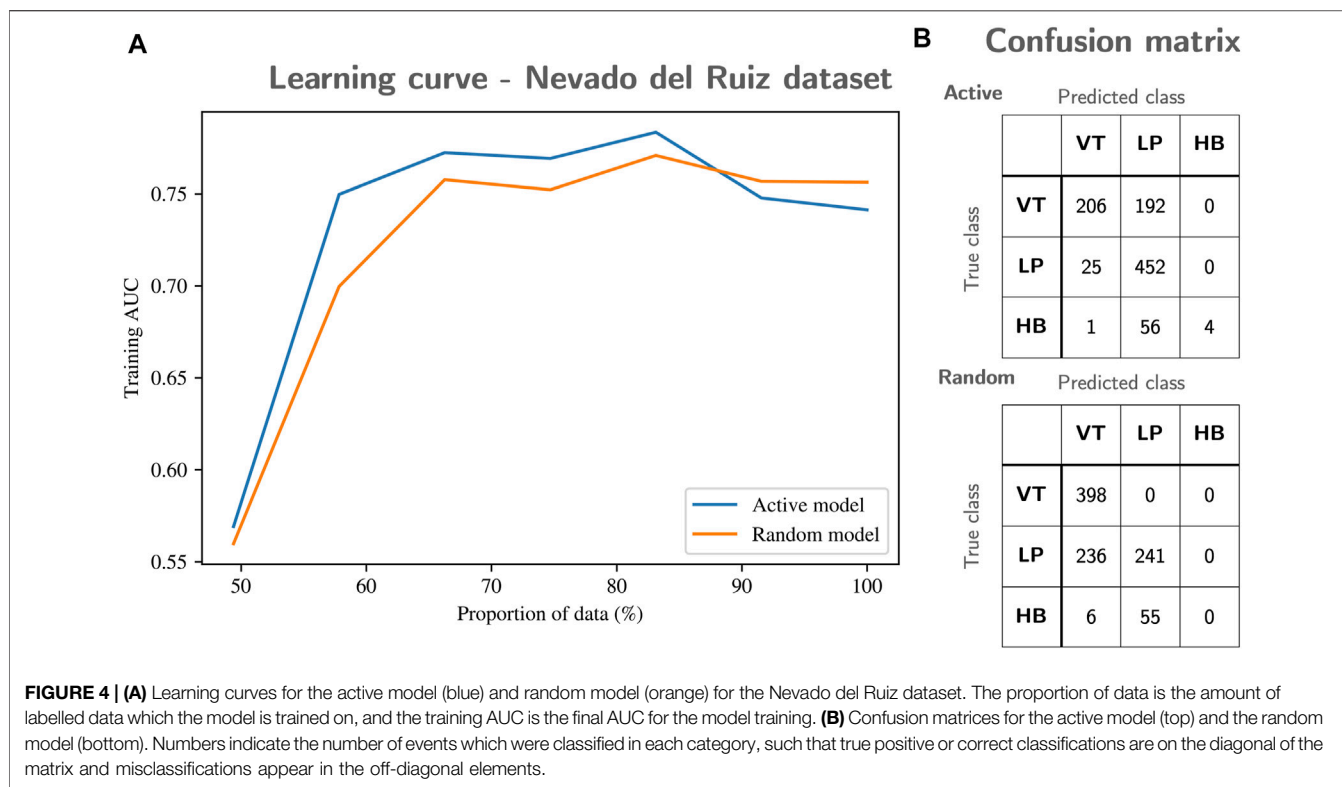
We apply a core-set selection approach to active learning after Sener and Savarese (2017), known as KCenterGreedy, illustrated in **Figure 2B**. The KCenterGreedy method works by selecting new data points which are the most dissimilar from any data point within the current training set. Deep learning models are formulated on batches of data, making labelling single data points logistically challenging, and meaning that obtaining a single label is not likely to significantly affect the model performance. The KCenterGreedy algorithm is implicitly designed to work on selecting batches of data. We choose this algorithm as it is extensible to any CNN-based classifier for volcano-seismic events.

**Figure 3** illustrates the process of the active learning experiment. The model is trained with an initial randomly-chosen batch of labelled waveforms. The full set of unlabelled data is then passed through the feature extraction layers of the network to generate the features for the algorithm. Within the feature space, the minimum pairwise Euclidean distance (a measure of feature similarity) between all datapoints to a point within the training set is then calculated. This calculation results in 0 distance for points already labelled within the training set. A new point which is the maximum distance away from any labelled point is selected and added to the set of labelled points. The minimum distances are then re-calculated with this point included (as unlabelled points located near the newly-chosen point may now be closer to a labelled point).

This process is repeated until the user-selected batch size is reached and all the newly chosen points are labelled. The user-selected batch size is the quantity of data selected in one active learning cycle as illustrated in **Figure 3**. Next, the model is retrained on all the data chosen so far and the AUC (Area Under the Curve, a measure of the ability of a classifier to distinguish between classes – see below) is re-evaluated. Once all the training data have been selected, the model is then tested on unseen data.

We split the full dataset into training, validation and testing datasets with a split of 2/3, 1/6 and 1/6 of the dataset respectively. For the models in this paper, the initial randomly-selected batch of waveforms is chosen to be approximately 50% of the total training size. We choose the active learning user-selected batch size as 216 datapoints. Experienced volcano-seismic analysts can classify in the region of hundreds of volcano-seismic events in a day (e.g., Thompson et al., 2020). Therefore, we pick a figure for selecting batches in the lower region of this estimate as a reasonable quantity to be labelled at once before the models can be re-trained with the new data included.

