Advanced signal processing for micro-seismic signal analysis

PhD Thesis

Jiangfeng Li

in

Electronic and Electrical Engineering

University of Strathclyde, Glasgow

May 2, 2023

To my family

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

Seismic events are brittle failures mainly attributed to the reduction in effective stress. They are typically nonstationary signals with small moments (known as microseismic events) and are frequently exposed to various types of ambient noise, resulting in low Signal-to-Noise Ratio (SNR) sensor readings. However, an efficient and accurate (micro)seismic monitoring system is highly demanded, especially for unstable slopes such as landslides. Unlike the volume of literature that focuses primarily on earthquakes or volcanic seismic activities, the research on unstable slope monitoring requires addressing more challenging signal recordings from events such as quake/slide-quake considering microseismic signals, rockfall, earthquake, and anthropogenic noise. Thus, advanced signal processing approaches must be regarded as for signal denoising, event detection, feature engineering, and classification. This thesis first proposes an end-to-end platform containing denoising via Graph-Based Bilateral Filter (GraphBF), detection via Neyman-Pearson lemma, and classification via Graph Laplacian Regularisation (GLR) to identify the potential (micro)seismic events from raw observations. Secondly, an evaluation of feature engineering for (micro)seismic signal classification is proposed; the contribution concentrates on feature space optimisation with graph learning. Eventually, this thesis proposes a novel deep learning-based multitask learning to classify the (micro)seismic with little domain knowledge. For all proposed methods, the competitive performance is demonstrated in terms of accuracy and efficiency compared to state-of-the-art approaches, with the datasets collected at ongoing landslides.

Contents

List of Figures

List of Tables

Glossary

$\mathbf{A}$	Graph weighted adjacency matrix
A	Dimension of second derivative component
$\mathbf{B}$	Spatial-domain second derivative
B	No. of classes
β	Standard deviation
C	No. of recording channels
$\mathbf{D}$	Graph degree matrix
e_k	Weight of feature k
$\mathcal{F}$	A set of feature vectors
$\mathcal{G}$	Graph
$\mathbf{g}$	Lagrange multiplier
$\Gamma(\cdot)$	Gamma function
h	No. of hypotheses
$\mathbf{I}$	Identity matrix
J	No. of recorded data samples
K	No. of the features
$\mathbf{L}$	Graph Laplacian matrix
l	Length of the window

λ	Detection threshold
$\mathcal{M}$	A set of velocity models
N	No. of graph vertices
$\mathcal{O}$	A subset of features
O	No. of provided velocity models
$\mathcal{P}$	A set of trainable probability tensors
P_D	Probability of detection of Neyman-Pearson lemma
P_{FA}	Probability of detection false alarm of Neyman-Pearson lemma
$\mathbf{r}$	Stacked signal
$\mathbf{s}$	Graph signal
σ	The scaling factors of graph adjacency matrix
$\mathbf{T}$	Wavelet-decomposed signal
$\mathbf{tinv}(\cdot)$	Student's t inverse cumulative distribution function
$\mathbf{U}$	Time-domain second derivative
$\mathbf{u}$	Sesimic signal displacement vector
$\mathcal{V}$	Graph node
$\mathbf{v}$	P-wave velocity
$\mathcal{W}$	A set of segment windows
$\mathbf{x}$	A vector dataset
y	Predicted label

Acronyms

3C	Three-Component
Acc	Accuracy
ANN	Artificial Neural Network
BPF	BandPass Filtering
CapsNet	Capsule Neural Network
CFS	Cross-correction Feature Selection
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DFT	Discrete Fourier Transform
DT	Decision Tree
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
FN	False Negative
FP	False Positive
FSLib	Feature Selection Library
GCN	Graph Convolutional Network
GLR	Graph Laplacian Regularisation
GraphBF	Graph-Based Bilateral Filter

GSP Graph Signal Processing

GTV Graph Total Variation

Inf-FS Infinite Feature Selection

KNN K-Nearest Neighbor

LDA Linear Discriminant Analysis

LIME Local Interpretable Model-Agnostic Explanation

MCM Multi-channel Coherency Migration

MLP Multilayer Perceptron Neural Network

NB Naive Bayes

PASSCAL Portable Array Seismic Studies of the Continental Lithosphere

PCA Principal Component Analysis

PDF Probability Density Function

RANSAC RANdom SAmple Consensus

ResNN Residual Neural Network

RF Random Forest

RNN Recurrent Neural Network

SDR Semi-Definition Relaxation

SMOTE Synthetic Minority Over-Sampling Technique

SNR Signal-to-Noise Ratio

STA/LTA Short-Term-Average/Long-Term-Average

STFT Short-Time Fourier Transform

SVM Support Vector Machine

TN True Negative

TP	True Positive
t-SNE	t-distributed Stochastic Neighbour Embedding
Vp	P-wave Velocity
WPD	Wavelet Packet Decomposition

Acknowledgement

I would like first to thank my supervisors, Dr. Lina Stankovic, Dr. Stella Pytharouli, and Prof. Vladimir Stankovic, for their long-term guidance and inspiration on my Ph.D. work. Without their endless help and encouragement, I would never have done this work. I would also like to thank Dr. Floriane Provost, Dr. Naomi Vouillamoz, Dr. Mingxiang Ye, Dr. Bochao Zhao, Dr. Kanghang He, Dr. Cheng Yang, Dr. Zhiwang Feng, Prof. Qingjaing Shi, Prof. Gene Cheng for their generous help, thoughtful advice, and comments. I would also like to thank all the CIDOM group members, especially David Murray and Jiaxin Jiang, for their assistance given during my Ph.D. work. It's my pleasure and honor to have the opportunity to work with them and learn from them.

This work might have never happened without the financial aid of the University of Strathclyde Faculty of Engineering Scholarship (2017-2020) and the University of Strathclyde Department of Electronic and Electrical Engineering Scholarship (2017-2020). This work is also supported by the SENSIBLE project funded by the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 734331 and EPSRC Prosperity Partnership research and innovation programme under grant agreement EP/S005560/1.

Finally, I would like to thank my family for all their help and support throughout this challenging work.

Chapter 1

Introduction

1.1 Slope Stability Monitoring

Slope stability failures are inconspicuous but worldwide daily threats, which secondary disasters: rockfall, topple, lateral spread, rotational slide, translational slide, and flow usually result in catastrophic consequences for life and property [5–8]. Slope instability can be caused by 1) external factors: deformation of the slope foundation, overloads and seepage of water on the crest of the slope, lack of drainage system inside the slope, induced vibrations (traffic), and seismic events; 2) internal factors: increasing of the water table and pore pressure, loss of shear strength for the soil (e.g., change of effective stress) piping and liquefaction [5, 6].

Slope stability prevention measures involve assessing the relative magnitude of resisting forces (stability) and driving forces (instability) to establish and enhance slope stability. Typically, the data acquisition for slope stability monitoring is mainly via field surveys, unmanned aerial vehicles, aerial platforms, satellite platforms, in-situ monitoring systems, and seismic reflection [9]. In particular, the in-situ monitoring systems demonstrate superiority with (near)real-time data recording to provide comprehensive information for fast-moving hazards such as

landslides.

Generally, slope stability monitoring approaches have some drawbacks: 1) field surveys can be time-consuming and laborious; 2) Unmanned aerial vehicles can be affected by environmental factors such as strong wind , *etc.*; 3) Aerial photography and Satellite platforms can have high cost and are susceptible to atmospheric conditions such as clouds, snow , *etc.* [9, 10]. While, the in-situ microseismic monitoring systems have the characteristic of simple installation and maintenance, and the in-situ sensor arrays deployed in strategic areas with arbitrary layouts can record the seismic signals from a range of sources within 10s km distance [9]. The research priorities of this thesis concentrate on analysing the microseismic or seismic signal recorded by seismic networks deployed locally with advanced signal processing and machine learning approaches.

1.2 (Micro)seismic Event Analysis

(Micro)seismic event could intuitively be described as a naturally sourced sudden brittle failure or an artificially triggered explosion, a rapid release of energy in a relatively negligible region that shakes the surrounding medium and propagate the waves in the distance [11].

Specifically, for example, 1) for volcanic seismology, the signal generation mechanism relates to the physical behavior of the volcano, such as the transport of gas and fluids (e.g., water, magma) and their interaction with solid rock contains: volcanic-tectonic, long-period signal, tremor, and quake [12]; 2) for the injection process (e.g., string shot, perforation shot, plug setting, ball drop, and sleeve opening) of hydraulic fracturing and carbon capture and storage, the source mechanism could be catalogued as volumetric, and shear double-couple component [11, 13]; 3) besides, (micro)seismic events could be derived from unstable slopes such as open-pit mining and landslide caused by the generated elastic

accumulation, rupture, friction and shear between soil particles [14] *etc.*

Seismic investigations of slope instabilities emerged in the 1960s and have been complemented during the last decades by an abundance of microseismic monitoring studies [14], which also benefit from the improvements in seismometer design (high sensitivity relative to other slope monitoring technologies), installation (low cost, and no need of high-power supply), and network density [9]. Generally, the critical motivator of (micro)seismic event analysis is to assist the geological hazards assessment (e.g., enhance the understanding of the underlying processes) to mitigate or prevent future failures and further reduce the social and economic consequences of damaging disasters [9].

(Micro)seismic signals can sometimes be the precursor of landslides with low dominant frequency and nonstationary characteristics. Consequently, obtaining a solid foundation of knowledge for the various (micro)seismic signals is critical to effectively predict major geological disturbances to minimise fatalities and infrastructure damage [14–16]. However, the open issues of (micro)seismic signal monitoring contain: 1) low SNR, especially for microseismic signals; 2) massive continuous recorded data raises difficulties for the following processing and analysis; 3) recordings contain multisource signals generated by non-landslide related sources (wind, rainfall, human activities, animals, *etc.*) as well as unstable ground; 4) expert knowledge required for signal processing, the involved human interventions lead to elevated uncertainty [4, 14, 15].

1.2.1 Detection

In seismic analysis, efficient and highly accurate seismic detection plays a crucial role. Generally, manually detecting signals of interest from continuous seismic recordings with expert knowledge is time-consuming, laborious, and subjective. The increasing trend in recent years has been the creation of algorithms that automatically detect seismic signals. Short-Term-Average/Long-Term-Average

(STA/LTA) is the most widely utilised approach, which slides the short and long windows with the preset lengths on the time/frequency domain amplitude, envelope, or the high order statistic features (e.g., skewness and kurtosis) [4, 17–21] of the recorded signal. The events are identified using the trigger and detrigger thresholds on the short and long-window average value ratio. Recently, STA/LTA and its modifications have been extensively utilised in earthquake [22], volcanic [23], and landslide-induced [4, 20, 21] events detection. However, the limitations of STA/LTA are obvious, for example, 1) the inappropriate parameter initialisation leads to false alarms [24]; 2) too sensitive to ambient noise [15, 25, 26]; 3) the window length and threshold selection processes can be time-consuming and inefficient [27,28]. Established on STA/LTA, the modified energy ratio detection method is proposed with the same length of the pre-and post-sample windows [29]. Later, three modified detection approaches established on the seismic attributes (energy ratio, fractal dimension, and entropy) are proposed [30]. The Akaike Information Criterion detection algorithm proposed by [31] is based on the concept that nonstationary seismic signals can be approximated by dividing an observed waveform into locally stationary segments, with each segment treated as an autoregressive process. The detection method known as matched filtering based on template matching necessitates prior knowledge of representative parent waveforms [32, 33].

Low SNR and varying ambient noise strength and dispersion present the critical obstacles to (micro)seismic event detection [28, 34]. Increasing the SNR by damping random ambient noise, the previous research employs stacked absolute values and the absolute value of the product of the amplitudes of the multi-channel recordings as the input for detection approaches [35, 36]. Strong ambient noise in any of the components (recording channels), however, degrades the quality of the stacked signals. Additionally, if the ambient noise is present just on one of the components, the result effectively suppresses it while also drastically

reducing the signal amplitude [34].

In general, as reviewed in [12, 14, 28], the current detection approaches are suitable for well-known routine seismic signatures but fail for unknown and unexpected low SNR microseismic events (e.g., landslide induced slide quakes, rockfall).

1.2.2 Feature Engineering

Recent researches indicate that feature engineering (feature construction, extraction, and selection) is a crucial step toward efficient signal classification, as a large set of features with redundant information could increase the processing time and cause classifier overfitting, multicollinearity, and suboptimal feature ranking at the selection stage [37]. A detailed review of feature construction for (micro)seismic events is provided in [15], where temporal, spectral, and cepstral features and combinations thereof are derived from the raw denoised measurements. Feature extraction [38], and selection [37] are commonly used for dimensionality reduction of the feature space, thus decreasing the required storage and testing and training time of the classifier.

As the most popular feature extraction (and dimensionality reduction) method, Principal Component Analysis (PCA) has been consistently shown to be effective for a range of (micro)seismic events detection and classification tasks. For example, (1) PCA is adapted with Artificial Neural Network (ANN) to classify microseismic events and quarry blast [38]; (2) utilising the PCA extracted features, the seismic events classification accuracy obtained by Support Vector Machine (SVM) classifiers with linear and Radial basis function kernels got improved [39]. While PCA is hampered by the high computational complexity of singular value decomposition, further calculated principle components cannot reliably identify the most crucial variables for information preservation and interpretability.

Feature selection methods are often categorised as filter-based (most popu-

lar for (micro)seismic analysis), wrapper-based, embedded, hybrid, and ensemble approaches. Filter-based methods, based on evaluating and selecting the features with various statistical tests, are model-agnostic, i.e., they can be applied to any learning algorithm to exclude irrelevant and redundant features, and are of lower complexity [40]. Thus, these methods have been widely explored in various (micro)seismic analyses. For example, (i) Information Gain, One Rule, Relief, Chi2 Discretisation, and uFilter filter feature selection approaches are utilised with a Gaussian Mixture Model classifier to classify volcanic-seismic signals [37], (ii) the Relief filter approach is proposed with SVM for classifying levee passive seismic signals in the earth dam [41], (iii) filter (mutual information and statistical dependence) methods and embedded (cross-validation and pruning) methods are explored to classify volcanic-seismic signals with K-Nearest Neighbor (KNN) and Decision Tree (DT) [42], (iv) Information Gain filter method is employed with ANN to predict earthquakes [43]. Although the aforementioned studies demonstrated that filter-based feature selection successfully enhances categorisation outcomes in various (micro)seismic signals, filter methods often cannot identify the discriminate features, such as those associated with long tail distribution. Specifically, for the features with long-tailed distribution, the feature value of the majority of samples may be concentrated on a small range, while the remaining feature value may have little to no information gain. The measure of similarity of the feature selection chosen depends on the data type of both the input and output variables [40]. This thesis chooses statistical ways to quantify similarities, such as mutual information or chi-square, *etc.* which are entropy-based metrics that are commonly used in feature selection attempts on classification tasks with numerical input and categorical output. [40]. However, these statistical-based filter feature selection approaches can be challenging to extract useful information from these long-tailed features [40]. Additionally, the filter methods do not eliminate multicollinearity (a statistical concept where several independent vari-

ables in a model are correlated), which could result in the selected features being sub-optimal for signal discrimination. Recently, wrapper methods are also employed to enhance volcanic-seismic signals classification [44] [45], based on the inferences from the classification model, and the performance surpasses the filter-based methods, while usually confined to a high level of computational complexity and subjects the model to overfitting.

1.2.3 Classification

Regarding (micro)seismic classification, the main challenges are 1) insufficient open access annotated datasets [15]; 2) the imbalanced catalogue of labelled events caused by the sparsity of events of interest [15]; 3) high similarities between unknown natural and anthropogenic "interfering" signals and events of interest in time or frequency domain [4]. Thus, manual classification is still widely used, e.g., (micro-)earthquake, block fall, rockfall, quarry blast, and multiple events are manually classified based on frequency range, amplitude, signal shape, and duration [46]. However, (micro)seismic events classification via visual inspection is a tedious process that necessitates detailed domain knowledge [46]. Recently, (micro)seismic events classification has evolved into a trend that depends on automatic machine learning approaches.

Automatic classification approaches such as SVM and Random Forest (RF) are usually proposed with some feature extraction or selection methods, as reviewed above. SVM can solve the high dimensional non-linear classification problem with a small number of training samples [47], which is utilised to distinguish long period event, tremor, and volcanic tectonic in [12, 48], earthquake and non-earthquake events in [49], *etc.* RF is parallelisable, performs well for high dimensional signals, quick in prediction/training, robust to outliers and non-linear data, handles imbalanced data, and low bias [50], and is employed for the classification of landslide (micro)seismic events, namely rockfall, slide quake, earthquake, and

natural/anthropogenic noise [4,9]. Other classification models such as Multilayer Perceptron Neural Network (MLP), Linear Discriminant Analysis (LDA), Naive Bayes (NB), KNN, and Feed-Forward Back-Propagation Neural Network are also investigated to distinguish volcanic-seismic events in [51, 52]. Conventional (micro)seismic event classification relies on handcrafted feature construction, extraction, and selection, commonly performed requiring extensive domain knowledge for the physical characterisation of events.

Conversely, deep learning is considered to provide an integration of detection, feature representation, and classification with competitive performance, assuming that a good representative dataset is available for training requiring little or no domain knowledge. Since the first successful Convolutional Neural Network (CNN) model was proposed, deep learning thriving in all aspects, which also prompted the development of seismic signal analysis [9]. Recently, plenty of CNN-based architectures are proposed for (micro)seismic signal classification with the input of 1) multi-channel recorded time series signal [53–58], 2) frequency domain signal or spectrogram [59–64], 3) handcrafted features [65, 66]. The emerging attention model and Graph Convolutional Network (GCN) is also employed for seismic classification task [67, 68].

However, the main issue with the current models is their lack of interpretability [9], which limits their application due to not only trust issues but also difficulties in debugging, improving performance, making fair comparisons across approaches and datasets, and using the results to improve general understanding of underlying seismic processes.

1.3 Dataset

1.3.1 Seismic Instruments

Seismometers are instruments that measure and record ground motions, such as seismic waves produced by earthquakes, nuclear explosions, and other seismic sources. Seismologists can map the Earth's interior and locate and measure the size of these various sources using seismic wave recordings.

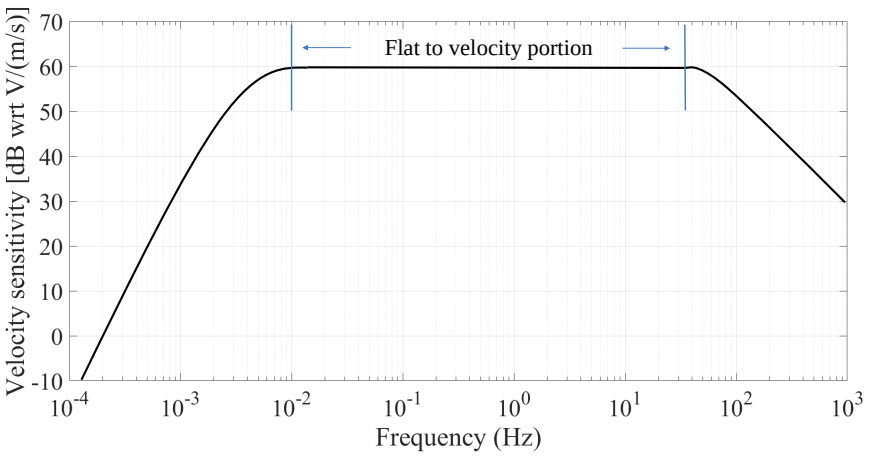


Figure 1.1: Flat-to-velocity of broadband Seismometer.

Regards to seismometers, two important parameters need to be clarified first: flat-to-velocity portion and corner periods.

- The flat spectrum response portion (as shown in Fig. 1.1) of a seismometer's response curve refers to the range of frequencies for which the seismometer's output is proportional to the velocity of ground motion, and not affected by the seismometer's natural frequency response and damping [69].

- The corner period (as shown in Fig. 1.2, where the unit of acceleration is the acceleration of gravity $g = 9.80m/s^2$) of the seismometer is defined as the period at which the voltage output of a seismometer starts to decrease due to

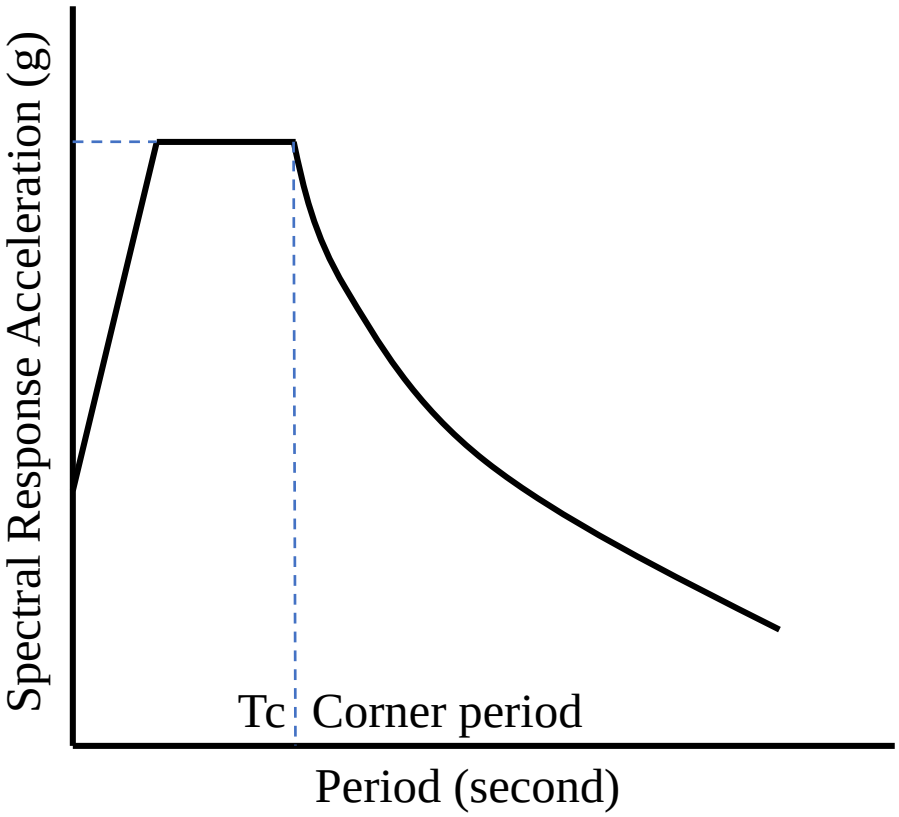


Figure 1.2: Corner period of Seismometer.

damping and is not proportional to the ground movement velocity anymore. The corner period is an important parameter because it determines the range of periods that the structure is most sensitive to [70].

The Incorporated Research Institutions for Seismology Portable Array Seismic Studies of the Continental Lithosphere (PASSCAL) Instrument Center supports the following types of seismometers [70]:

- Broadband Sensors - are Three-Component (3C) seismometers capable of sensing ground motions over a wide frequency range and are commonly used in passive seismic monitoring. The flat-to-velocity portion of the bandwidth typically ranges from $0.01 - 25\,\text{Hz}$.

- Intermediate Sensors - are 3C seismometers with corner periods ranging from $30 - 40\,\text{s}$ (as opposed to the $120 - 240\,\text{s}$ of PASSCAL's truly broadband fleet). Similar to broadband sensors, these sensors can detect ground motions with much longer periods than their corner periods if the long-period amplitudes are sufficient.

- Short-Period Sensors - are rugged seismometers that cover higher-frequency bands (typically $1 - 100+$ Hz) and can be used in passive and active-source experiments. The sensors can be either feedback seismometers that require power or conventional, passive seismometers requiring no external power.

- High-Frequency Sensors - are extremely rugged seismometers that cover even higher frequency band $4.5 - 100+$ Hz. These sensors, also known as geophones, are most commonly utilised in active-source experiments.

- Accelerometers - also known as strong-motion sensors, are designed to measure the large amplitude, high-frequency seismic waves typical of large local earthquakes and operate in the frequency band $0 - 100+$ Hz.

Figure 1.3: Typical components of a local short-period seismic array: 4 seismometers, data logger (in yellow box), power supply (in red box), geophone cables.

The typical sensor array utilised for microseismic monitoring, as shown in Fig. 1.3, comprises a battery unit (red box), a data logger (yellow box), and four sensors. The seismometers are usually formatted as a triangular array. The intervals are arranged in tens to several hundreds of meters, with one 3C sensor (the larger one) and three vertical one-component sensors deployed at the triangle's center and vertexes.