We compare the active learning experiment to an experiment where the next batch is randomly-selected. The performance metric we use to compare the performance of the models is the Area Under the Receiver Operating Characteristic Curve (AUC-ROC, referred to as AUC). The ROC curve is a plot of the False Positive Rate (FPR) vs the True Positive Rate (TPR) of a model at various threshold values. The AUC is computed by integrating under this curve, whereas model accuracy is computed by setting a FPR threshold and



**TABLE 2 |** Testing performance (AUC and accuracy) for the models presented in this study.

Dataset	Model type	Testing AUC	Testing accuracy (%)
Nevado del Ruiz	Actively-sampled	0.8848	70.73
	Randomly-sampled	0.8626	68.27
Llaima	Actively-sampled	0.9898	93.32
	Randomly-sampled	0.9332	94.82

computing the percentage of examples which were correctly classified. A higher AUC corresponds to a higher TPR at lower FPR and therefore better model performance. For comparing models, the AUC is a preferable metric to other metrics such as the training accuracy as the AUC is independent of a threshold which, in the case of accuracy, can vary between models.

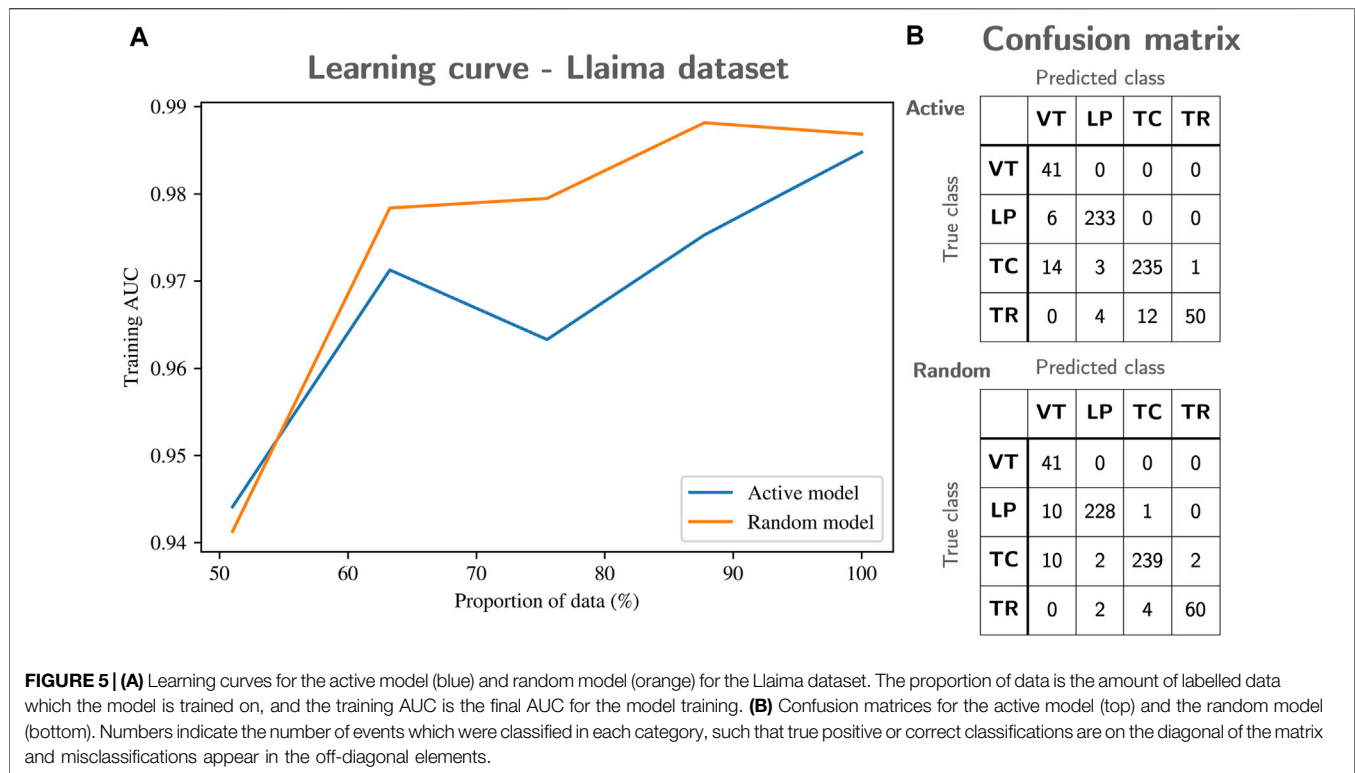
### 3 RESULTS

To assess the results of the active learning frameworks described in Section 2, we compare the model performance via the AUC against the proportion of the dataset which is being used for model training, where greater AUC corresponds to better model performance. As illustrated in Figure 3, the model is trained with a subset of labelled data points and new training points are added to the labelled set either through active or random selection. Hence, the proportion of labelled points increases over the course of the experiment. We compare the active learning selection of training data against a random selection of training data. For a

successful active learning approach, the models will achieve higher performance at a lower proportion of training data than the randomly-selected case.

#### 3.1 Nevado del Ruiz Dataset

Figure 4 illustrates the active learning and randomly-selected learning curves for models trained on the Nevado del Ruiz dataset. Table 2 indicates the testing AUC and accuracy of the model. The active learning model achieves greater training AUC at a smaller proportion of the dataset used for training with an AUC comparable to that achieved at the final training step obtained with just approximately 60% of the training datapoints labelled. The randomly-sampled model achieves greater AUC than the actively-sampled model once approximately 85% of the training data are labelled. However, the active-sampling testing AUC of 0.8848 is greater than the random-sampling testing AUC of 0.8626 when tested on the same set of unseen datapoints. The final testing accuracy of the active learning model, at 70.73%, is higher than the randomly-sampled model at 68.27%. These testing accuracies are comparable to



classification accuracies achieved on noisy seismic data (e.g., Cortés et al., 2021) but lower than those achieved for the less noisy Llama dataset.