1.3.2 Dataset Utilised in this Thesis

SZ10, PG

The datasets overview for SZ10, PG15, and PG16 are provided in [14]. The SZ10 data was collected with a sampling rate of Fs = 1000 Hz, and the two PG datasets with a sampling rate of Fs = 500 Hz. These datasets have been pre-processed and manually labelled with expert knowledge as described in [14]. The datasets use SZ to represent Super-Sauze (Southeastern France) and PG to refer to Pechgraben (Upper Austria). Seismic events in the datasets have been identified based on the arrival times of P-waves and their durations. Since the sensor array was placed at the same site but at separate times for collecting the PG15 and PG16 datasets, they were combined and named the PG dataset for the following analysis.

In the SZ10 dataset, nine different types of seismic events are distinguished: local earthquake, distant (teleseismic) earthquake, distant slide quake, nearfield low-frequency slide quake, nearfield microearthquake, multiple tremors (rockfall/near repeater), dispersive tremor, undefined sinusoidal event (harmonic), and calibration shot. The PG dataset consists of six classes: local earthquake, distant earthquake (teleseismic), distant slide quake, multiple tremors (rockfall/near repeater), undefined sinusoidal event (harmonic), and calibration shot. Note that the calibration shot is an artificially induced signal. Accurately detecting and classifying these signals will help assess the performance of the proposed ap-

proaches.

Table 1.1: No. of events in the SZ10 and PG datasets.

Class name	Type of event	SZ10	PG
1.1	Local earthquake	50	7
1.2	Distant earthquake (Teleseismic)	6	9
2.2	Distant slide quake (Moderate distant quake)	39	6
2.3	Nearfield low frequency slide quake	12	0
2.4	Nearfield microearthquake	7	0
3.1	Multiple event (Rock fall/near repeater)	15	5
3.2	Dispersive tremor	7	0
3.4	Undefined sinusoidal event (Harmonic)	15	4
9	Calibration shots	11	5
99	Undefined signal (Environmental noise)	12	6
Total number of events		174	42

Events associated with earthquakes, both local and distant, can be the preliminary indications of a landslide. The dominant frequency of the local earthquake is around $1-20\,\text{Hz}$, while for distant earthquake, can be under $5\,\text{Hz}$.

Quake events, characterised by the low dominant frequency of $5-50\,\text{Hz}$, include distant and nearfield slide quakes and nearfield microearthquakes. They all have small durations, usually, around $2\,\text{s}$ and up to $10\,\text{s}$ for microearthquakes. Additionally, microearthquake events can typically be recognised by their distinct P and S phases.

A relatively long duration characterises tremors from $2\,\text{s}$ up to several minutes. Because the frequency content overlaps with ambient noise caused by vehicles or other moving objects, the tremors can be misinterpreted [71].

Calibration shots artificially carried out in the study area are usually utilised to evaluate amplitude attenuation patterns empirically. Accurately detecting and classifying these signals could assess the proposed approaches in this research.

Table 1.2: No. of events in the SZ13, 14, and 15 datasets.

	No. of events			
Type of event	**Total**	SZ13	SZ14	SZ15
Rockfall	**401**	66	325	10
Slide Quake	**234**	132	97	5
Earthqauke	**388**	138	87	163
Natural/Anthropogenic noise	**351**	190	104	57
Total No. of events	**1374**			

SZ13, SZ14 and SZ15

The raw waveform of SZ13, SZ14, and SZ15 are publicly available under (`https://seismology.resif.fr/networks/#/MT`) accessed on 8 January 2022, in the periods (11 October to 19 November 2013 (SZ13), 10 to 30 November 2014 (SZ14), and 9 June to 15 August 2015 (SZ15)). The sensor arrays formed as two permanent triangular shapes (array A named SZC and array B named SZB) with 40 m sensor intervals, which was monitored by the French Landslide Observatory (Observatoire Multi-disciplinaire des Instabilités de Versants) OMIV. Each array comprises one 3C center site broadband seismic recorder (RefTek 130S-01) and three vertical 1-component Short-Period sensors (Noemax and Sercel L4C) with a sampling rate of Fs = 250 Hz, and a flat response in the range $5 - 100$ Hz. Thus, each sensor array contains six channels; the layouts as illustrated in Fig. 1.3. A catalogue of manually verified events, detected with spectrogram analysis, is provided [4]. This thesis uses the same labels for the four classes in the catalogue: 401 "Rockfall" events, 234 "Slide Quake" events, 388 "Earthquake" events, and 351 "Natural/Anthropogenic noise" events, totaling 1374 events (see Table 1.2).

In general, the data utilised in the thesis was recorded at the ongoing clay-based landslides, consisting of the endogenous landslide seismicity (e.g., slide quake and rockfall) and signal attributed to external causes (e.g., earthquake and natural/anthropogenic noise), which distinguished through the mechanism of generation. With the taxonomy based on event moment, the recorded signal can

be catalogued as microseismic signals with negative moments (typically between -3 and 0) and seismic signals with higher moments [11]. Despite the lack of rigorous signal moment measurement, most research claims that the endogenous seismicities at the landslide are microseismic signals [4, 14].

Although the data utilised in this thesis is recorded at the clay-based landslides, the proposed signal-processing approaches and machine-learning techniques are developed for general time series seismic signals. In general, this thesis explores the (semi)-supervised learning techniques, which can be trained with or without unlabelled data. Depending on the quality of the data and the features selected, it may be possible to use these techniques on soils that are not composed of clay [72]. In general, different types of soil can affect seismic signals in different ways, thus the selection of pertinent data and features, as well as the adjusting of the parameters of the machine learning model, need to be adapted to the various soil conditions [72]. For example, using one-dimensional convolutional neural networks in combination with unsupervised classification techniques to extract seismic waveform features, several studies have tested machine learning methods for seismic signal analysis on various types of soils and have produced promising classification results [72]. Hence, after proper adaptation and retraining, the seismic signal analysis methods established by machine learning techniques on clay-based soils may be extended to other non-clay-based soils.

1.4 Research Aim and Objectives

The main aim of this research is to explore and analyse the landslide-induced (micro)seismic events to facilitate the understanding of slope deformation. This thesis is not focused on predicting landslides. Rather, its main objective is to detect landslides and classify the underlying factors characterising landslides through the analysis of time series seismic signals. Specifically, landslides are destructive

and unpredictable, which are considered the most common geological hazards, and at least 17% of worldwide natural hazard fatalities can be attributed to landslides [73], and the landslide-induced microseismic signal such as rockfall, slide quake, and tremor *etc.* are the internal factors of the slope instabilities. The proposed (micro)seismic signal analysis approaches in this thesis, motivated by emerged Graph Signal Processing (GSP) and deep learning algorithms, cover denoising, detection, feature engineering, and classification to assist the understanding of landslide deformation.

The main research objectives are listed below:

- Develop an effective and end-to-end platform for the landslide-induced (micro)seismic events to understand the ongoing slope deformation and provide the reference for decision-making for engineers with little or no geophysics expertise. The proposed highly accurate end-to-end platform should enable efficient monitoring, further providing timely information.
- Explore efficacious feature selection and extraction methods with predefined handcrafted features to improve the predictive performance of signal classification to enhance the understanding of the characteristic of the landslide-induced (micro)seismic events.
- Develop a generic deep learning-based classification method that combines the (micro)seismic wave propagation physic information with expert knowledge inference and enhances the model interpretability.

1.5 Contribution of the Thesis

The major contributions of this thesis to achieve the research objectives can be summarised as follows:

1. This research develops an automatic end-to-end platform comprising signal denoising, event detection, feature construction, and classification, to efficiently and accurately monitor landslide-induced (micro)seismic events. The experimental results are evaluated utilising the datasets collected at the ongoing landslides, demonstrating the competitiveness of the proposed platform w.r.t state-of-the-art.

    1.1 To improve the signal SNR, this research novelty adapts GraphBF on (micro)seismic recordings, and the hyper-parameters are tuned with the characterisation of the ambient noise distribution.

    1.2 To simplify the detection, this research applies a low-complexity thresholding method based on the Neyman-Pearson lemma to detect all the verified (micro)seismic events with few false detections.

    1.3 To obtain the catalogue of the detected events, this research adapts the GLR classifier, which is known as semi-supervised learning and performs well with limited training samples.

2. To facilitate the understanding of signal characterisation, this research provides a (micro)seismic events detection with multi-channel continuous recordings and graph-based feature weight optimisation and classification.

    2.1 To mitigate the impact of variable ambient noise signal recordings, this research proposes a detection scheme that combines Multi-channel Coherency Migration (MCM) and Neyman–Pearson lemma so that the seismic events recorded by a portion of the channels are readily identifiable, which is more in line with practical application scenarios.

    2.2 This research adapts the graph-based feature weight optimisation approach for (micro)seismic event classification and proposes a novel graph kernel bandwidth optimisation to learn the optimal representation graph.

2.3 This research assesses the proposed detection scheme and graph-based feature weight optimisation, respectively, on the catalogued landslide-induced (micro)seismic events and the continuous recordings (24-28/November/2014) verified by an expert.

2.4 This research explores in detail the impact of feature engineering (filter, wrapper, embedded-based feature selection approaches, feature extraction with PCA, and adapted graph-based feature weight optimisation) on landslide-induced (micro)seismic event classification and concludes that with graph smoothness, the features highlighted with adapted graph-based feature weight optimisation are more discriminative.

2.5 Following graph-based feature weight optimisation, This research contributes a feature recommendation list of rockfall, slide quake, earthquake, and natural/anthropogenic noise occurrences, which summarise the most distinct characteristics of each of the aforementioned classes.

3. To investigate the question of "if physical information could enhance the interpretation while improving the classification model's performance", this research novel employs the (micro)seismic wave equation and propagation velocity model for classification. Specifically, this research proposes a multitask learning scheme that includes: 1) a novel signal representation learning/reconstruction task with two CNN architectures to estimate the signal propagation temporal and spatial characteristics from the real-field (micro)seismic recordings; 2) a classification task with the novel CNN architecture.

1.6 Organisation of the Thesis

The remainder of the thesis is organised as follows:

Chapter 2 reviews the background, application, and general definition of GSP.

Chapter 3 introduces the general background and the open challenges for the automatic end-to-end (micro)seismic analysis platform, comprising signal denoising, detection, feature engineering, and classification. Then the performance of the proposed methods is compared with state-of-the-art approaches.

Chapter 4 starts with the general review of feature engineering (feature construction, extraction, and selection) on (micro)seismic analysis. Then the proposed graph-based alternative feature weight optimisation is introduced to assist with graph-based classification. Finally, the performance of the proposed approach is evaluated with benchmarking methods.

Chapter 5 expends the interpretability of proposed graph-based feature weight optimisation and classification with handcrafted features.

Chapter 6 first introduces a novel multitask learning scheme with seismic wave equation and velocity model for landslide-induced rockfall, slide quake, earthquake, and natural/anthropogenic noise classification. Then the model interpretability, enhanced by the detailed physical information, is introduced.

Chapter 7 concludes this thesis and justifies the remaining challenges in (micro)seismic signal processing.

1.7 Publications

Journal Articles

J1 Li. J, Ye. M, Stankovic. L, Stankovic. V, and Pytharouli. S, Domain knowledge informed multitask learning for landslide induced seismic classification. Submitted to *IEEE Geoscience and Remote Sensing Letters*, Dec. 2022.

Contribution: Experiments design, Algorithm implementation, Results analysis, Formal analysis, Draft writing.

The publication relates to Chapter 6

J2 Li. J, Stankovic. L, Stankovic. V, Pytharouli. S, Yang. C, and Shi. Q, Graph-based feature weight optimisation and classification of continuous seismic sensor array recordings. *Sensors*, Dec. 26. 2022.
Contribution: Experiments design, Algorithm improvement, and implementation, Results analysis, Draft writing, Formal analysis.
The publication relates to Chapter 4

J3 Li. D, Li. J, Zeng. X, Stankovic. V, Stankovic. L, Xiao. C, and Shi. Q, Transfer learning for non-intrusive load monitoring of multi-objects in smart building. *Applied Energy*, Oct. 24. 2022.
Contribution: Literature review, Editing and write-up.
The publication motivates the deep learning based classification for Chapter 6

J4 Li. J, Stankovic. L, Pytharouli. S, and Stankovic. V, Automated Platform for Microseismic Signal Analysis: Denoising, Detection, and Classification in Slope Stability Studies. *IEEE Transactions on Geoscience and Remote Sensing,* vol. 59, no. 9, pp. 7996-8006, 2020.
Contribution: Experiments design, Algorithm implementation, Results analysis, Formal analysis, Draft writing.
The publication relates to Chapter 3

J5 Herrero-Bermello. A, Li. J, Khazaei. M, Grinberg. Y, Velasco. A, V, Vachon, Vachon. M, Cheben. P, Stankovic. L, Stankovic. V, Xu. D, Schmid. J and Alonso-Ramos. C, On-chip Fourier-transform spectrometers and machine learning: a new route to smart photonic sensors. *Optics letters,* vol. 44, no. 23, pp. 5840-5843, 2019.
Contribution: Evaluation of SVM with proposed approach.
The SVM implementation for classification, supporting methodology for benchmarking in Chapter 3

Conference Papers

C1 Li. D, Li. J, Zeng. X, Stankovic. V, Stankovic. L, and Shi. Q, Non-intrusive load monitoring for multi-objects in smart building. *IEEE 2021 International Balkan Conference on Communications and Networking (BalkanCom)*, Sep. 2021.

Contribution: Literature review, Proofreading.

The publication motivates the deep learning based classification for Chapter 6

C2 Li. J, Yang. C, Stankovic. V, Stankovic. L, and Pytharouli. S, Graph-based micro-seismic signal classification with an optimised feature space. *2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2020.

Contribution: Experiments design, Algorithm implementation, Results analysis, Formal analysis, Draft writing.

The publication relates to Chapter 4

C3 Li. J, Stankovic. L, Pytharouli. S, and Stankovic. V, High-accuracy real-time microseismic analysis platform: case study based on the super-sauze mud-based landslide. *Geoconvention*, Feb. 2020.

Contribution: Experiments design, Algorithm implementation, Results analysis, Formal analysis, Draft writing.

The publication relates to Chapter 3

C4 Herrero-Bermello. A, Li. J, Khazaei. M, Grinberg. Y, Velasco. A, V, Vachon, Vachon. M, Cheben. P, Stankovic. L, Stankovic. V, Xu. D, Schmid. J and Alonso-Ramos. C, Smart on-chip Fourier-transform spectrometers harnessing machine learning algorithms. *SPIE Photonics West OPTO*, Feb. 2020.

Contribution: Evaluation of SVM with the proposed approach.

The SVM implementation for classification, supporting methodology for benchmarking in Chapter 3

Chapter 2

Graph Signal Processing

Graph Signal Processing (GSP) is a vibrant branch of signal processing considered to be powerful for large-scale structured data such as network and graph representation, processing, analysis, and visualisation, which is recently booming in applications such as social, energy, transportation, sensor, and neuronal network [74, 75]. Graphs are generic data representation forms that are useful for describing the geometric structures of data domains; the weight associated with each edge represents the similarity between the two connected vertices; the connectivities and edge weights are either dictated by the physics of the problem at hand or inferred from the data [75].

The emerging field of GSP merges algebraic and spectral graph theoretic concepts with computational harmonic analysis to process such signals on graphs [75]. Due to the superiority in exploiting the underlying connectivity information, GSP has evolved into multidisciplinary research that contains shifting, filtering, classification, clustering, Fourier transform, and interpolation, *etc.*; furthermore, it promotes the applied research in data mining and signal processing problems, ranging from denoising and data compression to classification, biomedical, and environmental data processing (see [74–78] and references therein).

Inspired by the initial success of GSP in many fields: time series denoising [79],

feature representation learning [80], graph learning for binary classification [81], graph-based clustering on eye tracker data analysis [82], and more applications in [83–85], in this thesis, the GSP-based automatic end-to-end seismic signal processing platform and feature weight optimisation scheme are proposed to perform low-complexity denoising, detection, feature learning and classification for the ongoing landslide induced (micro)seismic signal. GSP principles can be used to perform semi-supervised learning, is particularly suitable for data classification with limited training samples and insufficient to build appropriate class models [15].

2.1 General Definition

This section introduces the basic concepts of GSP, which are used in the remainder of the thesis.

In the following context of the thesis, $\mathbf{X}^\top$, $\mathbf{X}^{-1}$, $\mathbf{X}^\#$ and $\mathbf{X}^*$ respectively denotes transpose, inverse, pseudo-inverse and the conjugate transpose of the matrix $\mathbf{X}$, and $X_{i,j}$ is the entry in the matrix at row i and column j. Vectors are denoted by lower-case bold letters, such as $\mathbf{x}$, with the i-th element x_i, and $\mathrm{x}_{i:I}$ denotes a sub-vector $[x_i, x_{i+1}, \ldots, x_I]^\top$, for $i < I$. The set is denoted using bold calligraphic letters, such as $\mathcal{M}$, where $|\mathcal{M}|$ denotes its cardinality.

This thesis focuses on analysing signals defined on an undirected, fully connected, weighted graph $\mathcal{G}= \{\mathcal{V}, \mathbf{A}\}$, that contains a set of vertices $\mathcal{V}$ with $|\mathcal{V}| = N$, and a weighted adjacency matrix $\mathbf{A}$. The entry of $A_{i,j}$ represents the weight of the edge usually measured with Euclidean distance-based Gaussian kernel weighting function in graph-based semi-supervised learning methods [75], defined as:

$$A_{i,j} = \exp\left(-\frac{\|v_i - v_j\|_2^2}{2\sigma^2}\right), \tag{2.1}$$

where σ is the scaling factor or kernel bandwidth and $\|\cdot\|_2$ is the l-2 norm also known as the Euclidean norm.

The combinatorial graph Laplacian, is defined as $\mathbf{L}{:=}\mathbf{D}-\mathbf{A}$, where the degree matrix $\mathbf{D}$ is a diagonal matrix, given by $D_{i,i} = \Sigma_j A_{i,j}$, and the normalised graph Laplacian matrix is defined as $\mathbf{L}_{\text{norm}} = \mathbf{D}^{(-1/2)}\mathbf{L}\mathbf{D}^{(-1/2)}$. Since $\mathbf{L}$ & $\mathbf{L}_{\text{norm}}$ are real and symmetric matrixes, their complete set of orthonormal eigenvectors, and associated eigenvalues, are usually considered as the frequency information on graph signal [75, 86].

The smoothness property of the data domain intrinsic structure is critical for graph data analysis. Usually, the notion of smoothness to the intrinsic structure of the underlying graph is measured with some discrete differential operators. The discrete p-Dirichlet form of the global smoothness for an undirected graph, with the given signal $\mathbf{s} \in \mathbb{R}^N$ is defined as [75]:

$$G_p(\mathbf{s}) := \frac{1}{p}\sum_{i\in N} \|\nabla_i \mathbf{s}\|_2^p = \frac{1}{p}\sum_{i\in N}\left[\sum_{j\in N} A_{i,j}(s_j - s_i)^2\right]^{\frac{p}{2}}, \tag{2.2}$$

where ∇ is the gradient operator, a signal or function $\mathcal{V} \to \mathbb{R}$ defined on the vertices of the graph could be represented as a vector, namely as graph signal $\mathbf{s} \in \mathbb{R}^N$, where the i-th component of the vector $\mathbf{s}$ represents the function value at the i-th vertex in $\mathcal{V}$. For instance, for a classification problem, graph signal $\mathbf{s}$ can be the catalogue, and the vertex $\mathbf{v}_i$ can be considered as the feature vector of sample i.

This thesis considers two widely used graph smoothness operators named: Graph Total Variation (GTV) minimisation ($p = 1$ in Eq. 2.2), and Graph (normalised) Laplacian Regularisation ((norm)GLR) ($p = 2$ in Eq. 2.2) which have been applied in denoising [79, 87], filtering [88], interpolation [89], *etc.*

Usually, the simplified form of (norm)GLR and GTV are provided as:

$$\text{tr}(\mathbf{HS}), \tag{2.3}$$

where $\text{tr}(\cdot)$ is the trace operator, $\mathbf{S} = \mathbf{s}\mathbf{s}^\top$. Namely, the matrix $\mathbf{H}$ in set $\mathcal{H} =$

$\{\mathbf{L}, \mathbf{L}_{\text{norm}}, \widetilde{\mathbf{A}}\}$, represents the (normalised) combinatorial graph Laplacian matrix for (norm)GLR and $\widetilde{\mathbf{A}} = (\mathbf{I} - \mathbf{A})^{\top}(\mathbf{I} - \mathbf{A})$ for the GTV ($\mathbf{I}$ is the identity matrix) respectively.

An example of a six-node (N= 6) undirected fully connected graph is illustrated in Fig. 2.1, where each $\mathbf{x}_i$ corresponds to a graph node $\mathbf{v}_i \in \mathcal{V}$ in Eq. 2.1 obtained from Iris flower data set representing the feature vector of the i-th sample, containing four features (sepal length/width, and petal length/width in cm). While the graph signal $\mathbf{s} \in \{0, 1, 2\}$ is the iris species Setosa, Versicolour, and Virginica could be considered as the catalogue for classification problem [90].

The thesis only considers GSP on an undirected fully connected graph. As illustrated in Fig. 2.1, the edges are first built as the connecting lines between nodes, then the weight of the edges measured with Eq. 2.1, the thicker edges with higher weights; noteworthy, this thesis does not consider graphs with self-loops; thus, the $A_{i,i} = 0$. With the embedded graph, the piecewise smoothness assumption is the core of classifying the catalogue of the graph node, which is usually measured by discrete differential operators as in Eq. 2.2. Generally, the measurement is the multiplication of the difference between the nodes $(s_j - s_i)$ and the edge weight $A_{i,j}$. Thus, intuitively the measurement is small when the signal $\mathbf{s}$ has similar values at neighboring vertices connected by an edge with a large weight, i.e., when it is smooth. Regarding the classification task, the graph nodes within the same catalogue are connected via high-weight edges, which represent the intrinsic structure of the data domain is smooth. Specifically, in Fig. 2.1, the above fully connected graph with unweight edges is less smooth ($G_2(\mathbf{s}) = 24$ with Graph Laplacian Regularisation and the edges take as $A_{i,j} = 1$) than the one below($G_2(\mathbf{s}) = 0.82$ with Graph Laplacian Regularisation, and the edges $A_{i,j}$ calculated as illustrate in Fig. 2.1), which is connected with weight edges.

GSP provides a convenient tool to address the classification problem for data in the irregular graph domain (unlike regular patterns such as stripes, checks,

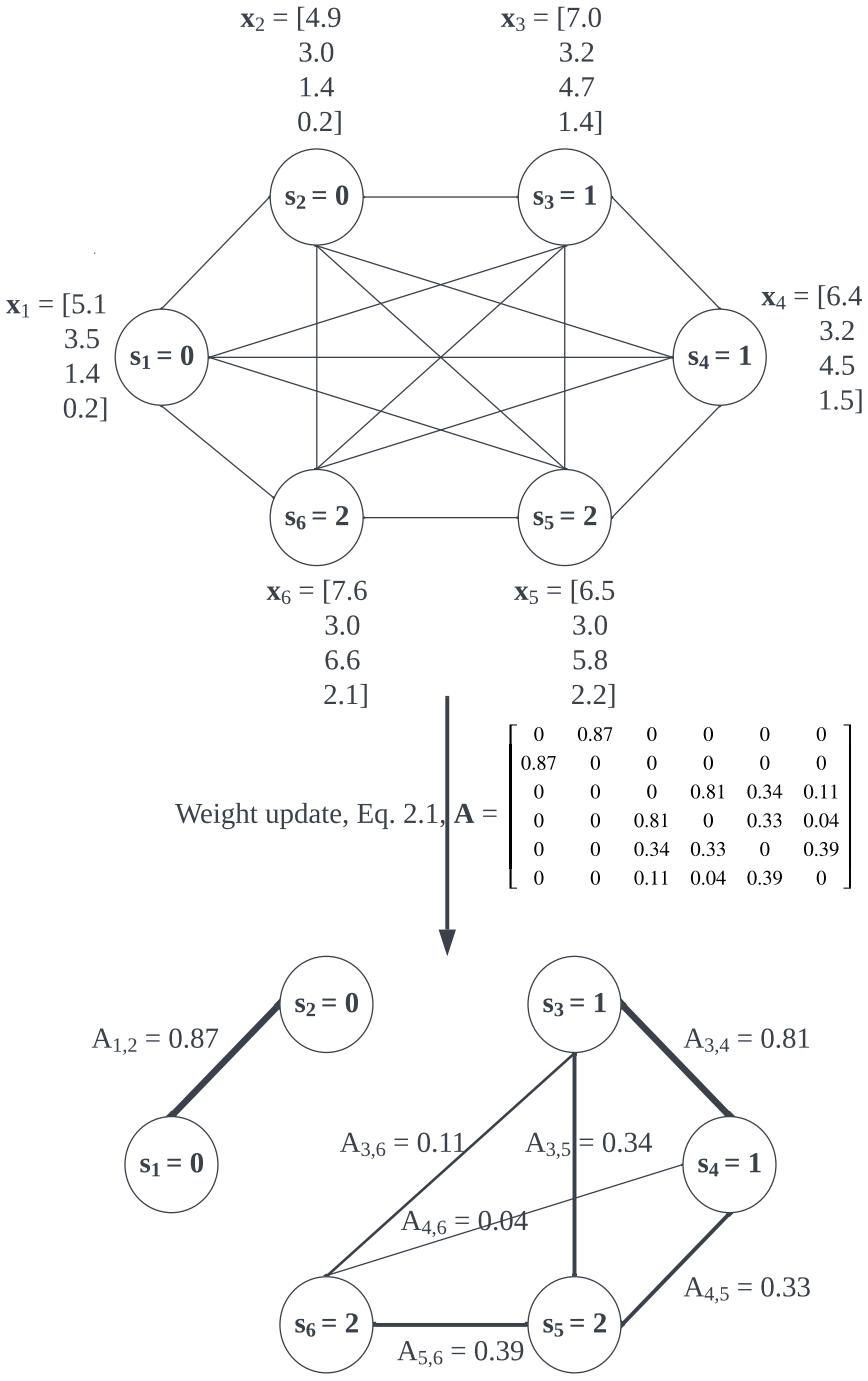


Figure 2.1: Graph example with six nodes

grids, and mosaics in 2D, as well as lattices, crystal structures, and periodic waves in 3D), and regularisation on graphs has emerged recently as a competitive model-based classification and clustering method [74]. The main idea of graph-based classification is to smooth the graph signal $\mathbf{s}$ (usually represents the catalogue) through regularisation via minimising Eq. 4.3 for (norm)GLR or GTV. The edges with high weight connecting the graph nodes (samples) are considered as belonging to the same category. The following chapter will introduce the graph-based classification for (micro)seismic events established on the field-recorded (micro)seismic signal.

Chapter 3

Automated Platform for (Micro)seismic Signal Analysis: Denoising, Detection and Classification in Slope Stability Studies

Chapter 2 introduced the background, notation, and general definition of GSP. Then an illustration (Fig. 2.1) is demonstrated to construct an undirected graph with the Iris flower data set. In this chapter, a GSP-based end-to-end early warning platform is designed to analyse (micro)seismic events. The experimental results evaluated with landslide-induced (micro)seismic datasets SZ10 and PG demonstrate the competitive performance regarding the accuracy and efficiency of the proposed GSP-based end-to-end platform concerning the state-of-the-art approaches. Specifically, due to the data availability, this chapter utilises the single channel recorded data (the vertical channel of S3.0 for SZ10, and the vertical channel of S2.4 for PG as illustrated in Fig. 1 [14]). This chapter is based on the material published in journal papers (J4 & J5) and conference papers (C3 & C4).

3.1 Introduction

Despite the significant difficulties discussed in Chapter 1, in-situ microseismic monitoring has attracted many academics to work on its advancement. Recently, there has been a dramatic increase in the number of studies published on in-situ microseismic monitoring [4, 28].

In general, landslide monitoring techniques include field surveys, unmanned aerial vehicles, aerial platforms, satellite platforms, seismic reflection, and in-situ microseismic monitoring systems [9]. Most of them suffer from high cost, limited spatial coverage, and susceptibility to atmospheric conditions, *etc.*; however, the in-situ microseismic monitoring systems have the advantage of being simple to install and maintain, and the monitoring systems are typically deployed in key locations throughout hazard-prone areas and record the (micro)seismic signal in the distance [9, 10].

With the momentarily released energy, (micro)seismic events are created, consisting of a succession of seismic waves traveling through the crust and may be contaminated by ambient noise. Investigating (micro)seismic events is difficult with low-energy (i.e., ac SNR) events. To explore the (micro)seismic events characteristic, this chapter proposes an automated end-to-end platform to assist the in-situ microseismic monitoring consisting of denoising, detection, and classification, which is efficient for (micro)seismic events processing.

3.2 Related Work

After publishing the research of this chapter in [15], the following works proposed the end-to-end workflow for seismic signal processing in [28, 91, 92]. A fully functional intelligent wireless geophone sensing system is designed in [91] for collection with the objective of real-time data transmission. Since the RF classifier demonstrates its effectiveness in numerous sectors of seismological data

processing, which has been utilised for real-time anthropogenic sources generated seismic events classification with 53 curated attributes [92]. Additionally, a one-step classification strategy with the RF classifier is proposed to separate the slope failure-related (micro)seismic events from the raw signal recordings, and the features retrieved from the continuously recorded data with the sliding window [28]. The challenge and the related work of (micro)seismic events detection, feature engineering, and classification are presented in Chapter 1 Section. 1.2. The remainder of this section presents the review of (micro)seismic denoising, which is also considered crucial for the end-to-end system design.

The low SNR characteristic is known as the primary challenge of (micro)seismic event analysis; due to (sub)surface seismic measurements being affected by: 1) ambient noise, 2) wave propagation-related noise (inc. surface wave and geologic noise), 3) sensor measurement noise and 4) data processing artifacts [93]. Thus, the recorded traces on which the signal of interest can be contaminated by the nonstationary noise raise the difficulty in distinguishing the signal from noise [14, 25]. (Micro)seismic denoising is a crucial step that could affect source location and mechanism determination and hazard assessment, especially in the mining and hydrocarbon settings [25].

The recent denoising approaches for (micro)seismic signal analysis contain 1) multi-channel stacking, 2) spectral analysis, and 3) signal decomposition and reconstruction. The recorded background noise is considered roughly the same order of magnitude as the micro-seismic signal of interest and not biased towards any particular frequency or amplitude and is normally distributed; simply stacking the multi-channel records could mitigate the stationary noise as in [94]; however, the performance is limited by the trace number. Spectral analysis is another commonly utilised denoising approach; for example, wavelet transform is used in [26, 54, 95–97], Short-Time Fourier Transform (STFT)-based denoising proposed in [97]. Besides wavelet, other signal transforms such as curvelet,

dreamlet, and shearlet have been proposed for seismic signal denoising as reviewed in [98]. In General, the recorded time domain signal is first transformed to the spectrum domain; then, various thresholding approaches are involved in minimising the noise amplitude; eventually, the denoised signal is obtained by converting the signal back to the time domain. Spectral analysis-based denoising suffers in efficiency, especially for (near)real-time (micro)seismic monitoring; it cannot discard the noise in which frequency bands overlap with the signal of interest. Signal decomposition is another branch for denoising, such as Ensemble Empirical Mode Decomposition (EEMD) employed in [99, 100]. Except for the denoising approaches above, deep learning is also considered for seismic analysis, in [54, 101]; however, the performance is highly affected by the volume of the training samples and not suitable for the landslide-induced (micro)seismic signal denoising.

The main drawbacks of the current denoising approaches are processing and parameter tuning complexity, assumption of a particular noise distribution, and insufficient evidence of suitability for field data with microseismic and tremor events with insufficient examples to train the models [15].

The efficient landslide-induced (micro)seismic processing problem is presented as a single-channel recorded signal denoising, detection, feature engineering, and classification problem, the proposed GSP-based end-to-end (micro)seismic signal processing in this chapter in the procedure for performing a low-complexity and accurate (micro)seismic signal measurement platform. The anomalous impact in the recorded time series waveform is considered a potentially (micro)seismic event, indexed by the nodes of an undirected graph, where each vertex corresponds to the signal sample for signal denoising or the extracted feature vector from detected events for classification. The weights of the edges connecting the vertices reflect the similarity between the nodes, i.e., the weights of the edges enable 'smoothing' the amplitudes of the connected samples for denoising or 'grouping'

the detected events with small feature distances. Specifically, this chapter utilises the GraphBF to smooth the raw signal for denoising. Concerning detection, the distribution difference between the potential event and the ambient noise is first analysed, and then the event trigger threshold is set up with the Neyman-Pearson lemma. Eventually, the detected signals are classified with the feature selected from the predefined handcrafted features. For the graph-based classifier, this chapter defines an optimisation problem that contains the regularisation term of the GLR as defined Eq. 4.3 in Chapter 2; that is, applying regularisation on the constructed graph signal to minimise the signal variation.

The following sections of this chapter provide a detailed description of the proposed platform with denoising, detection, feature engineering, and classification.

3.3 Methodology

According to the flowchart in Fig. 3.1, the proposed automated early warning system consists of denoising, detection, feature construction and selection, and classification in bold. Generally, the system's input is the recorded time series signal, and eventually, the labelled seismic events are provided. The following subsections describe each of these elements in the proposed flowchart.

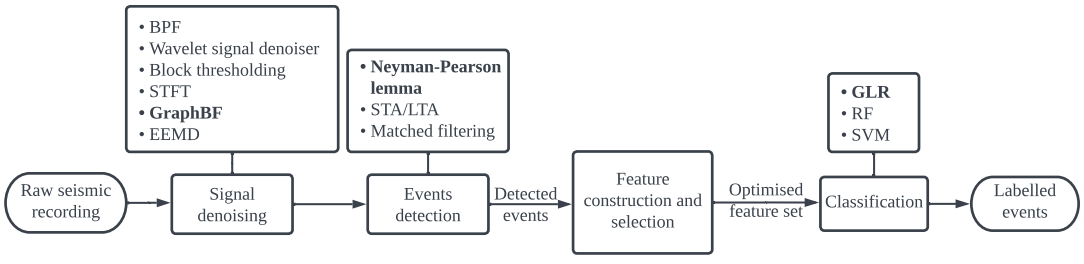


Figure 3.1: System Flow Chart.

3.3.1 Event Denoising

Ambient Noise Distribution

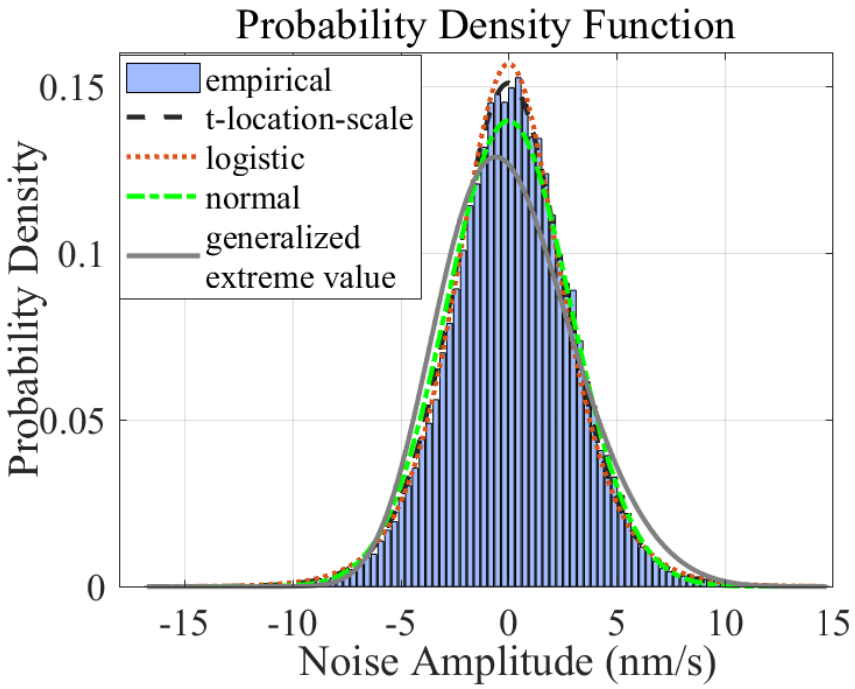


Figure 3.2: Probability density function curve fit

While synthetic seismic noise tends to be modeled as Gaussian in the literature, it has been observed that the noise present in the field dataset (see Section 1.3.2), after BandPass Filtering (BPF), has the best fit to the t-location-scale distribution, as shown in Fig. 3.2, which is more prone to outliers or heavy tails than the normal distribution. Except t-location-scale distribution, this thesis also considers a list of other distributions as in Table 3.1, and the fitness is measured with Bayesian and Akaike information criteria. The t-location-scale distribution is a mixture of normal distributions with gamma mixing weights, and therefore the variance follows the gamma distribution with parameter $\frac{n}{2}$, and n is considered as the degree of freedom. The Probability Density Function (PDF) is given as:

Table 3.1: List of considered distribution

List of considered distribution	
1.	Beta
2.	Birnbaum-Saunders
3.	Exponential
4.	Extreme value
5.	Gamma
6.	Generalized extreme value
7.	Generalized Pareto
8.	Inverse Gaussian
9.	Logistic
10.	Log-logistic
11.	Lognormal
12.	Nakagami
13.	Normal
14.	Rayleigh
15.	Rician
16.	t location-scale
17.	Weibull

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}\left(\frac{n}{n+x^2}\right)^{\frac{n+1}{2}}, \tag{3.1}$$

where $\Gamma(\cdot)$ known as gamma function, defined as: $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$.

Signal Denoising

Since the dominant frequencies of microseismic events are relatively low and expected to fall in the $(1-20\,\mathrm{Hz})$ range [14], the first step is to bandpass filter the raw signals to remove low-frequency and high-frequency components of coherent noise, with the Butterworth filter. However, regarding the spectrum, the background noise and the seismic signal of interest might have some frequency components, standard BPF does not sufficiently denoise the signal, and therefore other denoising approaches are needed [14].

GraphBF denoising: Given a bandpass filtered and sampled signal $\mathbf{x}$, the undirected graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ is constructed, where $\mathcal{V}$ is the set of nodes, each corresponding to one sample in $\mathbf{x}$, and $\mathbf{A}$ is defined as the weighted adjacency

matrix in Eq. 2.1 as described in Chapter 2.

Note that $A_{i,j}$ reflects correlation between samples x_i and x_j of $\mathbf{x}$. Let $\mathbf{D}$ be a diagonal matrix, given by $D_{i,i} = \Sigma_j A_{i,j}$. Then the output of GraphBF, for a noisy input $\mathbf{x}$ is given by Eq. 3.2 [79]:

$$\arg\min_{\mathbf{s}} \frac{1}{2} \left\| \mathbf{s} - \mathbf{x} \right\|_2^2 + \alpha \frac{1}{2} \left\| \mathbf{s} - \mathbf{D^{-1}As} \right\|_2^2. \tag{3.2}$$

Note that the first term of Eq. 3.2 maintains the similarity between the denoised signal $\mathbf{s}$ and the input noisy signal $\mathbf{x}$, the second term is the smoothness prior that tends to minimise the difference between the restored signal and its neighbours, and α is the trade-off factor. This optimisation problem has a closed-form solution [79, 102]:

$$\hat{\mathbf{s}} = (\mathbf{I} + \alpha(\mathbf{I} - \mathbf{D}^{-1}\mathbf{A})^*(\mathbf{I} - \mathbf{D}^{-1}\mathbf{A}))^{-1}\mathbf{x}, \tag{3.3}$$

where $\mathbf{I}$ is the identity matrix. $\hat{\mathbf{s}}$ is the denoised signal that is passed to the event detection block.

The GraphBF closed-form solution involves two parameters: the trade-off factor α, set heuristically to 300, and kernel bandwidth σ, optimised based on Neyman-Pearson Hypothesis Testing [103] set as the detection threshold λ, discussed in Section 3.3.2.

3.3.2 Event Detection

Given two hypotheses $H_0 : f_0(x)$ and $H_1 : f_1(x)$, the Neyman-Pearson lemma decides on the valid hypothesis based on probabilities of detection P_D and false alarm P_{FA} defined as:

$$P_D = \int_{\mathbf{x}} f_1(x)dx, \quad P_{FA} = \int_{\mathbf{x}} f_0(x)dx. \tag{3.4}$$

Specifically, the detection objective is to maximise P_D while keeping P_{FA} within a small boundary θ, heuristically set to 0.01. The objective function can

then be defined as:

$$\max_{\mathbf{x}}(P_D - \mathbf{g}(P_{FA} - \theta)), \tag{3.5}$$

where $\mathbf{g}$ is the Lagrange multiplier, and

$$P_{FA} = \int_{\lambda}^{+\infty} p_a(a|H_0)da, \tag{3.6}$$

where $a \sim \mathcal{N}(0, l\beta^2)$ follows the noise distribution shown in Fig. 3.2 with l and β, representing, respectively, the length of the window and standard deviation of the signal in the window.

Given the PDF of the noise Eq. 3.1, the $p_a(a|H_0)$ and P_{FA} defined as:

$$p_a(a|H_0) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi nl\beta^2}\,\Gamma(\frac{n}{2})}\left(\frac{nl\beta^2}{nl\beta^2 + a^2}\right)^{\frac{n+1}{2}}, \tag{3.7}$$

$$P_{FA} = \int_{\lambda}^{+\infty} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi nl\beta^2}\,\Gamma(\frac{n}{2})}\left(\frac{nl\beta^2}{nl\beta^2 + a^2}\right)^{\frac{n+1}{2}} da, \tag{3.8}$$

$$\lambda = \sqrt{l\beta^2}\mathrm{tinv}(1 - P_{FA}, n), \tag{3.9}$$

where $\mathrm{tinv}(\cdot)$ represents the t inverse cumulative distribution function, and n is the degree of freedom of the t-location-scale distribution, obtained after the PDF curve fitting.

Note that the detector output is the linear detection threshold λ. All signal amplitudes larger than λ are considered events. All detected events shorter than five samples are discarded to exclude detection errors caused by sensor noise or electrical failure. This duration was chosen to include short-duration ($\leq 2\,\mathrm{s}$) slide quake-like signals to a wide variety of longer-duration tremor-like radiations ($\geq 2\,\mathrm{s}$ to several min). In addition, if the time difference between two consecutive detected events is shorter than $0.5\,\mathrm{s}$, these events are merged into a single event. This step is beneficial for correctly detecting Class 3.1, e.g., multiple events (Rock

fall/ near repeater).

For benchmarks, the following adjustable STA/LTA parameters are tuned as per [104]: 1) short-term window length as 1 s and 0.5 s, 2) long-term window length as 50 s, 3) STA/LTA trigger threshold level set to 2, and 4) STA/LTA de-trigger threshold level set to 0.8. For the matched filtering, the trigger threshold is set as five times the standard deviation value of the cross-correlation, as suggested in paper [32]. With simple thresholding and cross-correlation, the matched filtering is relatively fast.

3.3.3 Feature Construction, Weighting, and Selection

Since the wave amplitude is highly affected by wave propagation attenuation and lack of event source locations as the benchmark, the filtered signal is preprocessed with Max-Min normalisation before feature construction.

The signals are initially described with $K = 99$ (micro)seismic features (Table 3.2) that are appropriate for characterising the signal of interest in terms of the temporal, spectral, and cepstral domain. Some of these traits have been described in previous studies [4,14,20,28,105], and they are commonly used in signal categorisation. The features are measured on the raw signal window without human interpretation.

The number of features that uniquely characterise each event type is reduced through feature selection, in contrast to [12], which employed the full feature set for classification. Feature selection reduces complexity and enhances the classification accuracy by preventing [106] from becoming overfit. To find the optimal identifying characteristics of the event categories, the machine learning feature selection MATLAB Toolbox: FSLib is implemented as Table 3.3 [107].

The subset of features that maximise classification accuracy result with five-fold cross-validation was selected. The feature selection method also returns a soft value score for each feature k, indicating the usefulness of the feature, which,

after normalisation, is used in the proposed GLR classifier, as e_k, described next.

Table 3.2: List of features drawn from the literature: temporal $\mathbf{s}(t)$, power signal $\mathbf{p}(t)$, envelope $\mathbf{e}(t)$, auto correlation function $\mathbf{ac}(t)$, spectral $\mathbf{f}(v)$, cepstrum domain $\mathbf{ce}(v)$ and envelope $\mathbf{ss}(t)$, temporal $\mathbf{es}(t)$ and spectral $\mathbf{fs}(v)$ with $(1-5\,\text{Hz})$, $(5-9\,\text{Hz})$, $(9-13\,\text{Hz})$, $(13-17\,\text{Hz})$ and $(17-20\,\text{Hz})$ passband. **PMF** refers to Probability Mass Function, **PSD** is the power spectral density of $\mathbf{f}(v)$.