**Figure 4** contains the confusion matrix of both models. For the actively-sampled model, the incorrect classifications are primarily the erroneous classification of VT events as LP events. For the randomly-sampled model the opposite occurs, where LP events are primarily misclassified as VT events. Both models classify the majority of hybrid events as LP events (though hybrids represent a small proportion of the dataset), with the random classifier failing to identify any hybrid events. Previous automatic classifiers applied to Nevado del Ruiz data had a high misclassification rate of hybrid events, and it was suggested that discriminating between hybrid and LP events is particularly difficult from observations made at a single recording station (Cárdenas-Peña et al., 2013). Additionally, the a-priori classification of hybrid events may be subject to human analyst bias (Langer et al., 2006).

### 3.2 Llama Dataset

**Figure 5** illustrates the active learning and randomly-selected learning curves for models trained on the Llama dataset. **Table 2** indicates the testing results of the models. The testing AUC of the active learning model, at 0.9898, is better than the testing AUC of the randomly-sampled model, at 0.9332. The testing accuracy is similar to the training and validation accuracy reported by Canário et al., 2020a, at 93.32% for the active learning model and 94.82% for the randomly-sampled model. The active learning approach achieves similar training AUC until approximately 60% of the

dataset is labelled, after which the randomly-selected model achieves greater AUC. The models approach a similar training AUC towards the end of model training.

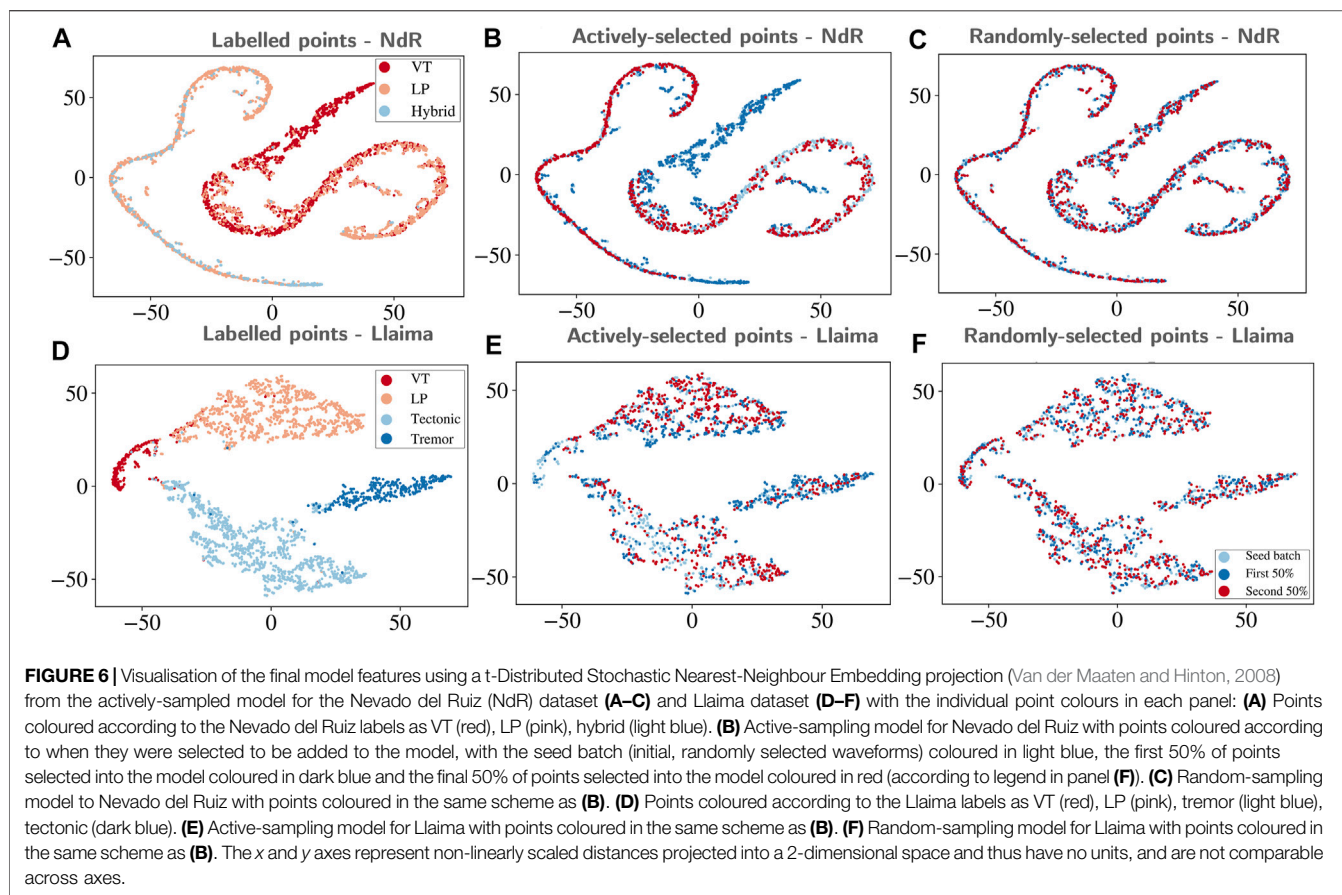
**Figure 5** includes the confusion matrix of the active learning compared to the randomly-sampled case. The correct classifications are similar between both models. For the active learning model, misclassifications are primarily tectonic events misclassified as VT events, and tremor events misclassified as tectonic events. For the randomly-sampled case, misclassifications of LP or tectonic events as VT are the main source of wrong classifications.