Parameter	Description
Temporal Feature	
T1	Duration [108] [4] [48] [109]
T2-4	Standard deviation (STD), Mean, Median of $\mathbf{s}(t)$ [48] [14] [109] [110] [111]
T5-7	Max, Mean, Median of $\mathbf{e}(t)$ [4] [14] [110]
T8	Rising duration $\mathbf{s}(t)$ [109]
T9	Decreasing duration $\mathbf{s}(t)$ [109]
T10	Entropy of $\mathbf{s}(t)$ $-\sum \mathbf{PMF}(\mathbf{s}(t)) \log \mathbf{PMF}(\mathbf{s}(t))$ [12] [112] [113] [111]
T11	Zero Cross Rate of $\mathbf{s}(t)$ [14] [111]
T12	STD at the decreasing part of $\mathbf{s}(t)$ $\sqrt{\frac{1}{(N-t_{max}-1)}\Sigma_{t_i=t_{max}}^{N}\Big(\mathbf{s}(t_i)-\mu\Big)^2}$ [4]
T13-15	Skewness of $\mathbf{s}(t)$, $\mathbf{p}(t)$ & $\mathbf{e}(t)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{x}(t_i)-\mu}{\text{STD}(\mathbf{x})}\right)^3$ [108] [4] [14] [109] [111]
T16-18	Kurtosis of $\mathbf{s}(t)$, $\mathbf{p}(t)$ & $\mathbf{e}(t)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{x}(t_i)-\mu}{\text{STD}(\mathbf{x})}\right)^4$ [108] [4] [14] [109] [111]
T19&20	Rate of attack of $\mathbf{s}(t)$ & $\mathbf{e}(t)$ [12]
T21&22	Rate of decay of $\mathbf{s}(t)$ & $\mathbf{e}(t)$ [12]
T23	Ratio of Max and mean of $\mathbf{e}(t)$ $\frac{T_5}{T_6}$ [109]
T24	Ratio of Max and Median of $\mathbf{e}(t)$ $\frac{T_5}{T_7}$ [109]
T25	Ratio of Max and STD of $\mathbf{e}(t)$ $\frac{T_5}{STD(\mathbf{e}(t))}$ [109]
T26	Ratio of (tmax)/(N-tmax) $\frac{t_{max}}{N-t_{max}}$ [4] [114] [109]
T27	Ratio of T18 and T20 $\frac{T_{18}}{T_{20}}$ [4]
T28-T32	Energy of $\mathbf{es}(t)$ $\sum_{i=1}^{N}\mathbf{es}(t_i)^2$ [4] [109]
T33-T37	Average power of $\mathbf{ss}(t)$ $\frac{1}{N}\sum_{i=1}^{N}\mathbf{ss}(t_i)^2$ [14]
T38&39	Energy of (1:N/3) and (N/3:N) of $\mathbf{ac}(t)$ $\sum_{i=1}^{\frac{N}{3}}\mathbf{ac}(t_i)^2 \quad \sum_{i=\frac{N}{3}}^{N}\mathbf{ac}(t_i)^2$ [4]
T40	Int-ratio of $\mathbf{ac}(t)$ $\frac{T_{38}}{T_{39}}$ [4] [109]
T41	Num of peaks of $\mathbf{ac}(t)$ [4] [109]
T42	duration of $\mathbf{ac}(t)$ $\max_t(\mathbf{ac}(t) < 0.2\max(\mathbf{ac}(t)))/(T_1)$ [114]
T43	Measure of location $\sum_{i=1}^{N} i\mathbf{p}(t_i)$ [113]
T44	Measure of dispersion $\sqrt{\sum_{i=1}^{N}(i-T_{43})^2\mathbf{p}(t_i)}$ [113]
T45	Measure of asymmetry $\frac{1}{T_{44}^3}\sum_{i=1}^{N}(i-T_{43})^3\mathbf{p}(t_i)$ [113]
T46	Measure of concentration around single value $\frac{1}{T_{44}^4}\sum_{i=1}^{N}(i-T_{43})^4\mathbf{p}(t_i)$ [113]

Table 3.2: *Cont.*

Parameter	Description
Spectral Feature	
F1	Absolate value of mean $\mathbf{f}(v)$ [108] [4] [109]
F2	Absolate value of max $\mathbf{f}(v)$ [4] [109] [110]
F3	Absolate value of median $\mathbf{f}(v)$ [4] [109]
F4	Variance of $\mathbf{f}(v)$ [4] [109] [14] [110] [111]
F5	Max envelop of $\mathbf{f}(v)$ $max(\mathbf{e}(\mathbf{f}(v))$ [114] [110]
F6	Num of peaks >0.75 bandwidth $\mathbf{f}(v)$ [4] [109]
F7	Dominate frequency [4] [108] [110]
F8	Spectral centroid [4] [109] [113] [110] [111]
F9	Int-ratio of $\mathbf{f}(v)$ $\frac{\sum_{i=1}^{\frac{N}{3}} \mathbf{f}(v_i)^2}{\sum_{i=\frac{N}{3}}^{N} \mathbf{f}(v_i)^2}$ [112]
F10-14	Kurtosis $\mathbf{fs}(v)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{fs}(v_i)-\mu}{\mathrm{STD}(\mathbf{x})}\right)^4$ [4] [109]
F15	Num of peaks $\mathbf{f}(v)$ [4]
F16-20	Energy of $\mathbf{fs}(v)$ $\sum_{i=1}^{N} \mathbf{fs}(v_i)^2$ [110] [109]
F21	Gamma 1 $\frac{\sum_{i=1}^{N} v_i \mathbf{f}(v_i)^2}{\sum_{i=1}^{N} \mathbf{f}(v_i)^2}$ [109]
F22	Gamma 2 $\sqrt{\frac{\sum_{i=1}^{N} v_i^2 \mathbf{f}(v_i)^2}{\sum_{i=1}^{N} \mathbf{f}(v_i)^2}}$ [109]
F23	Gamma 3 $\sqrt{\mid F21^2 - F22^2 \mid}$ [109]
F24	Mean frequency $\frac{\sum_{i=1}^{N} \mathbf{PSD}(v_i) v_i}{\sum_{i=1}^{N} \mathbf{PSD}(v_i)}$ [114]
F25	Frequency bandwidth $2\sqrt{\frac{\sum_{i=1}^{N} \mathbf{PSD}(v_i) {v_i}^2}{\sum_{i=1}^{N} \mathbf{PSD}(v_i)} - F_{24}^2}$ [114]
F26	Minimal frequency $\min_v(\mathbf{PSD}(v) < 0.2max(\mathbf{PSD}(v)))$ [114]
F27	Maximal frequency $\max_v(\mathbf{PSD}(v) < 0.2max(\mathbf{PSD}(v)))$ [114]
F28	Gyration radius $\sqrt{\frac{m_3}{m_2}}$ m_i is the ith moment [4]
F29	Spectral centroid width $\sqrt{{F_8}^2 - {F_{28}}^2}$ [4] [111]
Cepstrum Feature	
C1	STD $\mathbf{ce}(v)$ STD($\mathbf{ce}(v)$) [12]
C2	Skewness of $\mathbf{ce}(v)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{ce}(v_i)-\mu}{\mathrm{STD}(\mathbf{x})}\right)^3$ [12]
C3	Kurtosis of $\mathbf{ce}(v)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{ce}(v_i)-\mu}{\mathrm{STD}(\mathbf{x})}\right)^4$ [12]
C4	Max value of $\mathbf{ce}(v)$ $max(\mathbf{ce}(v))$ [12]
C5-14	First 10 $\mathbf{ce}(v)$ [12] [111]
Acoustics Feature(Linear prediction filter coefficient)	
A1-10	10 Linear Prediction Coefficients [114]

Table 3.3: Feature Selection methods

Feature Selection Methods	
1.	Infinite Latent Feature Selection
2.	Infinite Feature Selection (Inf-FS)
3.	Eigenvector Centrality Feature Selection
4.	Minimum Redundancy Maximum Relevance
5.	Relieff-based feature selection
6.	Mutual Information-based feature selection
7.	Feature Selection via Concave Minimisation (FSV)
8.	Laplacian Score
9.	Recursive Feature Elimination
10.	Fisher/Correlation Score
11.	Unsupervised Discriminative Feature Selection
12.	Local Learning-Based Clustering
13.	Feature Selection with Adaptive Structure Learning
14.	Dependence Guided Unsupervised Feature Selection
15.	Unsupervised Feature Selection with Ordinal Locality
16.	Loss Object - for calculating loss functions

3.3.4 Event Classification

To classify the target signals, the GLR-based classifier as discussed in Chapter 2 is utilised, as Eq. 4.3, the matrix $\mathbf{H}$ represents $\mathbf{L}$ for GLR. Specifically, the GLR-based classification known as semi-supervised learning can be defined as minimises $\mathbf{s}^\top \mathbf{L}\mathbf{s}$, i.e., one that finds the smoothest graph signal.

For the classification task, an undirected graph is designed, $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is the set of nodes and $\mathbf{A}$ is defined as the weighted adjacency matrix, with $A_{i,j}$ being the weight of the edge connecting nodes i and j. Each event identified by the event detection approach is assigned to a node in $\mathcal{V}$. Let N be the total number of events/nodes. The weight $A_{i,j}$ reflects the correlation between the nodes i and j (that is, between the features constructed and selected in event i and j). Following [80]; the Gaussian kernel function is set as:

$$A_{i,j} = \exp\Big\{ -\sum_{k=1}^{K} e_k \frac{(f_k(i) - f_k(j))^2}{2\sigma^2} \Big\}, \tag{3.10}$$

where K is the number of features after the feature selection, $f_k(i)$ is the value of the k-th feature of event i, σ is the kernel bandwidth, which is set as 1 in this chapter and e_k is the weight of k-th feature obtained from the soft score of the feature selection methods as mentioned before. Then, the combinatorial Laplacian matrix of the graph can be calculated as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal matrix, defined in Section 3.3.1.

Let n be the number of training events. The definition of graph signal s_i for event i as:

$$s_i = \begin{cases} +1, & \text{if Event } i \text{ belongs to the Class and } i \leq n \\ -1, & \text{otherwise and } i \leq n \\ 0, & \text{for } n < i \leq N. \end{cases} \tag{3.11}$$

Signal classification with GLR is established on the assumption that the nodes that belong to the same class are connected with high-weighted edges; hence, if the set of classification labels is considered a graph signal, this signal should change smoothly across the graph. By finding the smoothest signal given constraints in Eq. 3.11, the unknown class labels (initially set to 0) are estimated. That is, the quadratic form of the global smoothness of the graph is minimised, often referred to as GLR term [75], which has a closed-form solution:

$$\begin{aligned} \min_{\mathbf{s}_{n+1:N}} \mathbf{s}^\top \mathbf{L} \mathbf{s} &= \underbrace{\begin{bmatrix} \mathbf{s}_{1:n}{}^\top & \mathbf{s}_{n+1:N}{}^\top \end{bmatrix}}_{\mathbf{s}^\top} \underbrace{\begin{bmatrix} \mathbf{L}_{1:n,1:n} & \mathbf{L}_{1:n,n+1:N} \\ \mathbf{L}_{n+1:N,1:n} & \mathbf{L}_{n+1:N,n+1:N} \end{bmatrix}}_{\mathbf{L}} \underbrace{\begin{bmatrix} \mathbf{s}_{1:n} \\ \mathbf{s}_{n+1:N} \end{bmatrix}}_{\mathbf{s}}, \\ \widetilde{\mathbf{s}}_{n+1:N} &= \mathbf{L}^{\#}_{n+1:N,n+1:N} \left(-\mathbf{s}_{1:n}^\top \mathbf{L}_{1:n,n+1:N} \right)^\top, \end{aligned} \tag{3.12}$$

where $^{\#}$ denotes pseudo-inverse matrix.

As in [115], once $\widetilde{\mathbf{s}}_{n+1:N}$ is calculated as above, for testing data, it is set to $+1$ if $\widetilde{\mathbf{s}}_{n+1:N} > 0$, or -1, otherwise.

3.4 Results and Discussion

This section first presents the metric utilised to evaluate the performance of the proposed workflow, followed by the results of each block in Fig. 3.1.

3.4.1 Metric

To evaluate the proposed end-to-end platform, the metrics that are commonly used for machine learning classification tasks are introduced; namely, precision, sensitivity or recall, and F1 score, defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Sensitivity or Recall} = \frac{TP}{TP + FN}, \tag{3.13}$$

$$\text{Accuracy}(Acc) = \frac{TP + TN}{TP + FP + TN + FN}, \quad \text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{3.14}$$

where TP, FP, FN, and TN are explained in Table 3.4, regards to detection and classification tasks. To evaluate the obtained results, the catalogue provided by authors [14] is considered as ground truth.

Table 3.4: Metric terminology

	Detection	Classification
True Positive (TP)	No.of correctly detected events	No.of positive class predictions that actually belong to the positive class
False Positive (FP)	No.of detected unidentified events	No.of positive class predictions that actually belong to the negative class
False Negative (FN)	No.of missed identified events	No. of negative class predictions that acutally belong to the positive class
True Negative (TN)		No. of negative class predictions that acutally belong to the negative class

3.4.2 Denoising and Event Detection Results

Signal Denoising Results

To test the denoising approach, five state-of-the-art denoising methods on the bandpass filtered signal have been implemented: wavelet-based denoising via Matlab toolbox, Block threshold [26], STFT [97], EEMD [99], and GraphBF [116]. Noting the bandpassed filtering denoising approach is considered the baseline in this section.

To provide a statistical measure of the denoising ability of the denoising approaches, the SNR is measured as in [117]:

$$\text{SNR} = \sqrt{\sum_{i=1}^{l}(x_i)^2} / \sqrt{\sum_{i=1}^{l}(no_i)^2}, \tag{3.15}$$

where l is the length of the window, x_i is the measured signal during the event, while no_i is the measured signal with ambient noise present only (no event).

Table 3.5: Denoising result (SNR dB)

Denoising method	class 1.1	class 1.2	class 2.2	class 2.3	class 2.4	class 3.1	class 3.2	class 3.4	class 9	average
BPF	2.68	3.24	4.54	5.96	9.57	9.81	7.23	5.02	35.13	9.24
Wavelet Signal Denoiser	12.34	7.79	17.77	23.72	15.93	30.18	12.49	14.54	37.21	19.11
Block thresholding [26]	9.18	9.55	14.48	18.24	22.44	24.13	18.73	13.92	39.94	18.96
STFT [97]	10.04	7.29	**26.29**	11.47	12.65	14.97	14.15	9.61	35.33	15.76
EEMD [99]	9.13	6.45	-2.68	8.05	9.34	13.22	8.01	12.87	35.13	11.06
GraphBF [116]	**13.45**	**12.33**	19.13	**27.06**	**24.27**	**30.76**	**25.24**	**22.39**	**47.59**	**24.69**

The bold content represents the proposed method and the denoising result with the highest SNR.

Table 3.5 presents the robustness of the proposed GraphBF and state-of-the-art benchmark denoising approaches, where it can be seen that the GraphBF outperforms all other methods for denoising all types of events. As expected, this is closely followed by the best-known wavelet denoising approach. The remaining

approaches performed poorly.

Additionally, this section objectively and quantitatively compares the denoising approaches for each class with Fig. 3.3. As observed in Fig. 3.3, wavelet denoising with Matlab toolbox, Block threshold denoising, and GraphBF visually show competitively good performance with consistency for all the classes, while EEMD denoising over-smooths the signal for the majority of classes and therefore the least favoured approach.

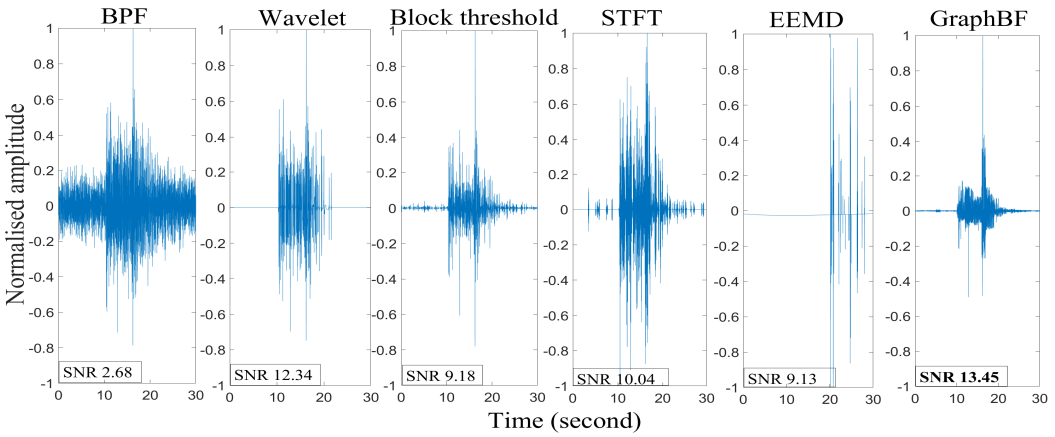
Signal Detection Results

Table 3.6: Event detection results for the SZ10 dataset.

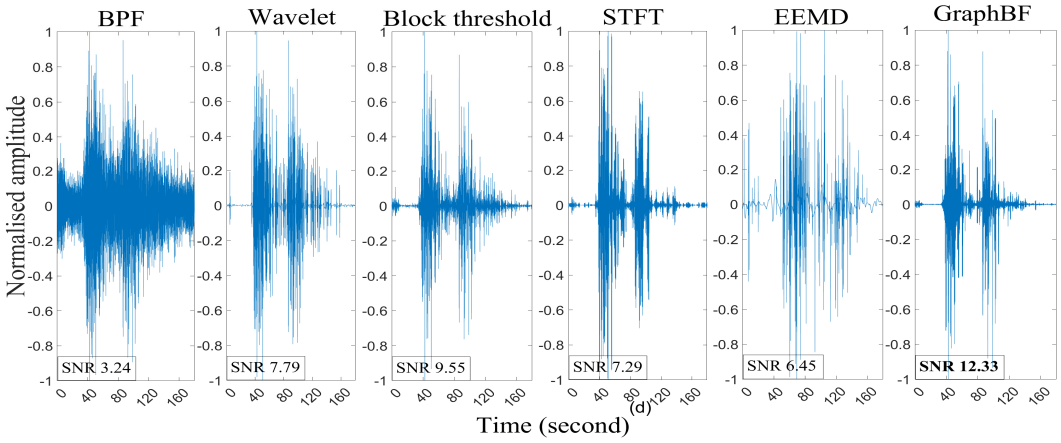
Detection method	TP	FP	Recall	Precision	F1 score
BPF					
Neyman-Pearson lemma	174	69	1	0.72	0.83
STA/LTA	157	45	0.9	0.78	0.83
Matched filtering [32]	169	91	0.97	0.65	0.78
BPF + GraphBF					
Neyman-Pearson lemma	**174**	**45**	**1**	**0.79**	**0.89**
STA/LTA	165	71	0.95	0.70	0.80
Matched filtering [32]	170	52	0.98	0.77	0.86
BPF + Wavelet Signal Denoiser					
Neyman-Pearson lemma	170	78	0.98	0.69	0.81
STA/LTA	171	339	0.98	0.34	0.50
Matched filtering [32]	172	84	0.99	0.67	0.80

The bold content represents the proposed methods.

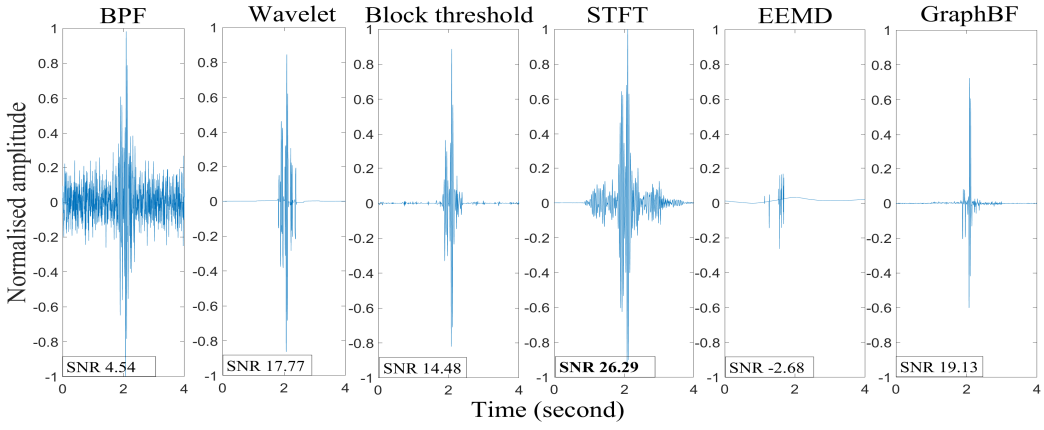
The performance of the proposed detection via Neyman-Pearson lemma and benchmark (STA/LTA and Matched filtering) detection approaches for the case of BPF only, without additional denoising, and two cases of BPF with the GraphBF and Wavelet Signal Denoiser, are shown in Table 3.6 and 3.7, for the SZ10 and PG datasets, respectively. As discussed in Section 1.2, the aim is to detect all labelled events without missing any (i.e., a recall of 1) and minimise the number of false positives (the more significant the precision, the better).



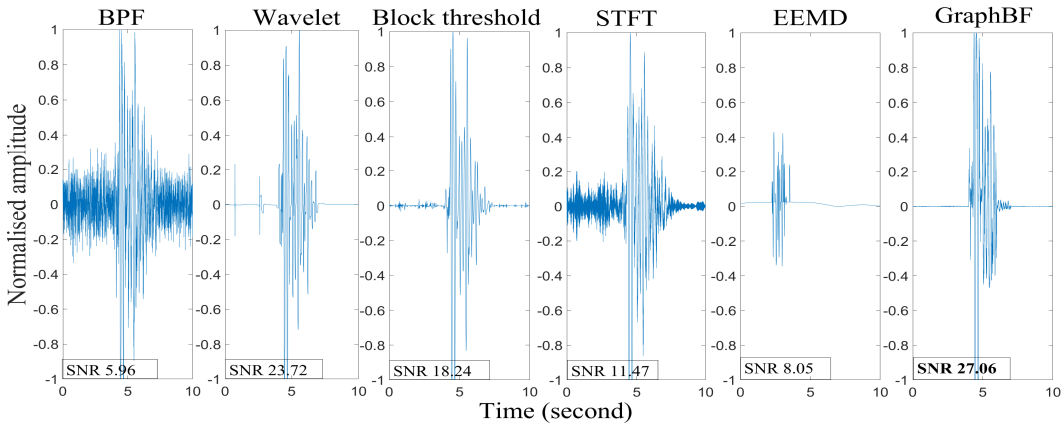
(a) local earthquake (class 1.1)



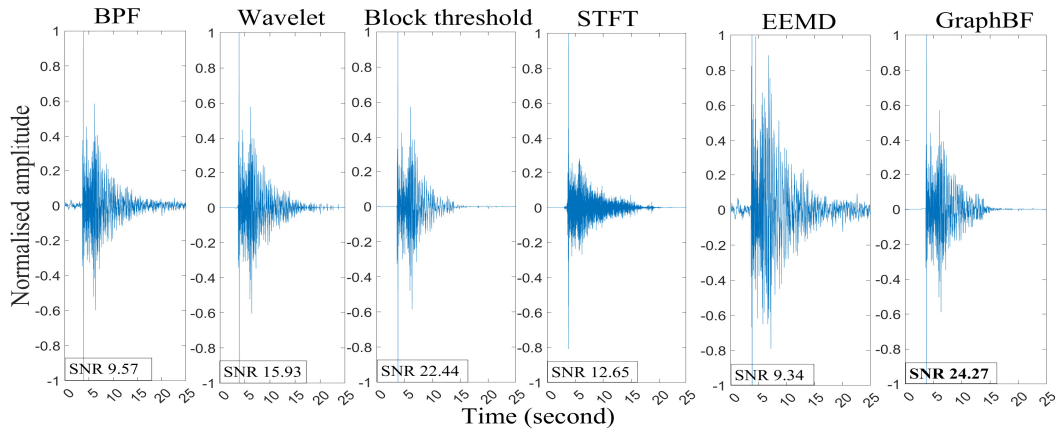
(b) Distant earthquake (class 1.2)



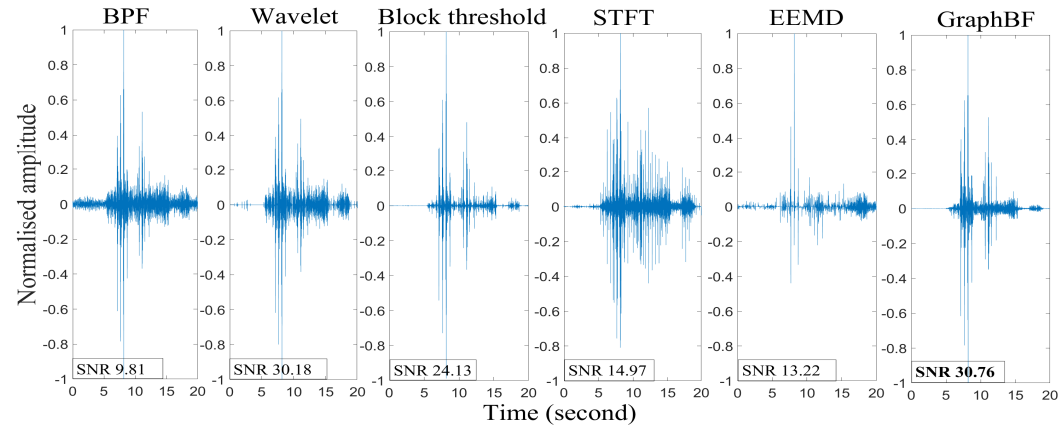
(c) Distant slide quake (class 2.2)



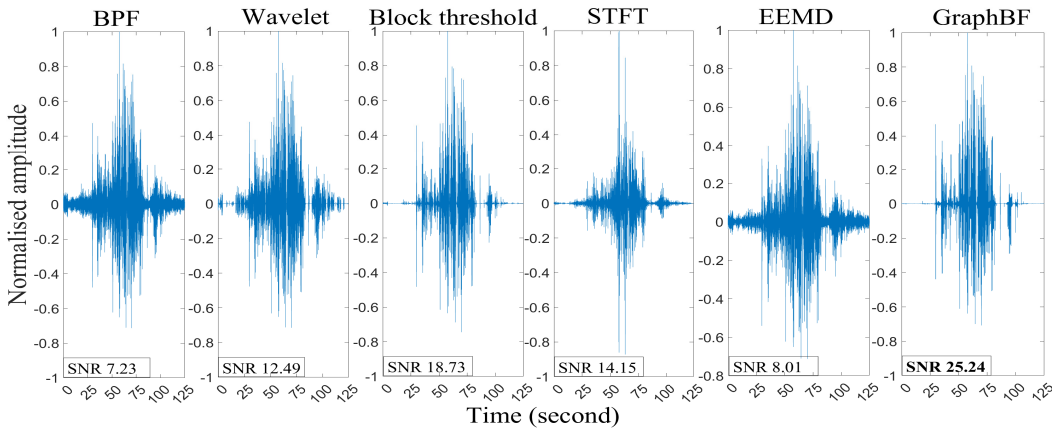
(d) Nearfield low frequency slide quake(class 2.3)



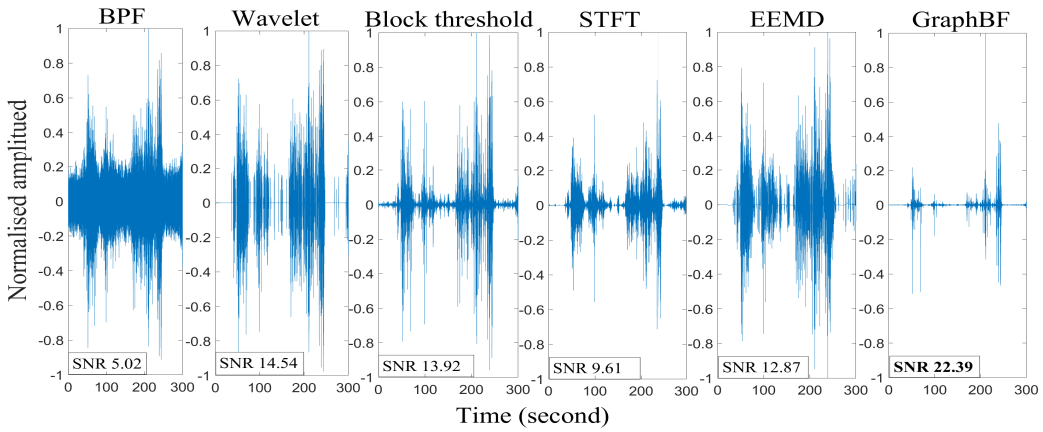
(e) Nearfield microearthquake (class 2.4)



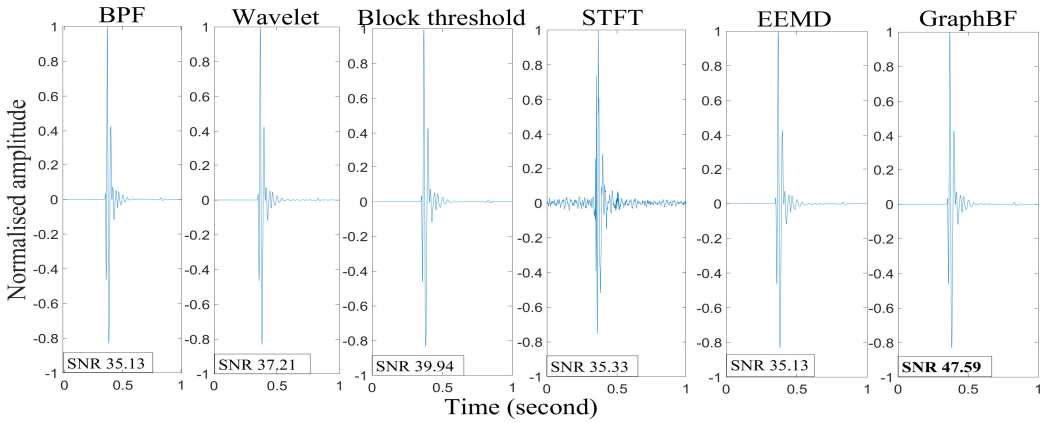
(f) Multiple event (class 3.1)



(g) Dispersive tremor (class 3.2)



(h) Undefined sinusoidal event (class 3.4)



(i) Calibration shot (class 9)

Figure 3.3: Visual denoising result for all target classes

Table 3.7: Event detection results for the PG dataset.

Detection method	TP	FP	Recall	Precision	F1 score
BPF					
Neyman-Pearson lemma	42	10	1	0.81	0.89
STA/LTA	36	35	0.86	0.51	0.64
Matched filtering [32]	40	21	0.95	0.66	0.78
BPF + GraphBF					
Neyman-Pearson lemma	**42**	**5**	**1**	**0.89**	**0.94**
STA/LTA	38	44	0.90	0.46	0.61
Matched filtering [32]	40	7	0.95	0.85	0.90
BPF + Wavelet Signal Denoiser					
Neyman-Pearson lemma	42	48	1	0.47	0.64
STA/LTA	41	56	0.98	0.42	0.59
Matched filtering [32]	40	34	0.95	0.54	0.69

The bold content represents the proposed methods.

As expected, the improvement due to additional denoising, besides standard BPF, can be observed through the consistent increase in the number of events detected (TPs and FPs) and improvement of the recall value for all three detection approaches. However, the increase in false alarms (FPs) is mitigated most effectively with the proposed GraphBF denoising approach in conjunction with the detection via the Neyman-Pearson lemma, as evidenced by the improved precision value. In general, the proposed detection via the Neyman-Pearson lemma has the best performance, closely followed by matched filtering. The STA/LTA provides lower detection accuracy, irrespective of the denoising method, because it tends to detect too many FPs.

Since GraphBF denoising, together with the proposed detection via Neyman-Pearson lemma, provides the highest detection accuracy for both datasets, without missing an event, they are therefore used for the next step: feature construction, selection, and event classification.

3.4.3 Feature Construction and Selection Results

Since different classes or event types are characterised by temporal, spectral, and cepstral features, selection techniques of Table 3.3 were implemented to narrow down the most pertinent or unique features for each class from all 99 feature candidates of Table 4.4, and using GLR classifier with 50% split training and testing data with five-fold cross-validation on the training data only to tune the parameters, resulting in Table 3.8.

Table 3.8: Feature selection result. Column 2 shows the selected features, notation as in Table 3.2, for each event type.

Type of event	Selected Features
Class1.1-Local earthquake	T44,T43,T27,T40,T24,F5,T22,F10,T13,F2 A8,T20,A9,T9,F11,F16,F13,T21,A7,T15 T12,T33,T2,T14,F14,T6,T1,T10,T23
Class1.2-Distant earthquake (Teleseismic)	C3,F3,F15,F6,F21,F24,T1,T9
Class2.2-Distant slide quake (Moderate distant quake)	F11,C4,T22,C6,T17,T24,T21,T12,T19,T11 T20,T25,T37,C11,A6
Class2.3-Nearfield low frequency slide quake	C8,T21,F28,C14,C10,A8,T3,T25,C2,T34
Class2.4-Nearfield microearthquake	C8,T14,T13,T12
Class3.1-Multiple event (Rock fall/near repeater)	T21, F7,F4,F3,T22,C11,C7,C4,A2,T12 F13,T11,F23,F25,T25,A1,T10,A3
Class3.2-Dispersive tremor	F27,T2,T20,T34,T8,F3,C5,F28,C1,C14 T10,F21,T42,C12,T6,A7,T50,F23,T5,T12 F7,C2,T7,C13,T25,F24,T38
Class3.4-Undefined sinusoidal event (Harmonic)	T1,F29,F8,T9,T8,F28,T35,T41,T12,T51
Class9-Calibration shots	F25,T38,C6,C10,F26,T25

For example, as mentioned in Section 1.3.2, Class 1.2, distant earthquake signal, has a low, dominant frequency; hence the feature selection method highly ranks feature F3 median value of the Discrete Fourier Transform (DFT) coefficient, which is often used in seismic analysis to estimate the dominant frequency of the seismic signal. The dominant frequency of a seismic wave is relevant to the energy emitted by the source. Small-magnitude seismic events have high fre-

quencies and higher-magnitude events have lower frequencies. The magnitude of a seismic event reflects the released energy by the seismic source. Large magnitude earthquakes require large faults, while small magnitude events can happen to smaller structural geological features [14]. Class 2.2, distant slide quake, has a consistent beginning, i.e., characterised by the following two features: rate of attack T19 and decay T20. Classes 2.3 and 3.1 have low-frequency content and short duration, which leads to the waveform-based feature T21 being highly ranked. Class 2.4, nearfield microearthquake, can be identified with distinguishable P and S-phases; hence the signal waveform-based features, features T12, T13, and T14, are highly ranked. Class 9 calibration shot has a unique waveform; therefore, waveform-based features T25 and T38 are picked by the feature selection method.

3.4.4 Classification Results

The classes are imbalanced distributed as shown in Table 1.1, which would have a negative impact on training the classification models. Indeed, some classes, such as distant earthquakes, nearfield microearthquakes, and dispersive tremors, have insufficient data to train the model. To overcome this, Synthetic Minority Over-Sampling Technique (SMOTE) is utilised to generate synthetic data directly in the feature domain [118]. SMOTE preprocessing is a standard tool for mitigating imbalanced data model learning by first confirming the oversampling number E, obtained with uniform class distribution, then selecting the instance of the minority class and the F nearest neighbors, and finally with E of the F instances, new samples are generated by interpolation [119]. The synthetic features are developed based on the features constructed described in Section 3.3.3 and only used for training the classifiers.

This chapter takes the one-against-all classification strategy with five-fold

cross-validation. Specifically, starting with Class 1.1, split the data into Class 1.1 events and non-Class 1.1 events. To distinguish Class 1.1 from the rest, the training data contains real non-Class 1.1 data with (different ratios from 10% to 70%) and an equal amount of synthetic Class 1.1 data. This way, the 'binary' classifier will be balanced. For example: after denoising with GraphBF and detection via Neyman-Pearson lemma, 219 events (174 actual events + 45 false alarms) are detected for the SZ10 dataset with 50 Class 1.1 events and 169 non-Class 1.1 events. Thus, 118 (i.e., 70% of 169) detected non-Class 1.1 events are utilised for training, and 118×99 synthetic features are generated for Class 1.1. The remaining $169 - 118 = 30\%$ of Non-Class 1.1 events, together with all detected 50 Class 1.1 events, are utilised for testing. Note that for testing, the events detected by the proposed detection approach are utilised, which also contain false positives.

The classification results for the SZ10 and PG datasets are shown in Table 3.9 and 3.10, respectively. Three classification methods are compared: the proposed GLR-based classifier, SVM, and RF classifiers, as described in Section 1.2 and 3.3.4. All three classifiers are fed with the same events detected by the proposed GraphBF+detection via Neyman-Pearson lemma. It can be seen that the GLR-based classifier is, for most classes, the most accurate classifier. This is expected since there is a high intra-class variability that GLR handles well. Table 3.9 shows that only Classes 2.2 and 3.1 have a relatively low F1 score for all three classifiers. The poor performance of Class 2.2 is explained by its small duration and false alarms during detection. Class 3.1 belongs to tremor-like signals, which are usually observed at strike-slip faults [14], having similar characteristics with ambient noise [14], and hence more difficult to distinguish.

The results for the PG dataset, as in Table 3.10, are worse for Classes 1.1, 3.1, and 3.4. This is expected since the PG dataset is smaller and contains recordings from two different monitoring periods (9 days in 2015 and 6 days in 2016), which

Table 3.9: SZ10 classification result (F1 score)

Class	training ratio	1.1	1.2	2.2	2.3	2.4	3.1	3.2	3.4	9	Mean
GLR	10%	0.53	1.00	0.58	0.87	0.39	0.43	0.71	0.75	0.62	**0.65**
	20%	0.68	1.00	0.75	0.80	0.93	0.50	1.00	0.86	0.62	0.79
	30%	0.84	1.00	0.82	1.00	0.93	0.58	1.00	0.89	0.85	**0.88**
	40%	0.88	1.00	0.79	1.00	0.93	0.72	0.92	0.97	0.93	**0.91**
	50%	0.85	1.00	0.93	1.00	1.00	1.00	0.82	1.00	1.00	**0.96**
	60%	0.88	1.00	0.80	0.92	0.93	1.00	0.93	0.79	0.93	**0.91**
	70%	0.95	1.00	0.93	1.00	0.93	0.83	1.00	0.93	0.93	**0.95**
SVM	10%	0.56	0.60	0.57	0.67	0.44	0.21	0.60	0.55	0.60	0.53
	20%	0.74	1.00	0.75	0.67	0.93	0.75	0.93	0.82	0.68	**0.81**
	30%	0.83	1.00	0.80	0.83	1.00	0.93	0.82	0.71	0.71	0.85
	40%	0.87	1.00	0.76	0.89	1.00	0.82	0.88	0.82	0.82	0.87
	50%	0.87	1.00	0.90	0.80	0.88	0.83	0.82	0.88	1.00	0.89
	60%	0.91	0.92	0.80	0.92	0.93	0.97	0.93	0.81	0.90	0.90
	70%	0.93	1.00	0.93	0.96	0.93	0.86	1.00	0.97	0.90	0.94
RF	10%	0.70	0.71	0.69	0.32	0.56	0.61	0.48	0.48	0.60	0.57
	20%	0.84	0.91	0.72	0.49	0.70	0.60	0.52	0.86	0.50	0.68
	30%	0.84	0.80	0.83	0.76	0.78	0.65	0.56	0.72	0.77	0.75
	40%	0.92	1.00	0.74	0.69	0.93	0.70	0.64	0.80	0.82	0.80
	50%	0.90	1.00	0.80	0.73	0.82	0.93	0.80	0.85	0.83	0.85
	60%	0.95	0.75	0.78	0.96	0.93	0.94	0.78	0.90	0.85	0.87
	70%	0.95	1.00	0.88	0.96	1.00	0.94	0.86	0.97	0.90	0.94

The bold content represents the highest F1 score for each training ratio.

Table 3.10: PG classification result (F1 score)

Class	training ratio	1.1	1.2	2.2	3.1	3.4	9	Mean
GLR	40%	0.88	0.94	1.00	0.91	1.00	0.91	**0.94**
	50%	0.53	0.88	0.86	1.00	0.86	0.67	0.80
	60%	0.86	0.95	1.00	0.83	0.86	0.83	**0.89**
	70%	0.93	0.78	1.00	0.83	1.00	1.00	0.92
SVM	40%	0.92	0.88	1.00	0.77	0.86	1.00	0.90
	50%	0.80	0.95	1.00	0.91	0.80	0.83	**0.88**
	60%	0.80	0.84	1.00	0.83	1.00	0.83	0.89
	70%	0.92	0.84	1.00	0.91	0.89	1.00	**0.93**
RF	40%	0.88	0.95	0.92	0.67	0.67	1.00	0.85
	50%	0.88	0.84	0.80	0.91	0.55	0.83	0.80
	60%	0.82	0.90	1.00	0.80	0.75	1.00	0.88
	70%	1.00	0.90	1.00	0.80	0.75	1.00	0.91

The bold content represents the highest F1 score for each training ratio.

are considered together. This is impacting multi-event classes, such as Classes 3.1 and 3.4. Overall, the best classification result is obtained by the GLR classifier with a balance split ratio (50%) for training and testing for the SZ10 dataset, while for the PG dataset, the split ratio of 40% with GLR obtained the highest F1 score.

3.4.5 Complexity Analysis

Table 3.11: The execution time for each approach.

SZ10 (5 min) Fs = 1000 Hz 300,000 samples	Approach	Time (s)
BPF + Denoising	BPF	0.04
	Matlab Wavelet app	0.11
	Block thresholding	254.65
	STFT	16.93
	EEMD	2.85
	GraphBF	36.62
Detection	Neyman-Pearson lemma	0.12
	Matched filter	0.12
	STA/LTA	2.25

Table 3.12: Feature selection and classification execution time in seconds

Class 1.1 50% train (168) test (135) Time (s)			
Workflow	SVM	RF	GLR
Feature selection	761.76	9195.8	959.04
No. optimised features	59	76	29
Classification training	0.64	1.91	0.06
Classification testing	0.06	0.13	

Table 3.11 and 3.12 show the execution time in seconds (s) for each approach to process a signal from the SZ10 dataset of 5 min duration. Note that the training in the classification step includes feature selection and the synthetic data generation steps before training the models. The execution time was tested based on class 1.1. All experiments were run on a desktop with i7-7700K CPU

and 16GB RAM using Matlab 2019b. From Table 3.11, it can be observed that although the GraphBF denoising approach is slower than benchmarks, the overall processing time of the proposed end-to-end system is low. Indeed, the system (BPF + GraphBF + detection via Neyman-Pearson lemma + GLR classifier) needs roughly 37 s to process 5 min of data; hence it can be implemented for efficient processing. Note that one-off feature selection and classification training are carried out offline, and only testing execution time is provided.

3.5 Summary

This chapter presents a novel automated end-to-end system that can denoise a raw recorded signal, detect seismic events, construct and select the optimal features characterising each event type and finally classify events into specific classes. Having described the noise distribution in the field datasets, the graph signal processing-based denoising and classification approaches are adopted, and a low-complexity Neyman-Pearson-based hypothesis testing approach is used for detection. The proposed methods are benchmarked against the state-of-the-art denoising, detection, and classification approach in the seismic processing literature, showing consistently better performance for each step in terms of improved denoising capability via SNR measure, improved detection capability by ensuring all actual events are detected while minimising false alarms, and improved classification via F1 score. The system is shown to be of low complexity via run-time results. Despite demonstrating landslide-induced (micro)seismic events, the methodology can be tuned to suit other data sets. The following chapter explores feature engineering, including feature embedding, to better understand and explain feature importance in classifying specific (micro)seismic events.

Chapter 4

Graph-Based Feature Weight Optimisation and Classification of Seismic Sensor Array Recordings

In Chapter 3, an automatic end-to-end (micro)seismic events processing platform containing graph-based denoising and classification and detection based on Neyman Pearson lemma. Though with the state-of-the-art feature selection methods, more focused feature sets are provided for classification. The existing approaches either do not consider the class labels or suffer from high computation complexity; therefore, there is no method for feature selection that is superior for all target signals. Additionally, Chapter 3 only considers the (micro)seismic signal recorded at a single channel. Though the proposed methods outperform the benchmarks, they depend heavily on the signal SNR of the chosen channel. To improve the detection performance, in this chapter, a novel multi-channel detection method is proposed based on the Neyman Pearson lemma and Multi-channel Coherency Migration(MCM) that considers the multi-channel recorded data. Furthermore, this chapter proposes a novel semi-supervised graph-based feature weight optimisation and classification method based on graph piece-wise smoothness to

improve the performance of feature selection. Specifically, the proposed methods were evaluated with datasets SZ13, SZ14, and SZ15, recorded by the sensor array A (SZC) as illustrated in Fig. 1 of [4]. This chapter is based on the material published in the journal paper (J2) and conference paper (C2).

4.1 Introduction

As demonstrated in Chapter 3, the proposed detection scheme outperforms STA/LTA and template matching detection methods by utilising the Neyman Pearson lemma and the probability distribution analysis of the ambient noise. While the proposed approach is based on single-channel recording, the obvious limitation is that correlations between data recording channels are not considered. Furthermore, feature selection with filter methods, wrappers, and embedded methods shows promising results in highlighting discriminate features and improving classification, but there are still limitations.

As discussed in [30], seismic detection algorithms can be applied to individual components (recording channels) of data recordings or their combinations. Utilising a components combination reduces data dimensionality and computational cost by providing a single attribute is beneficial in obtaining a unique estimation of seismic arrival times from multi-channel recorded data, which is suitable for real-time monitoring applications. The absolute value of the amplitudes product and the stack of absolute value is employed as input to STA/LTA algorithms to increase the SNR by damping random noise for seismic detection [35, 36]. According to [30], both methods improve the waveform pattern of seismic events, but the performance is affected by strong noise in any of the components. The product effectively suppresses ambient noise while significantly reducing signal amplitude if it is present only on one of the components. A one-step multi-channel detection and classification approach is proposed by first sliding the predefined

length window on the multi-channel data stream; then with the 55 handcrafted features, which are constructed from each window of each channel to classify the windows into noise, slope failure, and earthquake utilising RF classifier with majority voting strategy [28]. This one-step approach suffers from time-consuming feature construction, making it unsuitable for real-time applications. Additionally, the significance of the features associated with the whole waveform of the signal is not analysed because this one-step approach only utilises the portions of the event's waveform that appear in the sliding window.

Cross-correlation is commonly used in seismology to assess the coherence of waveforms from different data recording channels. MCM method is proposed for detection that focuses the source energy by utilising the stacked coherencies between different receiver (data recording channel) pairs [120]. The coherency between all possible receiver (data recording channel) pairs is measured by normalising the covariance matrix of the recorded waveforms. This chapter proposes a novel multi-channel seismic detection method that merges MCM and the Neyman-Pearson lemma to identify potential events, even when the moment is low (microseismic).

Compared to (un)supervised learning with the feature mentioned above selection, semi-supervised learning only requires a small number of labelled event samples, which could reduce the effects of human error due to labelling uncertainty while maintaining relatively high accuracy [121]. Motivated by recent successes of graph-based semi-supervised learning, mainly focused on 2 or 3-dimension image data [80, 81], this chapter explores the graph-based semi-supervised learning on feature weighting due to its ability to handle classes with arbitrary signal distribution generating a smooth feature subspace. Graph learning refers to finding a signal representation via a graph, relying on statistical or spectral graph methods based on data observations to represent the signal in a low-dimensional subspace [122]. Graph spectral-based feature weight learning of [80] assigns an

importance score to each feature, assuming feature independence. In [80], constructed/extracted feature vectors are embedded onto a representation graph, where the distances between detected signals are assessed with the feature vectors and a critical parameter known as graph kernel bandwidth, which is usually manually set. Appropriate estimation of graph kernel bandwidth is challenging but essential for graph signal representation. Some recent studies on optimising graph kernel bandwidth, reported in [81], [123], [124], are either tied to a specific problem or might be affected by the randomness of feature pair selection. In [125], an iterative, alternating feature learning and classification approach is proposed for characterising slide quake, earthquake, tremor, and calibration shot established on the single channel recorded signal.

This chapter proposes a comprehensive and integrated (micro)seismic monitoring workflow for the continuous multi-channel recorded signal at an ongoing landslide, which consists of multi-channel event detection with linear coherency analysis via MCM, graph-based feature weight optimisation, and classification. Specifically, with the continuous recordings from multiple sensors in an array, the proposed system first detects potential events with the coherence analysis MCM of the multi-channels and identifies the strongest signal component/channel for feature construction. Then, the graph-based feature weight optimisation is designed for landslide-induced event classification, comprising, in addition to the earthquake, endogenous events (e.g., rockfall) and (micro)seismic events related to landslide processes (e.g., fissure formation) referred to as slide quake. Technically, the iterative, alternating feature weight optimisation and classification approach of [80] (see also [125]) is modified with a new dual problem to reduce the computational complexity of the algorithm.

The remaining sections are organised as Section 4.2 introduces the proposed methodology, including multi-channel detection and graph-based feature weight optimisation and classification. Section 4.3 presents the experimental results of

the proposed methods compared with state-of-the-art methods. Section 4.4 provides the most discriminated features concluded by the proposed approach. Eventually, Section 4.5 summarises the work of this chapter.

4.2 Methodology

This section describes the methodology as the workflow shown in Fig. 6.2, and the blocks within will be covered in the following content.

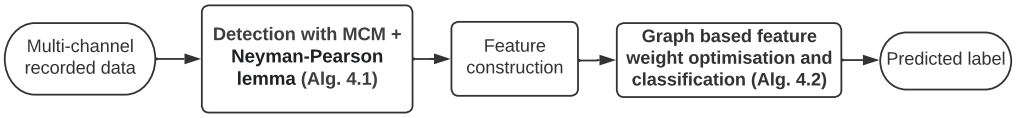


Figure 4.1: Workflow of the proposed system. The proposed Algs. 4.1 and 4.2 are in bold.

4.2.1 Multi-Channel Detection

Seismic monitoring is regularly performed with multiple sensors deployed over the area of interest to continuously record the activities over vast distances of the order of kilometers. However, depending on the relative distance of the source to the sensors, some sensors may not record a particular event at a sufficiently large SNR to be identifiable. Combining the readings from multiple sensors has been shown to improve event detection, e.g., by stacking signals from various channels as in [120]. These events tend to be localised, channel stations are relatively close for such studies, and therefore, the time difference in signal arrival at different events is negligible for multi-channel event detection and classification.

For the detection stage, the multi-channel recorded signals donated as $\mathbf{X} \in \mathbb{R}^{C \times J}$, where C and J, respectively, represent the number of recorded channels (this chapter considers vertical channels only) and signal samples in that channel. The pre-processing steps contain normalising and filtering the recorded signal to

minimise the effect of signal attenuation and measurement noise. As the first step of detection, the recorded data $\mathbf{X}$ segmented into non-overlapped length-l windows $\mathcal{W} = \{\mathbf{W}_1, ..., \mathbf{W}_I\}$, $\mathbf{W}_i \in \mathbb{R}^{C \times l}$ where the total number of windows is $I = \frac{N}{l}$. Then, this chapter analyses the linear coherence across each channel $c_p \in \{1, ..., C\}$ (i.e., traces from different deployed stations) within window $\mathbf{W}_i$ utilising MCM to form stacked signal $\mathbf{r}$. This chapter only considers the recorded vertical channels.