## 4 DISCUSSION

### 4.1 Model Comparison

As discussed in **Section 3**, the models trained with actively-sampled data achieved better training AUC than the models trained with randomly-sampled data for the Nevado del Ruiz dataset. The Nevado del Ruiz active learning model achieved similar training AUCs at 55% of the data labelled to 100% of the data labelled, indicating the potential of an active learning approach to reduce the amount of labelling required for training a model. However, the improvement for using the active sampling was significantly smaller for the Llama dataset. To contextualise these results, we can investigate the structure of the two datasets. **Figure 6** is a 2-dimensional projection of the features from the final model using t-distributed Stochastic Nearest Neighbour Embedding (t-SNE), with individual features coloured by the order in





which they are selected into the model. The t-SNE is a probabilistic machine learning algorithm which is designed to preserve the high-dimensional distance when the data are projected into two dimensions, such that points which are near in high-dimensional space remain near in the resulting two-dimensional representation (Van der Maaten and Hinton, 2008). This projection is achieved by minimising the Kullback-Leibler divergence of a high-dimensional joint probability and a low-dimensional joint probability that a point would pick another as its neighbour.

**Figures 6A,D** indicate some separability of the features based on the true label for the Nevado del Ruiz dataset, though the LP events show some overlap with both VT and hybrid events. There is excellent separability of the features based on the true label for the Llaima dataset. For the Nevado del Ruiz data, the hybrid events plot in a similar space to the LP events when projected which could explain the misclassification of hybrids as LP events which was discussed in **Section 3.1**. **Figures 6C,F** represent the point selection for the random models and we observe that the colours are evenly distributed across the space, which is consistent with the selection method as points are randomly sampled.

**Figures 6B,E** represent the point selection for the active learning models. For the Nevado del Ruiz active learning model (**Figure 6B**), a greater concentration of dark blue points (indicating earlier selection into the training set) is

observed in the primarily-hybrid and primarily-VT clusters. This observation indicates that a greater proportion of characteristic events of those classes are included at an earlier stage in the experiment, which could explain the better performance of the active learning model with less training data. For the Llaima active learning model (**Figure 6E**), the dark blue points (indicating earlier selection of the point into the training set) are primarily in tremor events and at the edges of the LP and tectonic clusters, with more red points concentrated in the middle of the LP and tremor clusters. This observation could explain the relatively poor performance of the active sampling on the Llaima dataset early in the active learning experiment, as the active sampler is primarily picking outliers of the event types and tectonic events, which encompass a range of non-volcanic source processes. These observations indicate that a diversity-based active learning approach could be more beneficial for datasets where the event types are less separable, such as the Nevado del Ruiz dataset, as the active selection more efficiently samples the characteristic events in each class than random sampling.

## 4.2 Applications for Active Learning

The active learning approach, such as that presented here, has practical applications for volcano monitoring and volcano seismology research. Seismic monitoring is one of the most common and earliest-employed tools to be deployed as part of a monitoring network (e.g., Tilling, 1989), and large seismic

catalogues are created for most well-monitored volcanoes (Newhall et al., 2017). Digitisation and curation of legacy seismic data can yield a large quantity of previously-unlabelled or analysed data, or inconsistently labelled data due to changes in monitoring network, observatory staff and/or procedures (Thompson et al., 2020). An active learning approach could be used in these contexts to accelerate the process of event classification to provide retrospective insights on the characteristics of seismicity prior to eruptions. Alternatively, active learning can be utilised at awakening volcanoes where limited (or no) prior seismic catalogue is available, e.g., the re-awakening of Chaitén Volcano, Chile in 2008 (Lara, 2009). In such cases, an active learning approach running in real-time would minimise the burden on observatory analysts and speed up event classification tasks, which are both crucial aspects of volcano crisis management.

The active learning framework presented here has the advantage of not being intrinsically linked to any specific classification scheme. Other methods, such as automated classifiers trained on existing waveforms, might rely on a consistent set of event types. However, depending on the volcanic system being studied, the set of geologically-relevant event types may vary. For example, glacially-derived seismic events, also known as “icequakes”, are not always included in classification schemes and can mimic other volcanic event types (Thelen et al., 2013). Though icequakes are not always included in analyst-reviewed catalogues, identifying and monitoring the occurrence of these events is key for hazard assessment as debris can be mobilised by water from glaciers (Carrasco-Nunez et al., 1993). Rockfall events can be linked to changes in volcanic activity. DeRoin and McNutt (2012) analyse the occurrence of rockfall events at Augustine Volcano, United States and found a greater occurrence of rockfall events preceding the 2005 eruption, which they attribute to higher ground instability due to increased steaming at the summit. Timely identification of diverse events such as icequakes or rockfalls is important so that they may either be removed from the seismic record or analysed for their relevance to volcanic activity. Unlike using an existing classifier trained on data from other volcanic systems, an active learning approach to classification can be easily expanded to accommodate more event types based on the geological setting and unique properties of the volcanic environment.

### 4.3 Limitations

The active learning strategies presented here assume that the labels provided by the human expert (also termed an “oracle” in the field of active learning) are correct. This assumption may not be perfect. Seismic waveforms can be modified depending on external effects which may include path effects, such as the soil/bedrock characteristics, or the station geometries. Indeed, Cortés et al. (2021) suggest that the agreement between seismic analysts is approximately 80%. Incorrect classifications are a source of noise, and Linville et al. (2019) show that for a tectonic seismic catalogue 70% of the machine learning classifier error may be attributed to analyst error during event labelling. A full analysis of

the effect of oracle noise in our data is beyond the scope of this study and, for the analysis presented here, we do not distinguish error due to analyst mislabelling and error due to misclassification by the machine learning model.

This study has involved the use of single-station data for classifying seismic events. In practice for many volcanoes, a seismic network comprising multiple stations is used for monitoring volcano-seismic activity. The analyst-reviewed labels used as the ground-truth labels may therefore be made with reference to multi-station information. As discussed above, path and site effects can modify waveforms and therefore inclusion of waveforms from multiple stations for event classification could mitigate this effect. The approach presented here could be extended to include waveforms from multiple stations as channels of input into the CNN. This multi-channel input approach has previously been applied to use a 3-channel waveform as input for seismic phase arrival detection (Lapins et al., 2021). Using channels from a variety of station locations in the network (for example, a combination of summit and distal stations) could mitigate against the location-specific events which include, but are not limited to, the attenuation of high-frequency seismic energy within the waveform (e.g., Clarke et al., 2021).