In [120], the MCM defined as:

$$
\begin{aligned}
r_i^m &= \frac{\sum_{t=1}^{l} \left[\left(\mathbf{W}_i^{c_1}(t) - \overline{\mathbf{W}}_i^{c_1}\right)\right] \left[\left(\mathbf{W}_i^{c_2}(t) - \overline{\mathbf{W}}_i^{c_2}\right)\right] \cdots \left[\left(\mathbf{W}_i^{c_p}(t) - \overline{\mathbf{W}}_i^{c_p}\right)\right]}{(l-1)\, \beta_i^{c_1} \beta_i^{c_2} \cdots \beta_i^{c_p}}, \\
r_i &= \sum_{m=1}^{M} r_i^m, \qquad \overline{\mathbf{W}}_i^{c_p} = \frac{1}{l} \sum_{t=1}^{l} \mathbf{W}_i^{c_p}(t) \ , \forall c_p \in \{1, ..., C\},
\end{aligned}
\tag{4.1}
$$

where r_i^m is the m-th multidimensional waveform coherency (p-dimension) among different data recording channels $c_1, c_2, ..., c_p$, $m \in \{1, ..., M\}$ and M is the total number of p-wise groups of channels), $\mathbf{W}_i^{c_p}(t)$ denotes the channel c_p of $\mathbf{W}_i$ with t samples, and $\beta_i^{c_p}$ is the standard deviation of the corresponding window. By utilising multidimensional waveform coherency, the total number of adequate information available for migration is improved from C to $M = \frac{C!}{p!(C-p)!}$ (! denotes factorial).

Next, similar to [15], the Neyman–Pearson lemma removes the stacked signals r_i that most likely contain only ambient noise with low SNR. After concatenating the remaining consecutive windows, new windows formed as $\widetilde{\mathbf{W}}_j$. For example, suppose that $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$, $\mathbf{W}_5$, and $\mathbf{W}_6$ are the remaining windows after detection via Neyman–Pearson lemma. $\widetilde{\mathbf{W}}_1$ is formed by concatenating $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$ and $\widetilde{\mathbf{W}}_2$ is formed by concatenating $\mathbf{W}_5$ and $\mathbf{W}_6$. In the final step, the best channel segments $\widetilde{\mathbf{w}}_j^{c^*}$ are automatically selected (i.e., the ones that maximise SNR among all channels within each $\widetilde{\mathbf{W}}_j$) as the detected events to feed to the

feature construction step (see Fig 6.2). The proposed multi-channel detection is summarised in Alg. 4.1.

Algorithm 4.1 Multi-channel detection

Require: Recorded multi-channel data $\mathbf{X}$; Window length l; No. of channels C;

Ensure: Detected events $\widetilde{\mathbf{w}}_j^{c^*}$;

1: **Preprocessing:** Filter $\mathbf{X}$ with a $(5-100\,\text{Hz})$ band-pass filter and split it into non-overlapping windows $\mathbf{W}_i$ of length l;
2: For each $\mathbf{W}_i$, use MCM [120] Eq. 4.1 to obtain r_i;
3: Set threshold $\lambda \leftarrow$ using Eq. 3.9;
4: Keep windows $\mathbf{W}_i$ for which $r_i > \lambda$, and concatenate all such consecutive windows, to form new window $\widetilde{\mathbf{W}}_j$ with $\widetilde{\mathbf{W}}_j^{c_p}$ denoting its c_p-channel;
5: **for** each window $\widetilde{\mathbf{W}}_j$ **do**
6: Identify channel c^* that maximises SNR across all $\widetilde{\mathbf{W}}_j^{c_p}$, $c_p \in \{1, ..., C\}$ [15];
7: **return** The detected events $\widetilde{\mathbf{w}}_j^{c^*}$;

The choice of window length l should consider the trade-off between noise suppression and time resolution [120]. In this chapter, the durations of the target signals vary from $0.5-100\,\text{s}$, while the majority are within the range $1-2\,\text{s}$, thus $l = 0.2\,\text{s}$ is chosen as the length of all windows in set $\mathcal{W}$. MCM is then used on $\mathbf{W}_i$ to obtain a stacked signal r_i as in [120] (Step 2), where the number of channels for signal linear coherency analysis is set as $p = 3$, as in Eq. 4.1. If the amplitude of the stacked signal r_i exceeds the calculated threshold λ (Step 3) obtained with Neyman–Pearson lemma, the windows are kept and concatenated to form $\widetilde{\mathbf{W}}_j$ (Step 4). Finally, for each $\widetilde{\mathbf{W}}_j$, the detected segments $\widetilde{\mathbf{w}}_j^{c^*}$ are obtained from a channel c^* that maximises the SNR over all vertical channels [15].

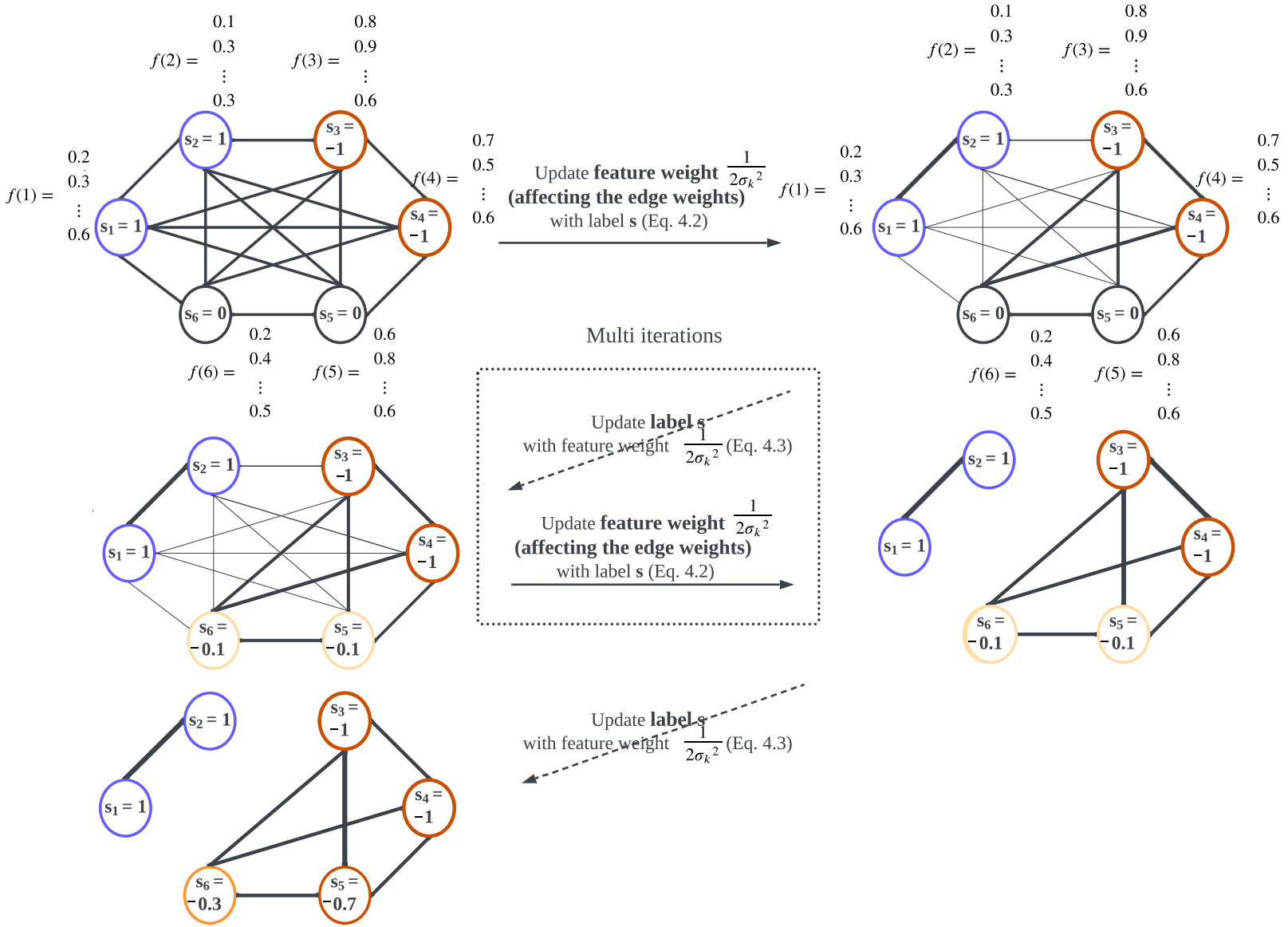


Figure 4.2: Graph-based feature weight optimisation and classification (schematic diagram), where the solid line represents Graph-based feature weight optimisation, and dashed line represents Graph-based classification.

4.2.2 Graph-Based Feature Weight Optimisation and Classification

After events are detected from multi-channel recorded data, as discussed in the previous subsection, the K-dimension feature set is constructed for each detected event $\widetilde{\mathbf{w}}_j^{c^*}$. The signal temporal, spectral, cepstral, and acoustic features are calculated, and the polarity attributes are calculated independently on the three-component seismometers. Then, the constructed features embed into a connected, "representation" graph, $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is the set of vertices and $\mathbf{A}$ is the graph adjacency matrix [15]. Each vertex in the graph $\nu \in \mathcal{V}$ corresponds to one of the detected events and is characterised by the corresponding K-dimensional feature vector. The graph needs to represent well the relationships between the events and is learned based on the importance of the features, as described next.

Let $f_k(i)$ represents the k-th feature of event i assigned to vertex ν_i. Then, the (i, j) entry in $\mathbf{A}$, $A_{i,j}$, i.e., the weight of the edge between nodes i and j, is set as Eq. 3.10 with $\sigma = \{\sigma_1, ..., \sigma_K\}$. From which, σ_k represents *the graph kernel bandwidth* for the k-th feature (i.e., $\frac{1}{2{\sigma_k}^2}$ is the weight of feature k). A too-small value of the bandwidth would lead to a poor representation of the local structures; conversely, a considerably high value could result in a coarse description of the data.

To construct the graph, this section assigns to each node a discrete graph signal $\mathbf{s}$ that carries the class label of the corresponding event in the training set and zero for nodes corresponding to the test set, as Eq. 3.11. Thus, if graph $\mathcal{G}$ captures the well correlation between the events, then the nodes with the same label will be connected by high-weight edges, that is, the N-length graph signal $\mathbf{s}$ will be piecewise smooth concerning $\mathcal{G}$ and the missing labels can be extrapolated (that are initialised to zero), e.g., via GLR, by finding the smoothest graph signal that fits the training data [125].

Fig. 4.2 provided as an illustration (a toy example) to further clarify the graph-based feature weight optimisation and classification approach, where an undirected graph with six nodes is built that corresponds to the detected events. Specifically, in Fig. 4.2, four nodes are used for training, i.e., $s_1, s_2 = 1$, $s_3, s_4 = -1$, denoting that the first two nodes correspond to Class 1 (blue circles) and the third and fourth nodes do not belong to Class 1 (red circles) events while nodes five and six (yellow) are used for testing and are initialised to $s_5, s_6 = 0$. As illustrated in Fig. 4.2, the whole workflow includes two steps: 1) Graph-based feature weight optimisation (shown as a solid line in Fig. 4.2) based on feature vector set $\mathcal{F}= \big\{\mathbf{f}(i), ..., \mathbf{f}(N)\big\}$, where $\mathbf{f}(i) = [f_1(i), ..., f_K(i)]^\top$; here the optimisation problem as in Eq. 4.2 to determine the feature weight $\frac{1}{2{\sigma_k}^2}$ and then the graph edge weights are obtained with Eq. 3.10 (a thick edge indicates that the connected events have high correlation), 2) graph-based classification, (dashed line in Fig. 4.2), with the optimised edge weights; the labels of the testing events s_5 and s_6 are obtained with Eq. 4.3. The above two steps alternatively and iteratively update the feature weights and event labels until the stopping criteria. These two steps are described next in more detail.

Graph-based feature weight optimisation

This subsection provides a detailed description of the adapted graph feature weight optimisation (the solid line step in Fig. 4.2), which is first proposed in [80]. Specifically, the algorithm, shown as Alg. 4.2, represents the iterative alternating binary classification via (norm)GLR [125] or with GTV [122], and graph-based feature weight optimisation via:

$$\arg\min_{\sigma_k} \sum_{i,j} \exp\left\{-\sum_{k=1}^{K} \frac{(f_k(i) - f_k(j))^2}{2\sigma_k^2}\right\} (\widetilde{s}_i - \widetilde{s}_j)^2 + U(\sigma), \tag{4.2}$$

where $\widetilde{s}_i$, $\widetilde{s}_j$ are the predicted graph signal, and $U(\sigma)$ is an indicator function, that returns 0 if all elements of $\sigma = [\sigma_1, \ldots, \sigma_K]$ are in the range $(0, 1]$, or ∞, otherwise. The algorithm is initialised to $\sigma_k = 0.7$ (Step 2) and then adapts the graph, i.e., σ_k, via feature selection and prioritisation by minimising Eq. 4.2 with gradient descent (Steps 7 and 10). The optimisation problem Eq. 4.2 subjects to $\sum_k \sigma_k \in (0, \gamma]$.

Algorithm 4.2 Alternative Graph-based Feature Weight Optimisation and Classification

Require: Constructed feature set $\mathcal{F}$;
Initial graph signal $\mathbf{s}$; Tolerance μ, ϵ; Constant γ; Step size θ;
Ensure: The predicted label $\widetilde{\mathbf{s}}_{n+1:N}$; Feature weight $\frac{\mathbf{1}}{\mathbf{2}\widetilde{\boldsymbol{\sigma}}^{\mathbf{2}}}$;

1: **Initialization**: $t = \epsilon$;
2: Initialise Graph kernel bandwidth σ;
3: **while** $t > \mu$ **do**
4: $\quad \mathbf{A} \leftarrow$ Eq. 3.10) and update $\mathbf{H}$;
5: $\quad \widetilde{\mathbf{s}}_{n+1:N} \leftarrow$ Eq. 4.5; $\hat{t} = \epsilon$;
6: $\quad m \leftarrow \widetilde{\mathbf{s}}^\top \mathbf{H} \widetilde{\mathbf{s}}$;
7: $\quad \nabla \leftarrow$ via gradient descent Eq. 4.2 [80]; $j = 0$;
8: $\quad$ **while** $\hat{t} > \mu$ **do**
9: $\quad\quad$ **if** $j > 0$ **then**
10: $\quad\quad\quad \nabla \leftarrow$ via gradient descent Eq. 4.2; [80]
11: $\quad\quad \forall k, \widetilde{\sigma}_k = \sigma_k - \theta \times \nabla$; $j = j + 1$;
12: $\quad\quad q = \widetilde{\mathbf{s}}^\top \mathbf{H} \widetilde{\mathbf{s}}$;
13: $\quad\quad$ Update $\mathbf{A} \leftarrow$ Eq. 3.10 and update $\mathbf{H}$ using $\widetilde{\sigma}_k$;
14: $\quad\quad \hat{q} = \widetilde{\mathbf{s}}^\top \mathbf{H} \widetilde{\mathbf{s}}$;
15: $\quad\quad \hat{t} = \hat{q} - q$;
16: $\quad \widetilde{\mathbf{s}}_{n+1:N} \leftarrow$ Eq. 4.5 with updated $\mathbf{H}$;
17: $\quad \hat{m} = \widetilde{\mathbf{s}}^\top \mathbf{H} \widetilde{\mathbf{s}}$;
18: $\quad t = \hat{m} - m$;
19: **return** The predicted label $\widetilde{\mathbf{s}}_{n+1:N}$ and feature weights $\frac{\mathbf{1}}{\mathbf{2}\widetilde{\boldsymbol{\sigma}}^{\mathbf{2}}}$;

Feature selection is performed iteratively between feature weight update (Step 11) and classification (Steps 12 and 14). The definition of $\mathbf{H}$ in Alg. 4.2 (Step 4) depends on the classifier used. Namely, $\mathbf{H}$ represents $\mathbf{L}_{\text{norm}}$ or $\widetilde{\mathbf{A}}$ for normGLR or GTV classifier. Note that $\sigma_k \geq 0$ represents the feature weight given to feature k, where smaller σ_k indicates that k-th feature is more important. Thus the algorithm inherently performs feature selection.

Graph-Based Classification

The graph classification problem, i.e., finding the smoothest graph that spans training and testing nodes and fits the training labels, can be solved via quadratic formulation as in [15], which is fast but with poor worst-case errors and requires finding a pseudo-inverse of potentially large Laplacian matrix. Alternatively, Nondeterministic Polynomial time quadratic constrained quadratic relaxation, and Semi-Definition Relaxation (SDR) are known to provide good error-bounded approximations [80]. Hence, the graph-based classifiers are solved with SDR, and following the idea of [80], with a general formation:

$$\begin{aligned} \min_{\mathbf{S},\mathbf{s}} \quad & \mathrm{tr}(\mathbf{HS}) \\ \text{s.t.} \quad & \mathbf{S}_{i,i} = 1, \quad i \in \{1, ..., N\} \\ & \begin{bmatrix} \mathbf{S} & \mathbf{s} \\ \mathbf{s}^\top & \mathbf{1} \end{bmatrix} \succeq 0, \quad s_i = \overline{s}_i, \quad i \in \{1, ..., n\}, \end{aligned} \tag{4.3}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $\overline{s}_i$ denotes the label of the i-th event in the training set, with $n < N$ denoting the size of the training set.

Optimisation problem Eq. 4.3 is of polynomial complexity, thus this section converts Eq. 4.3 to its dual problem [80], defined as:

$$\begin{aligned} \hat{\mathbf{g}} = \arg\max_{\mathbf{g}} & -\mathbf{1}^\top \mathbf{g} \\ \text{s.t.} \quad & \mathbf{H} + \mathrm{diag}(\mathbf{g}) \succeq 0, \end{aligned} \tag{4.4}$$

where $\mathbf{g}$ is the Lagrange multiplier. With the obtained optimised $\hat{\mathbf{g}}$, the optimised solution of Eq. 4.3 is equivalent to finding the infimum of Lagrange function:

$$
\begin{aligned}
&\mathbf{s}^\top(\mathbf{H}+\mathrm{diag}(\hat{\mathbf{g}}))\mathbf{s}-\hat{\mathbf{g}}1\\
&=\mathbf{s}^\top\hat{\mathbf{H}}\mathbf{s}-\hat{\mathbf{g}}1\\
&=\begin{bmatrix}\mathbf{s}_{1:n}^\top & \mathbf{s}_{n+1:N}^\top\end{bmatrix}\begin{bmatrix}\hat{\mathbf{H}}_{1:n,1:n} & \hat{\mathbf{H}}_{1:n,n+1:N}\\ \hat{\mathbf{H}}_{n+1:N,1:n} & \hat{\mathbf{H}}_{n+1:N,n+1:N}\end{bmatrix}\begin{bmatrix}\mathbf{s}_{1:n}\\ \mathbf{s}_{n+1:N}\end{bmatrix}-\hat{\mathbf{g}}1\\
&=\mathbf{s}_{1:n}^\top\hat{\mathbf{H}}_{1:n,1:n}\mathbf{s}_{1:n}+(\mathbf{b}^\top+\mathbf{c})\mathbf{s}_{n+1:N}+\mathbf{s}_{n+1:N}^\top\hat{\mathbf{H}}_{n+1:N,n+1:N}\mathbf{s}_{n+1:N}-\hat{\mathbf{g}}1,
\end{aligned}
\tag{4.5}
$$

where $\hat{\mathbf{H}}=\mathbf{H}+\mathrm{diag}(\hat{\mathbf{g}})$, $\mathbf{b}=\hat{\mathbf{H}}_{1:n,n+1:N}^\top\mathbf{s}_{1:n}$ and $\mathbf{c}=\mathbf{s}_{1:n}^\top\hat{\mathbf{H}}_{1:n,n+1:N}^\top$. By setting the first-order derivative of the Lagrange function Eq. 4.5 to zeros, the solution obtained as $\widetilde{\mathbf{s}}_{n+1:N}=\frac{-\hat{\mathbf{H}}_{n+1:N,n+1:N}^{-1}(\mathbf{b}+\mathbf{c}^\top)}{2}$.

4.3 Experimental Setup and Results

4.3.1 Multi-Channel Detection

Raw recorded data from the vertical channels are fed into Alg. 4.1 that selects the optimal channel segments that are least affected by noise and remove the segments with low signal activity, where noise is predominant. An example of Alg. 4.1 output is shown in Fig. 4.3 & 4.4 for four bandpass filtered vertical channels for the same window. With threshold $\lambda = 800nm^3/s^3$ calculated with Eq. 3.9, two stacks at times 05:05:13 and 05:06:59 are deemed of interest. These are highlighted in the blue rectangular box (kept windows) in Fig. 4.4, where the amplitude of each channel is scaled to $[-1, 1]$. Step 6 identified the optimal channel (the one with the highest SNR) c^* is ch5 for both windows $\widetilde{\mathbf{W}}_{5742}$ and $\widetilde{\mathbf{W}}_{5743}$; thus, the segments $\widetilde{\mathbf{w}}_{5742}^5$ and $\widetilde{\mathbf{w}}_{5743}^5$ (in red rectangular box) are selected.

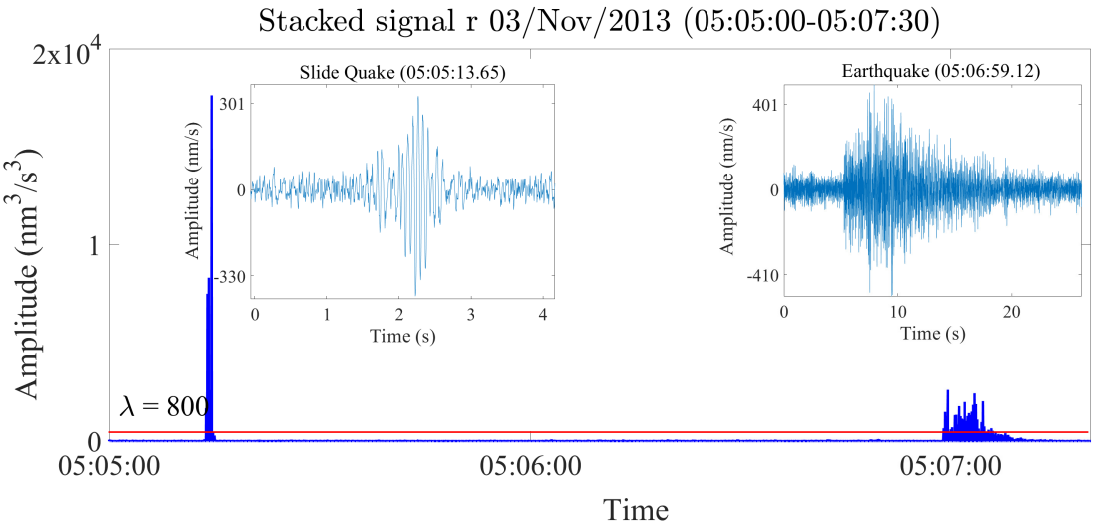


Figure 4.3: Example of stacked signal **r** as shown in (Alg. 4.1, step 2 & 3) (Period: 03/November/2013 (05:05:00-05:07:30)).

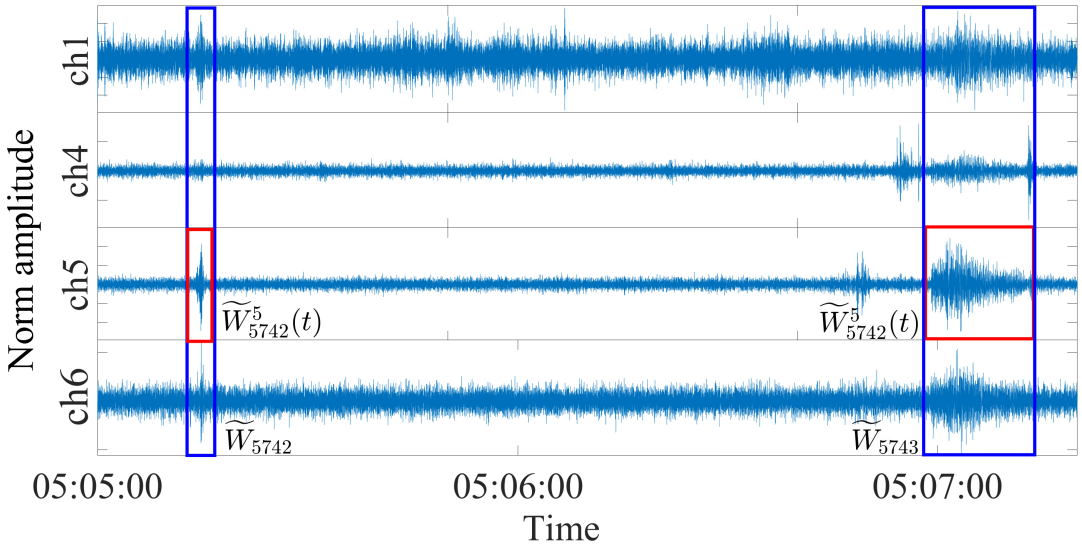


Figure 4.4: Example of the kept $\widetilde{\mathbf{W}}_{5742\&5743}$ (blue rectangular box) & $\widetilde{\mathbf{w}}^5_{5742\&5743}$ (red rectangular box) as given in Alg. 4.1, step 4 (Period: 03/November/2013 (05:05:00-05:07:30)).