Additionally, we have not considered any change in event classes during this study, using events which are from a limited time range. Cárdenas-Peña et al. (2013) used Hidden Markov Models to classify seismic activity at Nevado del Ruiz data and select features by choosing the set of features which contained the most time-varying information. We do not make any adjustments to our model for the time-dependent variation of seismicity. We suggest that, if models are trained over a longer period than presented here (e.g., 10s of years), examples are selected from the full extent of the period covered by the training dataset to account for any variability over time.

## 5 CONCLUSION

We present an application of active learning to volcano-seismic event classification. We apply a method of active learning designed to select an optimum training dataset from a large pool of unlabelled training data. We combine this active learning with a CNN-based volcano-seismic event classifier and apply the model to two volcano-seismic datasets from Nevado del Ruiz, Colombia and Llaima, Chile. For both datasets, the active learning models achieve greater model performance when tested on unseen data than models trained using randomly-selected data. The active learning approach to data selection achieved greater training AUC at a smaller fraction of data for the Nevado del Ruiz dataset; however, for the Llaima dataset the gain from using an active learning approach was significantly smaller. To contextualise these results, we compare the separability of features for both dataset and suggest that the diversity-based active learning approach presented here can provide particular utility in accelerating model performance for datasets where the event types are less separable, as at Nevado del Ruiz. We discuss the

potential applications of an active learning approach, which include the analysis of large amounts of unlabelled data from legacy datasets, rapid event classification at awakening volcanoes with no prior catalogue, or where volcano-specific event types (such as rockfall events) are widespread to the random selection of training data.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The Llaima dataset used in this paper are available from Canario et al. (2020b). Nevado del Ruiz seismic data are curated by SGC and are available on request (<https://www.sgc.gov.co>). Requests to access these datasets should be directed to <https://www.sgc.gov.co>.

## AUTHOR CONTRIBUTIONS

GM performed the analysis in this chapter with supervision from TM, DP, DC, and MR and wrote the manuscript. JL provided

access to the Nevado del Ruiz catalogue. All authors contributed towards revision of the manuscript.

## FUNDING

This research was facilitated by a Natural Environment Research Council (NERC) Studentship (NE/L002612/1). DC was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). DP and TM are supported by the NERC Centre for Observation and Modelling of Earthquakes, Volcanoes and Tectonics (COMET).

## ACKNOWLEDGMENTS

Nevado del Ruiz data was obtained from the Observatorio Vulcanológico y Sismológico de Manizales, Servicio Geológico Colombiano. Llaima data was obtained from Canário et al. (2020). Code to reproduce the figures presented in this paper is available from Manley (2021). We thank three reviewers for their comments on an earlier version of the paper.