Detection on Catalogued Events

To demonstrate the effectiveness of Alg. 4.1, the section benchmarked its performance against the single channel detection approach of [15] using catalogued events [4]. As shown in Table 4.1, compared to [15], Alg. 4.1 detected more catalogued events (shown as True positives), missed fewer catalogued events (shown as False negatives) and detected fewer additional events (shown as False positives) that are not present in the catalogue. The 23 events missed by Alg. 4.1 comprised 2 rockfalls, 3 slide quakes, 1 earthquake, and 17 noise events. The 28555 detected events that were not catalogued are not necessarily non-endogenous or non-seismic and could have been missed during manual labelling at the cataloguing stage. Therefore, Alg. 4.1 is effective in maximising the detection of catalogued events, missing fewer events, and minimising the detection of uncatalogued events.

Table 4.1: Detection on catalogued events for entire period

Approach	TP	FN	FP
Detection [15]	978	396	108288
Proposed detection (Alg. 4.1)	1351	23	28555

Detection on Continuous Data Recorded Between 24-28/November/2014

To further validate the performance of the proposed multi-channel detection, an expert was invited to manually identify the (micro)seismic events on multi-channel continuous recorded data between 24-28/November/2014. The period was chosen because, according to the catalogue of [4], all four events are present in this period (65 rockfalls, 18 slide quakes, 23 earthquakes, and 14 noise), which were detected with an STA/LTA algorithm applied in the frequency domain.

As illustrated in Table 4.2, 1006 events are newly identified by the expert, which was missed with STA/LTA algorithm applied in the frequency domain [4]. 614 of these 1006 events have been detected with proposed Alg. 4.1, and these

Table 4.2: Continuous detection result verification for the selected period.

Approach	Catalogued	Not-catalogued
Detection [4]	120	0
Manual expert detection	120	1006
Proposed detection (Alg. 4.1)	119	614

include 6 rockfalls, 3 slide quakes, 2 earthquakes, and 603 noise events that were missed by STA/LTA in [4]. This further demonstrates the effectiveness of the proposed detection technique. Additionally, the majority of the 392 (392=1006–614) missed occurrences by the proposed multi-channel detection were 295 noise events and only 45 rockfalls, 48 slide quakes, and 4 earthquakes. The difficulty of recognising rockfall and earthquake events arises since they are regarded as endogenous landslide seismicities with varying SNR.

4.3.2 Feature Engineering

This section compares the output of conventional feature extraction and selection approaches, as discussed in Chapter 1, to that of the adapted graph feature weight optimisation as Alg. 4.2, described in Section 4.2.2. Specifically, in Alg. 4.2, the hyperparameters are set as $\mu = 0.01$, $\epsilon = 10000$, and θ as the one over the Lipschitz constant.

Case Study 1: Feature Engineering with SZ10 Dataset

Leveraging on prior microseismic event characterisation, [14] as well as time-series acoustic signal classification [12], the constructed feature set contains 48 features, shown in Table 4.3, to characterise (micro)seismic events arising from landslides. The 48 candidate features include spectral features (features 1-15), temporal features (features 16-42), and cepstral features (features 43-48) Table 4.3. The task is to automatically select an optimal subset of the 48 candidate features and classify the related events into one of four classes (earthquake, slide quake, tremor, and calibration shot) (see Section 1.3.1).

Table 4.3: Candidate features: $\mathbf{y}(t)$ is the seismic signal in time domain, $\mathbf{p}(t)$ is the power signal, $\mathbf{e}(t)$ is the signal envelope, $\mathbf{q}(v)$ is the frequency domain signal and $\mathbf{c}(v)$ is the cepstrum.

1.Mean of $\mathbf{q}(v)$	17.Root mean square $\mathbf{y}(t)$	33.Rate of attack $\mathbf{y}(t)$
2.Max of $\mathbf{q}(v)$	18.Standard deviation $\mathbf{y}(t)$	34.Rate of attack $\mathbf{e}(t)$
3.Median of $\mathbf{q}(v)$	19.Max value of $\mathbf{y}(t)$	35.Rate of decay $\mathbf{y}(t)$
4.Variance of $\mathbf{q}(v)$	20.Average power of $\mathbf{y}(t)$	36.Rate of decay $\mathbf{e}(t)$
5.Max envelope PSD of $\mathbf{q}(v)$	21.Skewness of $\mathbf{y}(t)$	37.Ratio1=30/31
6.Frequency 95%bandwidth	22.Skewness of $\mathbf{p}(t)$	38.Ratio2=30/32
7.Frequency 50%bandwidth	23.Median value of $\mathbf{y}(t)$	39.Ratio3=30/STD($\mathbf{e}(t)$)
8.Dominate frequency	24.Kurtosis of $\mathbf{y}(t)$	40.Ratio4=$(t_{max}-1)/(N-t_{max})$
9.Median normal $\mathbf{q}(v)$	25.Kurtosis of $\mathbf{p}(t)$	41.Onset of $\mathbf{y}(t)$
10.Variance normal (v)	26.Entropy of $\mathbf{p}(t)$	42.Offset of $\mathbf{y}(t)$
11.Int-ratio [109]	27.Entropy of $\mathbf{y}(t)$	43.Std of $\mathbf{c}(v)$
12.Number of peaks of $\mathbf{q}(v)$	28.Kurtosis of $\mathbf{e}(t)$	44.Mean of $\mathbf{c}(v)$
13.Gamma 1 [109]	29.Skewness of $\mathbf{e}(t)$	45.Skewness of $\mathbf{c}(v)$
14.Gamma 2 [109]	30.Max of $\mathbf{e}(t)$	46.Kurtosis of $\mathbf{c}(v)$
15.Gamma 3 [109]	31.Meam of $\mathbf{e}(t)$	47.Dominate Frequency of $\mathbf{c}(v)$
16.Duration $\mathbf{y}(t)$	32.Median of $\mathbf{e}(t)$	48.Max of $\mathbf{c}(v)$

Firstly, similarly to the previous chapter, the bandpassed filter denoising and Neyman-Pearson lemme detection resulted in 223 detected events (56 earthquakes (class 1.1 and 1.2 in Table 1.1), 58 slide quakes (class 2.2, 2.3 and 2.4 in Table 1.1), 37 tremors (class 3.1, 3.2 and 3.4 in Table 1.1), 11 calibration shots (class 9 in Table 1.1), and 12 undefined events) plus 49 false alarms for SZ10 dataset.

Note that Section 3.4.2 concludes that the detection via Neyman-Pearson lemma outperforms the commonly used detection approach STA/LTA. As in Section 3.4.4, after feature construction, the SMOTE algorithm is implemented to generate synthetic feature samples based on all 48 features to balance the training set [118]. The dataset is split into 30% for testing and 70% for training, where the testing set does not contain any synthetic features.

As Alg. 4.2, in this section, the hyperparameter γ is set as 48 (48 is the dimension of the features, i.e., $K = 48$). The benchmarks utilised in this section contain the feature selection method in [107] Infinite Feature Selection (Inf-FS) (as used in Section 3.3.3) and two classifiers - SVM and RF - as used in [12] and [4], respectively. Inf-FS is a filter-based unsupervised feature ranking method

that outputs features sorted by importance. Then, cross-validation is utilised to pick the optimal subset of features with the highest classification accuracy performance. Finally, SVM or RF classifiers are used for classification based on the selected features. In addition, the harmonic function-based graph feature learning method of [123] is also implemented for benchmarking.

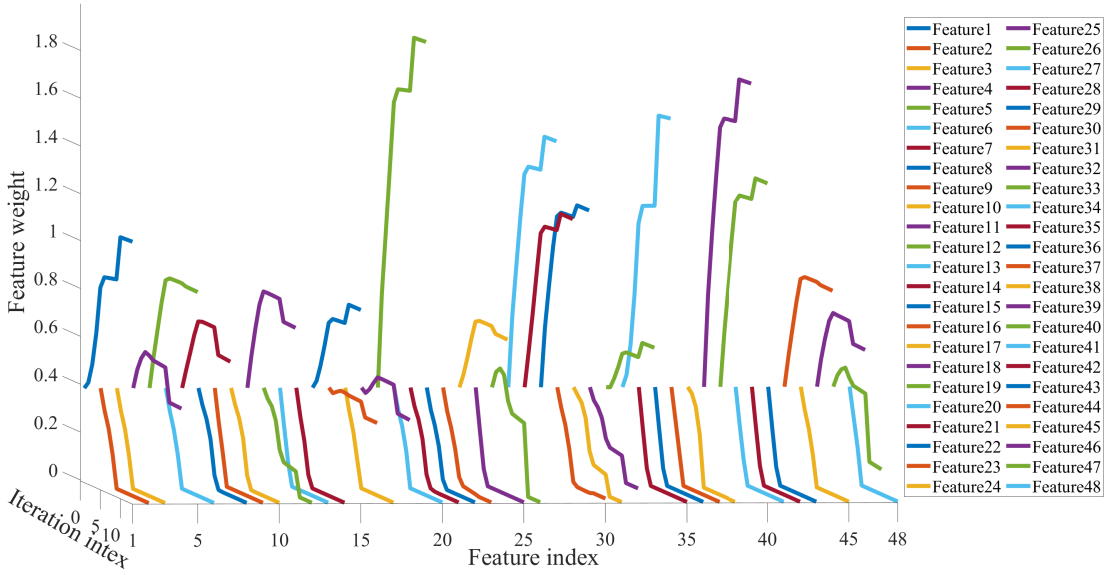


Figure 4.5: Variation of feature weights $\frac{1}{2\tilde{\sigma}^2}$, for all 48 features, with iteration index.

Fig. 4.5 shows how the feature weights for all 48 features vary across iterations of Alg. 4.2, which presents as the averaged feature weight value across four classes. Specifically, the feature weight axis represents the importance of the features which vary with the iterations, and the feature index refers to the features listed in Table. 4.3. For example, the weight of the first feature (Mean of $\mathbf{q}(v)$) increases with the iterations, and by the end of 10 iterations, the weight value has increased from 0.4 to 1. It can be observed that the variations in feature weights between iterations until convergence is reached (the curves become flat) around the tenth iteration, or even earlier for most features. Furthermore, eventually, only 22 features with non-zero weight once convergence is reached, showing a

dimensionality reduction of over 50%.

Case Study 2: Feature Engineering with the Datasets SZ13, SZ14 and SZ15

For each detected event segment from Alg. 4.1, $K = 119$ features are constructed, as listed in Table 4.4. With PCA feature extraction approach, the top 95% of the principal components are selected, amounting to, on average, 44 principal components per class, which are then used by the benchmarked SVM and RF classifiers.

Table 4.4: List of features drawn from the literature: temporal $\mathbf{s}(t)$, power signal $\mathbf{p}(t)$, envelope $\mathbf{e}(t)$, auto correlation function $\mathbf{ac}(t)$, spectral $\mathbf{f}(v)$, cepstrum domain $\mathbf{ce}(v)$, template rockfall $\mathbf{r}(t)$, slide quake $\mathbf{q}(t)$, earthquake $\mathbf{ea}(t)$, noise $\mathbf{n}(t)$, and envelope $\mathbf{es}(t)$ with (5 −10 Hz), (10 −50 Hz), (5 −70 Hz), (50 −100 Hz) and (5 −100 Hz) passband. **PMF** is Probability Mass Function.

Parameter	**Description**
Temporal Feature	
T1	Duration [4, 48, 108, 109]
T2	STD of $\mathbf{e}(t)$ [14, 48, 109–111]
T3	RMS between the decreasing signal and $\mathbf{J}(t) = Y_{max} - \frac{Y_{max}}{N-t_{max}}t$ [4]
T4	Optimum point of separation [110]
T5–7	Max, Mean, Median of $\mathbf{e}(t)$ [4, 14, 110]
T8, 9	Rising Decreasing duration $\mathbf{s}(t)$ [109]
T10	Entropy of $\mathbf{s}(t)$ $-\sum \mathbf{PMF}(\mathbf{s}(t)) \log \mathbf{PMF}(\mathbf{s}(t))$ [12, 111–113]
T11	Zero Cross Rate of $\mathbf{s}(t)$ [14, 111]
T12, 13, 16, 19	Max CrossCor ($\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$) [110]
T14, 15	Skewness of $\mathbf{p}(t)$ & $\mathbf{e}(t)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{x}(t_i)-\overline{\mathbf{x}}}{\sigma}\right)^3$ [4, 14, 108, 109, 111]
T17, 18	Kurtosis of $\mathbf{p}(t)$ & $\mathbf{e}(t)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{x}(t_i)-\overline{\mathbf{x}}}{\sigma}\right)^4$ [4, 14, 108, 109, 111]
T20,21,25,26	2dNormCrossCorAb ($\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$) [110]
T22	Rate of decay of $\mathbf{e}(t)$ [12]
T23	Ratio of Max and mean of $\mathbf{e}(t)$ $\frac{T_5}{T_6}$ [109]
T24	Ratio of Max and Median of $\mathbf{e}(t)$ $\frac{T_5}{T_7}$ [109]
T27,40,47,48	No.peaks 2dNormCrossCorAb ($\mathbf{q}(t)$,$\mathbf{r}(t)$,$\mathbf{ea}(t)$,$\mathbf{n}(t)$) [110]
T28–T32	Energy of $\mathbf{es}(t)$ $\sum_{i=1}^{N}\mathbf{es}(t_i)^2$ [4, 109]
T33-37	kurtosis of $\mathbf{es}(t)$ [4, 14, 108, 109, 111]
T38&39	Energy of (1:N/3) and (N/3:N) of $\mathbf{ac}(t)$ $\sum_{i=1}^{\frac{N}{3}}\mathbf{ac}(t_i)^2$ $\sum_{i=\frac{N}{3}}^{N}\mathbf{ac}(t_i)^2$ [4]
T40	Int-ratio of $\mathbf{ac}(t)$ $\frac{T_{38}}{T_{39}}$ [4, 109]
T41	No. of peaks of $\mathbf{ac}(t)$ [4, 109]
T42	duration of $\mathbf{ac}(t)$ $\max_t(\mathbf{ac}(t) < 0.2\max(\mathbf{ac}(t)))/(T_1)$ [4]
T43	Measure of location $\sum_{i=1}^{N} i\mathbf{p}(t_i)$ [113]
T44	Measure of dispersion $\sqrt{\sum_{i=1}^{N}(i-T_{43})^2\mathbf{p}(t_i)}$ [113]
T45	Measure of asymmetry $\frac{1}{T_{44}^3}\sum_{i=1}^{N}(i-T_{43})^3\mathbf{p}(t_i)$ [113]
T46	Measure of concentration around single value $\frac{1}{T_{44}^4}\sum_{i=1}^{N}(i-T_{43})^4\mathbf{p}(t_i)$ [113]

Table 4.4: *Cont.*

Parameter	Description
T49–52	No.peaks 2dNormCrossCorReal ($\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$) [110]
T53–56	2dCrossCorAb ($\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$) [110]
T57–60	2dCrossCorReal ($\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$) [110]
Spectral Feature	
F1–2	Absolute value of mean, max $\mathbf{f}(v)$ [4, 108, 109]
F3–4	Central frequency of the 1st and 2nd quartile [4]
F5	Max envelop of $\mathbf{f}(v)$ [4, 110]
F6	No. of peaks >0.75 bandwidth $\mathbf{f}(v)$ [4, 109]
F7	Dominate frequency [4, 108, 110]
F8	Spectral centroid [4, 109–111, 113]
F9–10	Median and variance of the normDFT [4, 14, 109–111]
F11	No. peaks of the autocorrelation $\mathbf{ac}(\mathbf{f}(\mathbf{v}))$ [4, 109]
F12-14,16	Energy of ([0,1/4],[1/4,1/2],[1/2,3/4],[3/4,1])*Fs [4]
F15	No. of peaks $\mathbf{f}(v)$ [4]
F17	Kurtosis of the Max of all DFTs [4]
F18,19	Mean ratio between the max and the mean and median of all DFTs [4]
F20,30,31	No. peaks in temporal evolution of the DFTs max, mean and median [4]
F21	Gamma 1 $\frac{\sum_{i=1}^{N} v_i \mathbf{f}(v_i)^2}{\sum_{i=1}^{N} \mathbf{f}(v_i)^2}$ [109]
F22	Gamma 2 $\sqrt{\frac{\sum_{i=1}^{N} v_i^2 \mathbf{f}(v_i)^2}{\sum_{i=1}^{N} \mathbf{f}(v_i)^2}}$ [109]
F23	Gamma 3 $\sqrt{\mid F21^2 - F22^2 \mid}$ [109]
F24	Mean frequency $\frac{\sum_{i=1}^{N} \mathbf{PSD}(v_i) v_i}{\sum_{i=1}^{N} \mathbf{PSD}(v_i)}$ **PSD** is the power spectral density of $\mathbf{f}(v)$ [4]
F25	Frequency bandwidth $2\sqrt{\frac{\sum_{i=1}^{N} \mathbf{PSD}(v_i) {v_i}^2}{\sum_{i=1}^{N} \mathbf{PSD}(v_i)} - F_{24}^2}$ [4]
F26	Minimal frequency $\min_v(\mathbf{PSD}(v) < 0.2max(\mathbf{PSD}(v)))$ [4]
F27	Maximal frequency $\max_v(\mathbf{PSD}(v) < 0.2max(\mathbf{PSD}(v)))$ [4]
F28	Gyration radius $\sqrt{\frac{m_3}{m_2}}$ m_i is the ith moment [4]
F29	Spectral centroid width $\sqrt{{F_8}^2 - {F_{28}}^2}$ [4, 111]
F32-33	Ratio F20:F30, ratio F20:F31 [4]
F34, 35	No. peaks in temporal evolution of the DFTs central and max frequency [4]
F36, 37	Distance of the temporal evolution of DFTs max and mean or median frequency [4]
F38–40	Distance between (1/4, or 3/4 and median), and (1/4, and 3/4) of all DFTs [4]
F41–44	2dCrossCorAbDWT with $\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$ [110]
F45–48	2dCrossCorRealDWT with $\mathbf{q}(t)$, $\mathbf{r}(t)$, $\mathbf{ea}(t)$, $\mathbf{n}(t)$ [110]

Table 4.4: *Cont.*

Cepstrum Feature	
C1	STD $\mathbf{ce}(v)$ [12]
C2	Skewness of $\mathbf{ce}(v)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{ce}(v_i)-\mu}{\mathrm{STD}(\mathbf{x})}\right)^3$ [12]
C3	Kurtosis of $\mathbf{ce}(v)$ $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{ce}(v_i)-\mu}{\mathrm{STD}(\mathbf{x})}\right)^4$ [12]
C4	Max value of $\mathbf{ce}(v)$ $max(\mathbf{ce}(v))$ [12]
C5	Cepstrum No. peaks (echo) $\mathbf{ce}(v)$
Acoustic Feature [15]	
A1	No. peaks Linear prediction filter coefficients
Polarity Feature [4]	
P1	Max eigenvalue of the covariance matrix
P2-3	Incidence angle, Polarization azimuth
P4-5	Degree of linear, Plane polarization

This section evaluates the filter, wrapper, and embedded feature selection methods from feature selection library FSLib available at (`https://www.mathworks.com/matlabcentral/fileexchange/56937-feature-selection-library`) accessed on 8 January 2022 [107]. For each class of signals, a sorted feature vector (high to low rank) is generated by each FSLib feature selection approach. The optimal feature space subset $\mathcal{O}$ and $|\mathcal{O}| \in [1, ..., K]$ is selected using ten-fold cross-validation at the classification training stage. The optimal $\mathcal{O}$, for each of the five used classifiers (RF, SVM, GLR, normGLR, and GTV), is the one that results in the highest validation accuracy for each classifier. If the highest accuracy score corresponds to several different feature subsets, then the optimal subset is chosen as the one with the minimum dimension. The feature selection result via FSLib shows the optimal feature selection approaches, which are embedded FSV ($|\mathcal{O}| = 45$ for rockfall, and $|\mathcal{O}| = 21$ for earthquake, i.e., 45 and 21 features are selected for these two classes, respectively), filter mutinfffs ($|\mathcal{O}| = 38$ for slide quake, and $|\mathcal{O}| = 53$ for noise)) and $\mathcal{O}$ is common to all 5 classifiers. Similarly, for the graph-based feature weight optimisation (GLR, normGLR, and GTV), the ten-fold cross-validation is adapted with training data to tune the

parameters in Alg. 4.2 where $\gamma = 119$, and the optimal feature weight $\frac{\mathbf{1}}{\mathbf{2}\widehat{\sigma}^{\mathbf{2}}}$ is the one which results in the best classification accuracy. Since the optimal feature weight is observed to be similar for all three classifiers, the average value over 10 folds from all three classifiers is presented.

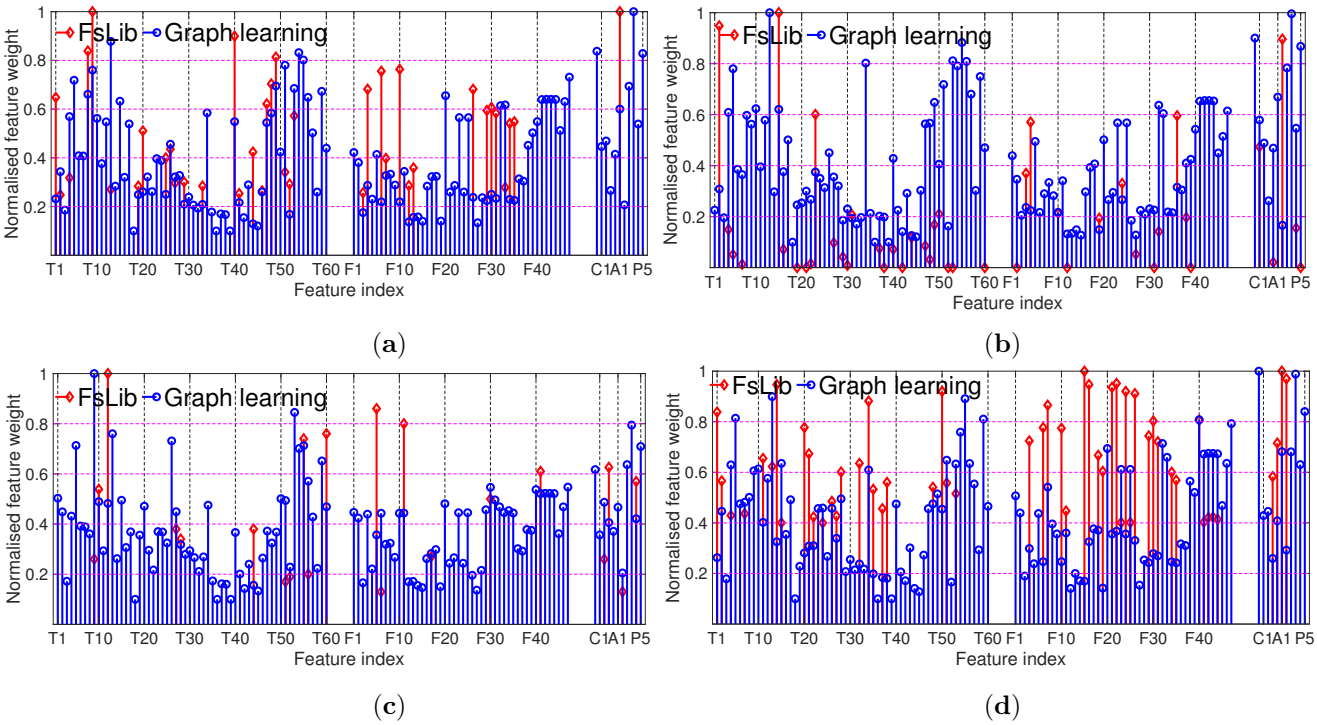


Figure 4.6: Normalised feature weights from the resulting set $\mathcal{O}$ for all 4 classes. Features of high importance include temporal (T), Spectral (F), Cepstrum (C), Acoustic (A), Polarity (P). (**a**) Feature weight for rockfall. (**b**) Feature weight for slide quake. (**c**) Feature weight for earthquake. (**d**) Feature weight for noise.