## REFERENCES

- Apolloni, B. (2009). Support Vector Machines and MLP for Automatic Classification of Seismic Signals at Stromboli Volcano. *In IOS Press*, 204, 116.
- Aytar, Y., Vondrick, C., and Torralba, A. (2016). Soundnet: Learning Sound Representations from Unlabeled Video. *Adv. Neural Inf. Process. Syst.* 29, 892–900.
- Baptie, B., Luckett, R., and Neuberg, J. (2002). Observations of Low-Frequency Earthquakes and Volcanic Tremor at Soufrière Hills Volcano, Montserrat. *Geol. Soc. Lond. Mem.* 21 (1), 611–620. doi:10.1144/gsl.mem.2002.021.01.30
- Battaglia, J., and Aki, K. (2003). Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes. *J. Geophys. Res. Solid Earth* 108 (B8). doi:10.1029/2002jb002193
- Bean, C. J., De Barros, L., Lokmer, I., Métaixian, J.-P., O' Brien, G., and Murphy, S. (2014). Long-period Seismicity in the Shallow Volcanic Edifice Formed from Slow-Rupture Earthquakes. *Nat. Geosci.* 7 (1), 71–75. doi:10.1038/ngeo2027
- Bueno, A., Zuccarello, L., Díaz-Moreno, A., Woollam, J., Titos, M., Benítez, C., et al. (2020). PICOSS: Python Interface for the Classification of Seismic Signals. *Comput. Geosciences* 142, 104531. doi:10.1016/j.cageo.2020.104531
- Buurman, H., and West, M. E. (2010). "Seismic Precursors to Volcanic Explosions During the 2006 Eruption of Augustine Volcano," in *The 2006 eruption of Augustine Volcano*, Editor J. A. Power, M. L. Coombs, and J. T. Freymueller (Alaska: U.S. Geological Survey Professional), Chap. 2, 1769, 41–57. Available at: [https://pubs.usgs.gov/pp/1769/chapters/p1769\\_chapter02.pdf](https://pubs.usgs.gov/pp/1769/chapters/p1769_chapter02.pdf).
- Canário, J. P., de Mello, R. F., Curilem, M., Huenupan, F., and Rios, R. A. (2020b). Llaima Volcano Dataset: In-Depth Comparison of Deep Artificial Neural Network Architectures on Seismic Events Classification. *Data Brief* 30, 105627. doi:10.1016/j.dib.2020.105627
- Canário, J. P., Mello, R., Curilem, M., Huenupan, F., and Rios, R. (2020a). In-depth Comparison of Deep Artificial Neural Network Architectures on Seismic Events Classification. *J. Volcanology Geothermal Res.* 401, 106881. doi:10.1016/j.jvolgeores.2020.106881
- Cárdenas-Peña, D., Orozco-Alzate, M., and Castellanos-Dominguez, G. (2013). Selection of Time-Variant Features for Earthquake Classification at the Nevado-Del-Ruiz Volcano. *Comput. geosciences* 51, 293–304. doi:10.1016/j.cageo.2012.08.012
- Carrasco-Núñez, G., Vallance, J. W., and Rose, W. I. (1993). A Voluminous Avalanche-Induced Lahar from Citlaltépetl Volcano, Mexico: Implications for hazard Assessment. *J. Volcanology Geothermal Res.* 59 (1-2), 35–46. doi:10.1016/0377-0273(93)90076-4
- Chan, T. H., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y. (2015). PCANet: A Simple Deep Learning Baseline for Image Classification?. *IEEE Transactions on Image Processing* 24 (12), 5017–5032.
- Chouet, B. A. (1996). Long-period Volcano Seismicity: its Source and Use in Eruption Forecasting. *Nature* 380 (6572), 309–316. doi:10.1038/380309a0
- Chouet, B. A., and Matoza, R. S. (2013). A Multi-Decadal View of Seismic Methods for Detecting Precursors of Magma Movement and Eruption. *J. Volcanology Geothermal Res.* 252, 108–175. doi:10.1016/j.jvolgeores.2012.11.013
- Chouet, B. A., Page, R. A., Stephens, C. D., Lahr, J. C., and Power, J. A. (1994). Precursory Swarms of Long-Period Events at Redoubt Volcano (1989–1990), Alaska: Their Origin and Use as a Forecasting Tool. *J. Volcanology Geothermal Res.* 62 (1-4), 95–135. doi:10.1016/0377-0273(94)90030-2
- Clarke, J., Adam, L., and van Wijk, K. (2021). LP or VT Signals? How Intrinsic Attenuation Influences Volcano Seismic Signatures Constrained by Whakaari Volcano Parameters. *J. Volcanology Geothermal Res.* 418, 107337. doi:10.1016/j.jvolgeores.2021.107337
- Cortés, G., Carniel, R., Lesage, P., Mendoza, M., and Della Lucia, I. (2021). Practical Volcano-independent Recognition of Seismic Events: VULCAN. Ears Project. *Front. Earth Sci.* 8, 702. doi:10.3389/feart.2020.616676
- Curilem, M., de Mello, R. F., Huenupan, F., San Martin, C., Franco, L., Hernández, E., et al. (2018). Discriminating Seismic Events of the Llaima Volcano (Chile) Based on Spectrogram Cross-Correlations. *J. Volcanology Geothermal Res.* 367, 63–78. doi:10.1016/j.jvolgeores.2018.10.023
- DeRoin, N., and McNutt, S. R. (2012). Rockfalls at Augustine Volcano, Alaska: The Influence of Eruption Precursors and Seasonal Factors on Occurrence Patterns 1997–2009. *J. Volcanology Geothermal Res.* 211–212, 61–75. doi:10.1016/j.jvolgeores.2011.11.003
- Gal, Y., and Ghahramani, Z. (2016). "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of The 33rd International Conference on Machine Learning Held in New York* (New York, USA: PMLR), 1050–1059.
- Harrington, R. M., and Brodsky, E. E. (2007). Volcanic Hybrid Earthquakes that Are Brittle-failure Events. *Geophys. Res. Lett.* 34 (6). doi:10.1029/2006gl028714
- Hellweg, M. (2000). Physical Models for the Source of Lascar's Harmonic Tremor. *J. Volcanology Geothermal Res.* 101 (1-2), 183–198. doi:10.1016/s0377-0273(00)00163-3

- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). *Improving Neural Networks by Preventing Co-adaptation of Feature Detectors*. arXiv preprint arXiv:1207.0580.
- Huppert, H. E., and Sparks, R. S. (2016). Compressible Vapour Flow in Conduits and Fractures. *J. Fluid Mech.* 802, 750–759. doi:10.1017/jfm.2016.298
- Jain, P., and Kapoor, A. (2009). “Active Learning for Large Multi-Class Problems,” in *Proceeding of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 20–25 June 2009 (IEEE), 762–769. doi:10.1109/cvpr.2009.5206651
- Jellinek, A. M., and Bercovici, D. (2011). Seismic Tremors and Magma Wagging during Explosive Volcanism. *Nature* 470 (7335), 522–525. doi:10.1038/nature09828
- Jolly, A. D., Lokmer, I., Thun, J., Salichon, J., Fry, B., and Chardot, L. (2017). Insights into Fluid Transport Mechanisms at White Island from Analysis of Coupled Very Long-Period (VLP), Long-Period (LP) and High-Frequency (HF) Earthquakes. *J. Volcanology Geothermal Res.* 343, 75–94. doi:10.1016/j.jvolgeores.2017.06.006
- Julian, B. R. (1994). Volcanic Tremor: Nonlinear Excitation by Fluid Flow. *J. Geophys. Res.* 99 (B6), 11859–11877. doi:10.1029/93jb03129
- Kiyasseh, D., Zhu, T., and Clifton, D. A. (2020). *SoCal: Selective Oracle Questioning for Consistency-Based Active Learning of Cardiac Signals*. arXiv preprint arXiv:2004.09557.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.
- La Femina, P. C., Connor, C. B., Hill, B. E., Strauch, W., and Saballos, J. A. (2004). Magma-tectonic Interactions in Nicaragua: the 1999 Seismic Swarm and Eruption of Cerro Negro Volcano. *J. Volcanology Geothermal Res.* 137 (1-3), 187–199. doi:10.1016/j.jvolgeores.2004.05.006
- Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A., and Page, R. A. (1994). Earthquake Classification, Location, and Error Analysis in a Volcanic Environment: Implications for the Magmatic System of the 1989–1990 Eruptions at Redoubt Volcano, Alaska. *J. Volcanology Geothermal Res.* 62 (1-4), 137–151. doi:10.1016/0377-0273(94)90031-0
- Langer, H., Falsaperla, S., Powell, T., and Thompson, G. (2006). Automatic Classification and A-Posteriori Analysis of Seismic Event Identification at Soufrière Hills Volcano, Montserrat. *J. volcanology geothermal Res.* 153 (1-2), 1–10. doi:10.1016/j.jvolgeores.2005.08.012
- Lapins, S., Goitom, B., Kendall, J. M., Werner, M. J., Cashman, K. V., and Hammond, J. O. (2021). A Little Data Goes a Long Way: Automating Seismic Phase Arrival Picking at Nabro Volcano with Transfer Learning. *J. Geophys. Res. Solid Earth* 126 (7), e2021JB021910. doi:10.1029/2021jb021910
- Lara, F., Lara-Cueva, R., Larco, J. C., Carrera, E. V., and León, R. (2021). A Deep Learning Approach for Automatic Recognition of Seismo-Volcanic Events at the Cotopaxi Volcano. *J. Volcanology Geothermal Res.* 409, 107142. doi:10.1016/j.jvolgeores.2020.107142
- Lara, L. E. (2009). The 2008 Eruption of the Chaitén Volcano, Chile: a Preliminary Report. *Andean Geology.* 36 (1), 125–129. doi:10.4067/s0718-71062009000100009
- Lesage, P., Mora, M. M., Alvarado, G. E., Pacheco, J., and Métaixian, J. P. (2006). Complex Behavior and Source Model of the Tremor at Arenal Volcano, Costa Rica. *J. Volcanology Geothermal Res.* 157 (1-3), 49–59. doi:10.1016/j.jvolgeores.2006.03.047
- Linde, A. T., and Sacks, I. S. (1998). Triggering of Volcanic Eruptions. *Nature* 395 (6705), 888–890. doi:10.1038/27650
- Linville, L., Pankow, K., and Draelos, T. (2019). Deep Learning Models Augment Analyst Decisions for Event Discrimination. *Geophys. Res. Lett.* 46 (7), 3643–3651. doi:10.1029/2018gl081119
- Londono, J. M. (2016). Evidence of recent deep magmatic activity at Cerro Bravo-Cerro Machín volcanic complex, central Colombia. Implications for future volcanic activity at Nevado del Ruiz, Cerro Machín and other volcanoes. *J. volcanology Geothermal Res.* 324, 156–168. doi:10.1016/j.jvolgeores.2016.06.003
- Lowe, D. R., Williams, S. N., Leigh, H., Connor, C. B., Gemmill, J. B., and Stoiber, R. E. (1986). Lahars initiated by the 13 November 1985 eruption of Nevado del Ruiz, Colombia. *Nature* 324 (6092), 51–53. doi:10.1038/324051a0
- Malfante, M., Dalla Mura, M., Metaxian, J.-P., Mars, J. I., Macedo, O., and Inza, A. (2018). Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives. *IEEE Signal. Process. Mag.* 35 (2), 20–30. doi:10.1109/msp.2017.2779166
- Manley, G. (2021). A Deep Active Learning Approach to the Automatic Classification of Volcano-Seismic Events (Code Repository). *Zenodo*. doi:10.5281/zenodo.5608754
- Manley, G. F., Pyle, D. M., Mather, T. A., Rodgers, M., Clifton, D. A., Stokell, B. G., et al. (2020). Understanding the Timing of Eruption End Using a Machine Learning Approach to Classification of Seismic Time Series. *J. Volcanology Geothermal Res.* 401, 106917. doi:10.1016/j.jvolgeores.2020.106917
- Marzocchi, W., Scandone, R., and Mulargia, F. (1993). The Tectonic Setting of Mount Vesuvius and the Correlation between its Eruptions and the Earthquakes of the Southern Apennines. *J. Volcanology Geothermal Res.* 58 (1-4), 27–41. doi:10.1016/0377-0273(93)90100-6
- McNutt, S. R., and Roman, D. C. (2015). “Volcanic Seismicity,” in *The Encyclopedia of Volcanoes* (Academic Press), 1011–1034. doi:10.1016/b978-0-12-385938-9.00059-6
- Moran, S. C., Freymueller, J. T., LaHusen, R. G., McGee, K. A., Poland, M. P., Power, J. A., et al. (2008). Instrumentation Recommendations for Volcano Monitoring at US Volcanoes under the National Volcano Early Warning System. *US Geol. Surv. Scientific Invest. Rep.* 5114, 47. doi:10.3133/sir20085114
- Nadeau, P. A., Palma, J. L., and Waite, G. P. (2011). Linking Volcanic Tremor, Degassing, and Eruption Dynamics via SO<sub>2</sub> Imaging. *Geophys. Res. Lett.* 38 (1). doi:10.1029/2010gl045820
- Naranjo, J. L., Sigurdsson, H., Carey, S. N., and Fritz, W. (1986). Eruption of the Nevado del Ruiz volcano, Colombia, on 13 November 1985: tephra fall and lahars. *Science* 233 (4767), 961–963. doi:10.1126/science.233.4767.961
- Neuberg, J. (2000). Characteristics and Causes of Shallow Seismicity in Andesite Volcanoes. *Philosophical Trans. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci.* 358 (1770), 1533–1546. doi:10.1098/rsta.2000.0602
- Neuberg, J., Luckett, R., Baptie, B., and Olsen, K. (2000). Models of Tremor and Low-Frequency Earthquake Swarms on Montserrat. *J. Volcanology Geothermal Res.* 101 (1-2), 83–104. doi:10.1016/s0377-0273(00)00169-4
- Neuberg, J. W., Tuffen, H., Collier, L., Green, D., Powell, T., and Dingwell, D. (2006). The Trigger Mechanism of Low-Frequency Earthquakes on Montserrat. *J. Volcanology Geothermal Res.* 153 (1-2), 37–50. doi:10.1016/j.jvolgeores.2005.08.008
- Newhall, C. G., Costa, F., Ratdomopurbo, A., Venezky, D. Y., Widiwijayanti, C., Win, N. T. Z., et al. (2017). WOVODat - an Online, Growing Library of Worldwide Volcanic Unrest. *J. Volcanology Geothermal Res.* 345, 184–199. doi:10.1016/j.jvolgeores.2017.08.003
- Pan, S. J., and Yang, Q. (2009). A Survey on Transfer Learning. *IEEE Trans. Knowledge Data Eng.* 22 (10), 1345–1359. doi:10.1109/TKDE.2009.191
- Powell, T. W., and Neuberg, J. (2003). Time Dependent Features in Tremor Spectra. *J. Volcanology Geothermal Res.* 128 (1-3), 177–185. doi:10.1016/s0377-0273(03)00253-1
- Power, J. A., Lahr, J. C., Page, R. A., Chouet, B. A., Stephens, C. D., Harlow, D. H., et al. (1994). Seismic Evolution of the 1989–1990 Eruption Sequence of Redoubt Volcano, Alaska. *J. Volcanology Geothermal Res.* 62 (1-4), 69–94. doi:10.1016/0377-0273(94)90029-9
- Rodgers, M., Pyle, D. M., and Mather, T. A. (2015b). “Statistical Analysis of Large Volcano Seismology Datasets to Determine Patterns of Volcanic Behaviour,” in *AGU Fall Meeting Abstracts*, San Francisco 2015, V14B–V01.
- Rodgers, M., Roman, D. C., Geirsson, H., LaFemina, P., McNutt, S. R., Muñoz, A., et al. (2015a). Stable and Unstable Phases of Elevated Seismic Activity at the Persistently Restless Telica Volcano, Nicaragua. *J. Volcanology Geothermal Res.* 290, 63–74. doi:10.1016/j.jvolgeores.2014.11.012
- Rodgers, M., Smith, P. J., Mather, T. A., and Pyle, D. M. (2016). Quiescent-explosive Transitions during Dome-Forming Volcanic Eruptions: Using Seismicity to Probe the Volcanic Processes Leading to the 29 July 2008 Vulcanian Explosion of Soufrière Hills Volcano, Montserrat. *J. Geophys. Res. Solid Earth* 121 (12), 8453–8471. doi:10.1002/2016jb013180
- Roman, D. C., and Cashman, K. V. (2006). The Origin of Volcano-Tectonic Earthquake Swarms. *Geol* 34 (6), 457–460. doi:10.1130/g22269.1
- Roman, D. C., De Angelis, S., Latchman, J. L., and White, R. (2008). Patterns of Volcanotectonic Seismicity and Stress during the Ongoing Eruption of the Soufrière Hills Volcano, Montserrat (1995–2007). *J. Volcanology Geothermal Res.* 173 (3-4), 230–244. doi:10.1016/j.jvolgeores.2008.01.014

- Roman, D. C., and Power, J. A. (2011). Mechanism of the 1996-97 Non-eruptive Volcano-Tectonic Earthquake Swarm at Iliamna Volcano, Alaska. *Bull. Volcanol* 73 (2), 143–153. doi:10.1007/s00445-010-0439-7
- Scarpetta, S., Giudicepietro, F., Ezin, E. C., Petrosino, S., Del Pezzo, E., Martini, M., et al. (2005). Automatic Classification of Seismic Signals at Mt. Vesuvius Volcano, Italy, Using Neural Networks. *Bull. Seismological Soc. America* 95 (1), 185–196. doi:10.1785/0120030075
- Sener, O., and Savarese, S. (2017). *Active Learning for Convolutional Neural Networks: A Core-Set Approach*. arXiv preprint arXiv:1708.00489.
- Settles, B. (2009). *Active Learning Literature Survey Technical Report*. Madison: University of Wisconsin-Madison, Department of Computer Sciences.
- Smailagic, A., Costa, P., Noh, H. Y., Walawalkar, D., Khandelwal, K., Galdran, A., et al. (2018). “Medal: Accurate and Robust Deep Active Learning for Medical Image Analysis,” in Proceeding of the 2018 17th IEEE international conference on machine learning and applications (ICMLA), Orlando, FL, USA, 17-20 Dec. 2018 (IEEE), 481–488. doi:10.1109/ICMLA.2018.00078
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a Simple Way to Prevent Neural Networks from Overfitting. *J. machine Learn. Res.* 15 (1), 1929–1958.
- Thelen, W. A., Allstadt, K., Angelis, S. D., Malone, S. D., Moran, S. C., and Vidale, J. (2013). Shallow Repeating Seismic Events under an alpine Glacier at Mount Rainier, Washington, USA. *J. Glaciol.* 59 (214), 345–356. doi:10.3189/2013jog12j111
- Thomas, M. E., and Neuberg, J. (2012). What Makes a Volcano Tick-A First Explanation of Deep Multiple Seismic Sources in Ascending Magma. *Geology* 40 (4), 351–354. doi:10.1130/g32868.1
- Thompson, G., Power, J. A., Braunmiller, J., Lockhart, A. B., Lynch, L., McCausland, W., et al. (2020). Capturing, Preserving, and Digitizing Legacy Seismic Data from the Montserrat Volcano Observatory Analog Seismic Network, July 1995–December 2004, July 1995–December 2004. *Seismological Res. Lett.* 91 (4), 2127–2140. doi:10.1785/0220200012
- Tilling, R. I. (1989). Volcanic Hazards and Their Mitigation: Progress and Problems. *Rev. Geophys.* 27 (2), 237–269. doi:10.1029/rg027i002p00237
- Van der Maaten, L., and Hinton, G. (2008). Visualizing Data Using T-SNE. *J. machine Learn. Res.* 9 (11), 2579–2605.
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Manley, Mather, Pyle, Clifton, Rodgers, Thompson and Londoño. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.