The resulting set $\mathcal{O}$ for both FSLib feature selection and graph-based feature weight optimisation (graph learning) for each class is shown in Fig. 4.6, in red and blue colour, respectively. The feature weights are normalised $[0, 1]$. It can be observed that a good mix of temporal, spectral, and cepstral features is selected for all classes. FSLib tends to discard the less discriminative features, unlike graph-based feature weight optimisation, which does not discard features but tends to give a much higher weight to the discriminative features, with the highest

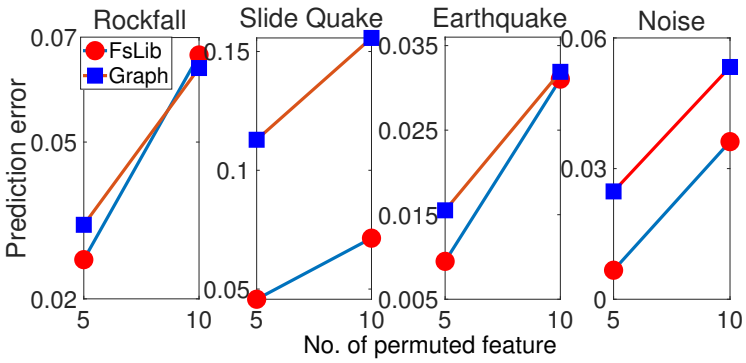


Figure 4.7: Permutation Feature Importance. The horizontal axis denotes the feature number starting from the most important feature for each of the four classes. The vertical axis is the decrease in sensitivity measure Δse_i

feature weight at 1 and the least discriminative feature weight around 0.1. The selected feature set $\mathcal{O}$ for FSLib and graph-based feature weight optimisation is very similar.

Next, the feature selection performance is evaluated via the permutation feature importance method by looking at how fast the prediction error increases after permuting the features. The larger the gradient, the higher the importance of the permuted features. Fig. 4.7 shows the results with the normGLR classifier. The random permutation of 5 and 10 most important features are considered (as per Fig. 4.6) that correspond to numbers 5 and 10, in the horizontal axis. Specifically, 1) with the highlighted optimal subset of features $\mathcal{O}$ both FSLib and graph-based feature weight optimisation, the top i ($i = 5$ or 10) features are permuted randomly forming a new feature set; 2) then, classification is performed with the newly formed feature sets, resulting in sensitivity score se_i; finally, 3) the prediction error is calculated as $\Delta se_i = se_0 - se_i$, where se_0 denotes the sensitive score without any permutation. The values Δse_i are then plotted in Fig. 4.7 against the number of permuted features i. Since the processing involves random permutations, se_i is obtained after averaging the result of 10 runs.

As shown in Fig. 4.7, permuting the selected features by both FSLib and

graph-based feature weight optimisation results in an increase in Δse_i, especially for slide quake, which indicates the effectiveness of the feature selection process. Furthermore, comparing graph-based feature weight optimisation and FSLib, it can be concluded that the highly ranked features by graph-based feature weight optimisation are more discriminative because they generally correspond to higher prediction errors due to the removal of these highly ranked features.

4.3.3 Classification

This section first evaluates the effect of the feature mentioned above engineering approaches (extraction via PCA, selection via FSLib, and graph-based feature weight optimisation) on classification performance for the classifiers of interest, considering only catalogued events. Afterward, the classification results for the proposed workflow are presented, as illustrated in Fig. 6.2 for continuous data.

Effect of Feature Engineering on Classification Performance

Case Study 1: Feature Engineering with SZ10 Dataset

Table 4.5 shows the classification performance expressed as F1-score, where 1 is the best performance, and 0 is the worst. The cases comparison includes when the complete set of 48 features is used (i.e., without any feature selection), non-iterative Inf-FS feature selection followed by the SVM or RF classifier, as well as the method of [123], with the proposed iterative alternating graph feature weight optimisation with two graph-based classifiers. It can be observed that there is no loss of performance when using feature selection compared to the complete set of features. Secondly, the proposed graph-based alternating classifiers and feature learning for all four (micro)seismic event classes provide the most accurate classification result. The improvement is especially significant for the slide quake class. The reason that benchmark [123] performs slightly worse than the proposed

method may potentially be because it does not promote smoothness within the label signal.

Table 4.5: Classification Result (F1-score)

	Classifier	Earthquake	Slide quake	Tremor	Calibration Shot
Full set of Features Table 4.3	SVM	0.78	0.78	0.91	0.9
	RF	0.86	0.9	0.89	0.87
	GLR	0.87	0.86	0.89	0.61
	normGLR	0.87	0.92	0.9	0.81
Inf-FS	SVM	0.9	0.84	0.92	0.8
	RF	0.9	0.88	0.91	0.9
[123]	[123]	0.91	0.91	0.88	0.8
Alg. 4.2	GLR	0.94	**0.95**	0.93	0.85
	normGLR	**0.95**	0.93	**0.95**	**0.92**

The bold content represents the highest F1 score for each class.

Table 4.6: Reproduced F1 score of Table 4.5. Classes 1, 2, 3, 4 correspond to Earthquake, Slide quake, Tremor, and Calibration Shot, respectively.

GLR	run1	run2	run3	run4	run5	run6	run7	run8	run9	mean	std	CV (%)	Table 4.5
class1	0.92	0.94	0.93	0.95	0.94	0.93	0.94	0.9	0.94	**0.93**	0.014	1.51	**0.94**
class2	0.95	0.94	0.93	0.95	0.93	0.94	0.93	0.94	0.92	**0.94**	0.011	1.17	**0.95**
class3	0.95	0.93	0.97	0.95	0.91	0.93	0.95	0.91	0.93	**0.93**	0.019	2.04	**0.93**
class4	0.88	0.85	0.85	0.81	0.81	0.85	0.88	0.85	0.85	**0.85**	0.022	2.59	**0.85**
normGLR	run1	run2	run3	run4	run5	run6	run7	run8	run9	mean	std	CV (%)	Table 4.5
class1	0.94	0.93	0.94	0.96	0.96	0.94	0.94	0.97	0.95	**0.95**	0.013	1.37	**0.95**
class2	0.94	0.92	0.92	0.93	0.92	0.94	0.93	0.93	0.92	**0.93**	0.008	0.86	**0.93**
class3	0.96	0.96	0.97	0.99	0.96	0.96	0.97	0.96	0.95	**0.96**	0.011	1.15	**0.95**
class4	0.96	0.96	0.96	0.96	0.96	0.92	1	0.88	0.92	**0.94**	0.033	3.51	**0.92**

The bold content represents the reproduced F1 score for each class.

Reproducibility and Reliability. Since the training set was randomly generated and the randomisation seed was not fixed, this section tests the reproducibility and reliability of the results obtained with graph-based classifiers with Alg. 4.2, as illustrated in Table 4.6. CV stands for the coefficient of variation (i.e., the ratio of the standard deviation to the mean), which shows the extent of variability concerning the mean of the population. Thus, the higher CV represents the greater dispersion. The results of 10 experimental runs are shown in

Table 4.6. It can be concluded that the difference between repeated results and the initial results is within-run precision. To assess reliability, it can be observed that CV for the experimental runs is under 5% and conclude that the GLR and normGLR models are reproducible and reliable.

Case Study 2: Feature Engineering with the Dataset SZ13, SZ14 and SZ15

Table 4.7: Sensitivity (mean and standard deviation (STD))

Classifier	Rockfall	Slide Quake	Earthquake	Noise
RF (PCA)	0.84(0.04)	0.76(0.05)	0.95(0.02)	0.80(0.03)
SVM (PCA)	0.85(0.04)	0.86(0.04)	0.96(0.02)	0.80(0.03)
Constructed feature set (see Table 4.4)				
RF	0.89(0.03)	0.84(0.04)	0.96(0.02)	**0.83(0.03)**
SVM	0.87(0.03)	0.86(0.04)	0.96(0.03)	0.82(0.03)
GLR	0.89(0.02)	0.85(0.03)	0.96(0.01)	0.70(0.03)
normGLR	0.88(0.02)	0.84(0.02)	0.96(0.01)	0.75(0.02)
GTV	0.89(0.01)	0.85(0.03)	0.96(0.01)	0.69(0.03)
Feature selection via FsLib				
classifier	Rockfall	Slide Quake	Earthquake	Noise
RF	0.88(0.03)	0.84(0.04)	0.96(0.02)	**0.83(0.03)**
SVM	0.89(0.03)	0.86(0.04)	0.96(0.02)	0.80(0.04)
GLR	0.91(0.01)	0.85(0.03)	**0.97(0.01)**	0.72(0.02)
normGLR	**0.92(0.02)**	**0.88(0.02)**	**0.97(0.01)**	0.79(0.02)
GTV	0.84(0.05)	0.84(0.04)	0.96(0.01)	0.75(0.02)
Feature selection via graph-based feature weight optimisation (see Alg. 4.2)				
classifier	Rockfall	Slide Quake	Earthquake	Noise
GLR	**0.92(0.01)**	**0.88(0.02)**	0.96(0.01)	0.76(0.02)
normGLR	0.91(0.01)	**0.88(0.02)**	**0.97(0.01)**	0.80(0.02)
GTV	**0.92(0.01)**	0.87(0.02)	0.96(0.01)	0.75(0.02)

The bold content represents the highest Sensitivity/Recall for each class.

With only catalogued event segments, this section adopts the one-against-all classification strategy with ten-fold cross-validation. The dataset was randomly split for training and testing for each class with a 70:30 ratio. The test is carried out 50 times under identical conditions to ensure the repeatability of results, and the mean and STD of the Sensitivity measure are shown in Table 4.7.

As in [4], the classification performance is presented with the sensitivity measure, equivalent to Recall, that is the ratio of correct events predicted over the total number of catalogued events for each class(see Eq. 3.13 and Table 3.4). The confusion matrixes are presented to explain misclassification, as shown in Table 4.8 & 4.9 utilising the constructed features without and with selection steps.

Table 4.8: Mean Confusion Matrix (Sensitivity) for 50 runs (feature construction only) [a]

classifier		Pred.Rockfall	Pred.Slide Quake	Pred.EQ	Pred.Noise
RF (PCA)	Ref.Rockfall	**0.84**	0.02	0.01	0.13
	Ref.Slide Quake	0.11	**0.76**	0.00	0.13
	Ref.EQ	0.02	0.01	**0.95**	0.02
	Ref.Noise	0.11	0.07	0.02	**0.80**
SVM (PCA)	Ref.Rockfall	**0.85**	0.02	0.00	0.13
	Ref.Slide Quake	0.05	**0.86**	0.01	0.08
	Ref.EQ	0.02	0.01	**0.96**	0.02
	Ref.Noise	0.10	0.08	0.02	**0.80**
RF	Ref.Rockfall	**0.89**	0.01	0.01	0.09
	Ref.Slide Quake	0.06	**0.84**	0.00	0.09
	Ref.EQ	0.02	0.02	**0.96**	0.01
	Ref.Noise	0.10	0.07	0.01	**0.83**
SVM	Ref.Rockfall	**0.87**	0.02	0.00	0.11
	Ref.Slide Quake	0.05	**0.86**	0.00	0.09
	Ref.EQ	0.01	0.02	**0.96**	0.01
	Ref.Noise	0.09	0.07	0.01	**0.82**
GLR	Ref.Rockfall	**0.89**	0.02	0.01	0.08
	Ref.Slide Quake	0.06	**0.85**	0.02	0.07
	Ref.EQ	0.02	0.01	**0.96**	0.02
	Ref.Noise	0.17	0.11	0.03	**0.70**
normGLR	Ref.Rockfall	**0.88**	0.02	0.01	0.09
	Ref.Slide Quake	0.06	**0.84**	0.01	0.09
	Ref.EQ	0.01	0.05	**0.96**	0.02
	Ref.Noise	0.15	0.09	0.02	**0.75**
GTV	Ref.Rockfall	**0.89**	0.02	0.02	0.07
	Ref.Slide Quake	0.06	**0.85**	0.03	0.07
	Ref.EQ	0.02	0.01	**0.96**	0.02
	Ref.Noise	0.15	0.10	0.05	**0.69**

[a] The predicted (Pred.) events are represented concerning the events of the reference catalogued (Ref.), EQ stands for the earthquake. The bold content represents the percentage of corrected classified events for each class.

Table 4.9: Mean Confusion Matrix (Sensitivity) averaged over 50 runs (FsLib and graph-based feature weight optimisation) [a]

Feature selection via FsLib					
classifier		Pred.Rockfall	Pred.Slide Quake	Pred.EQ	Pred.Noise
RF	Ref.Rockfall	**0.88**	0.01	0.01	0.09
	Ref.Slide Quake	0.07	**0.84**	0.00	0.09
	Ref.EQ	0.02	0.01	**0.96**	0.01
	Ref.Noise	0.09	0.08	0.00	**0.83**
SVM	Ref.Rockfall	**0.89**	0.01	0.01	0.09
	Ref.Slide Quake	0.04	**0.86**	0.01	0.09
	Ref.EQ	0.01	0.01	**0.96**	0.01
	Ref.Noise	0.10	0.08	0.01	**0.80**
GLR	Ref.Rockfall	**0.91**	0.02	0.01	0.07
	Ref.Slide Quake	0.05	**0.85**	0.01	0.10
	Ref.EQ	0.01	0.01	**0.97**	0.01
	Ref.Noise	0.14	0.13	0.01	**0.72**
normGLR	Ref.Rockfall	**0.92**	0.01	0.00	0.06
	Ref.Slide Quake	0.04	**0.88**	0.00	0.07
	Ref.EQ	0.01	0.01	**0.97**	0.01
	Ref.Noise	0.11	0.10	0.01	**0.79**
GTV	Ref.Rockfall	**0.84**	0.02	0.06	0.08
	Ref.Slide Quake	0.05	**0.84**	0.04	0.07
	Ref.EQ	0.01	0.02	**0.96**	0.01
	Ref.Noise	0.11	0.10	0.05	**0.75**
Feature selection via graph-based feature weight optimisation					
classifier		Pred.Rockfall	Pred.Slide Quake	Pred.EQ	Pred.Noise
GLR	Ref.Rockfall	**0.92**	0.01	0.00	0.06
	Ref.Slide Quake	0.07	**0.88**	0.00	0.05
	Ref.EQ	0.03	0.01	**0.96**	0.01
	Ref.Noise	0.14	0.09	0.01	**0.76**
normGLR	Ref.Rockfall	**0.91**	0.01	0.00	0.08
	Ref.Slide Quake	0.06	**0.88**	0.00	0.06
	Ref.EQ	0.01	0.01	**0.97**	0.01
	Ref.Noise	0.11	0.08	0.01	**0.80**
GTV	Ref.Rockfall	**0.92**	0.01	0.00	0.07
	Ref.Slide Quake	0.08	**0.87**	0.01	0.04
	Ref.EQ	0.03	0.01	**0.96**	0.01
	Ref.Noise	0.15	0.08	0.02	**0.75**

[a] The predicted (Pred.) events are represented concerning the events of the reference catalogued (Ref.), EQ stands for the earthquake. The bold content represents the percentage of corrected classified events for each class.

RF classifier with feature construction and selection was implemented with the catalogued dataset as in [4]. Sensitivity results in [4] (rockfall 0.94, slide quake 0.93, earthquake 0.94, and noise 0.92) were provided for a balanced testing set comprising 70 events per class, whilst the testing set contains 30% of the catalogued events for each class, which is a more realistic scenario. The 71 constructed features utilised for classification included nine network geometry attributes, such as the station with the highest SNR, which was not considered since it lacks generalisation. It is noted that the mean sensitivity over all classes falls to 0.9 without these network geometry attributes, ranging from 0.86 to 0.94. Our replicated RF results, as observed on the RF with FSLib selected features in Table 4.7, are in agreement.

As shown in Table 4.7 and Table 4.8, there is no performance benefit performing PCA for either RF or SVM classifier compared to using constructed features. The only advantage is lower complexity due to dimensionality reduction: 119 features vs. 44 principal components fed to the classifier. Indeed, with only constructed features fed to the classifier, all classifiers have similar performance. However, the additional FSLib feature selection step does not appear to improve the performance of RF or SVM, but it does significantly improve the performance of graph-based classifiers. The benefit of feature selection for SVM and RF is dimensionality reduction by more than 50% of the feature set per class from 119 to $|\mathcal{O}| = 45$ for rockfall, 38 for slide quake, 21 for earthquake, and 58 for noise. As expected, it can be found that better performance improvement for feature selection via the proposed graph-based feature weight optimisation for the graph-based classifiers. Thus, feature selection is considered a beneficial step for classification for performance improvement and complexity reduction. The performance improvement due to proposed Alg. 4.2 can be explained via the confusion matrix shown in Table 4.9. Compared to Table 4.8, for rockfall, the more discriminate feature selection shows that rockfall and slide quake are not

confused with earthquake anymore.

The catalogued nature/anthropogenic noise events are miscellaneous, caused by human-made activities, including footsteps to environmental conditions such as storms, with relatively not-very-distinct features compared to the other 3 classes, as shown in Fig. 4.6. Thus, the noise class has the worst sensitivity, but even noise performance is improved with graph-based feature weight optimisation since noise signals are less likely to be confused with rockfalls. Earthquakes usually have high SNR, with distinct P & S wave arrivals, which make them less likely to be confused with any other class, resulting in around 0.96 sensitivity for all classifiers, with or without feature selection. As stated in [4], and as observed in Table 4.9, it is sometimes difficult to distinguish slide quake from small-volume rockfall, rockfall, and footstep (noise).

Performance Comparison of the Graph-Base Classifiers Against the Others

The performance of graph-based classifiers (norm)GLR and GTV are highly dependent on the graph kernel bandwidth σ in Eq. 3.10. A too-small value of the bandwidth would lead to a poor representation of the local structures; conversely, a considerably high value could result in a coarse description of the data. As illustrated in Table 4.7, with constructed features, graph-based classifiers have similar performance to RF and SVM for rockfall, slide quake, and earthquake, when graph kernel bandwidth is set as in [81]. However, with optimised feature weights obtained with benchmarked feature selection and graph-based feature weight optimisation, the potential of graph-based classifiers is maximised (i.e., the classification sensitivity increases by about 3% for rockfalls and slide quakes). For instance, the classification sensitivity for rockfalls with GLR classifiers reaches 92% (graph-based feature weight optimisation), 91% (benchmarked feature selection FSLib) compared to 89% (constructed feature only). In con-

clusion, graph-based classifiers outperform RF and SVM with appropriate graph kernel bandwidth (feature weight).

Classification with Multi-Channel Detection (Alg. 4.1) on Continuous Data Recorded (24-28/November/2014)

Table 4.10: Classification results of the catalogued events in [4] from continuous data [a]

	Pred.Rockfall	Pred.Slide Quake	Pred.EQ	Pred.Noise
Ref.Rockfall (65)	**60 (0.92)**	0	0	5
Ref.Slide Quake (18)	1	**15 (0.83)**	0	2
Ref.EQ (23)	0	0	**23 (1.00)**	0
Ref.Noise (13)	1	1	0	**11 (0.79)**

[a] The predicted (Pred.) events are represented with respect to the events of the reference catalogued (Ref.). EQ stands for earthquake. The bold content represents the number and the percentage of corrected classified events for each class.

This section looks at the more realistic scenario of classifying events from the continuous (vs. only catalogued) dataset, which includes detection misses and uncatalogued events as discussed in Section 4.3.1. The selected period 24-28/November/2014, contains the most catalogued events, with 65 rockfalls, 18 slide quakes, 23 earthquakes, and 14 noise. After detection, as per the proposed workflow of Fig. 6.2, only 1 noise event was missed, and detected 614 uncatalogued events. With the 119 features constructed as shown in Table 4.4 for each event detected by Alg. 4.1, the classification training set is formed as described in Section 4.3.3, without the catalogued event in the selected period. Specifically, the feature weight obtained by the graph-based feature weight optimisation is utilised to distinguish the catalogued events with the normGLR classifier, resulting in classification performance indicated as a confusion matrix in Table 4.10. Performance is similar to that observed in Table 4.9.

As described in Section 4.3.1, the expert manually evaluates the result after detecting (micro)seismic events from the continuously recorded signal. Here the

Table 4.11: Classification results of the additional manually expert-verified events from continuous data [a]

	Pred.Rockfall	Pred.Slide Quake	Pred.EQ	Pred.Noise	Pred.Others
Ref.Rockfall (71)	**60 (0.85)**	0	0	10	1
Ref.Slide Quake (21)	1	**15 (0.71)**	0	4	1
Ref.EQ (25)	0	0	**24 (0.96)**	1	0
Ref.Noise (616)	32	2	65	**257 (0.42)**	260

[a] The predicted (Pred.) events are represented with respect to the events of the reference catalogued (Ref.). EQ stands for earthquake. The bold content represents the number and the percentage of corrected classified events for each class.

performance of the adapted graph-based feature weight optimisation and classification is assessed with expert verification over the selected period. The results are shown in Table 4.11, where another class considered not belonging to the four initial types of events is introduced. The adapted feature weight optimisation and classification workflow are resilient to rockfall and slide quake with only a minor reduction in classification sensitivity, according to Table 4.11. Additionally, it is no surprise that the classification results for the earthquake are unaffected; however, the situation is different for noise-type occurrences, which could be compensated with classification post-processing.

4.4 Feature Recommendation

Following the comprehensive classification performance evaluation, it can be observed that feature selection, rather than feature extraction via PCA, led to effective dimensionality reduction of the constructed feature set and performance improvement for different classes and classifiers. This section discusses which handcrafted optimised feature sets characterise slope failure endogenous events, including rockfalls, slide quakes, and earthquakes. The following content lists the standard highlighted features observed by recent research and the proposed graph-based feature weight optimisation in bold. The notations of the features correspond to the ones in Table 4.4.

With the RF classifier, the distinguishing attributes or features of the four classes in datasets SZ13, SZ14, and SZ15 are identified as the following (without distinguishing unique features per class): Duration (T1), Ratio between ascending and descending time (**T8&9**), Energy in the first third part of the autocorrelation function (T38), Energy of the signal filtered in $50-100$ Hz (T31), Mean and max of the DFT (**F1&2**), Central Frequency of the 2nd quartile (**F4**), Energy in ([0, 1/4], [1/4, 1/2])*Fs (F12&13), Frequency at the max (F27), No. Peaks in the curve showing the temporal evolution of the DFTs max, mean and median (F20&30, **F31**), Ratio between F20 and F31 (**F33**), No. Peaks in the curve of the temporal evolution of the DFTs central frequency (**F34**) and Polarization azimuth (**P3**) [4].

In the attempt to perform classification of slope failures, earthquakes, and noise on continuous data via an RF classifier, the following eight most distinct features (without distinguishing unique features per class) are highlighted: Spectral gyration radius (F28), Spectral centroid (F8), Central frequency of the first quartile (F3), Variance of the normalised DFT (F10), Frequency at the maximum of the DFT (F27), Frequency at spectrum centroid (F29), Energy of the last two-thirds of the autocorrelation function (T39) and Energy of the seismic signal in the frequency band of $1-3$ Hz (**T28**) [28].

Furthermore, for similar terrain and slope failures, through visual observation without automatic feature selection and classification, qualitatively, the distinguishing features for each of the four classes of interest are characterised [14]. Next, the distinguishing attributes identified by [14] are presented, and we attempt to map them into the equivalent notation used in Table 4.4. *Rockfalls*: the falling block impacts produce spikes or jolts in the waveforms, which are visible both in the signal waveform as cigar shapes (final impact) (**T15**, T18, **T33**, and **T34**) and in the power spectral density function for most of the events (**T17**). *Slide quakes*: short-duration (T1) (last less than 5 second) earthquake-like sig-

nals, with clearly discernable, trackable wave packets, emergent first arrivals, and undistinguishable P and S waves (**T2**, **T15**, **T22**, T33-36, and **T37**). *Earthquakes*: well-studied and potential landslide triggers that produce medium to long-duration signals (**T1**, **T9&10**, T38&39, and T41&42) with distinct P and S wave high impact arrivals. *Natural/Anthropogenic noise*: high-frequency range ($> 50\,\text{Hz}$) (**F12**, F13&14, **F16**) characteristics due to the shallow installation of seismometers in clayey materials; duration, phase, and velocities of noise signals are not identifiable; furthermore, the noise waveform amplitude attenuation patterns are incoherent.

While the discriminative features identified for landslide-induced events and earthquakes in [4, 14, 28] are not the same, they all show that temporal and spectral attributes are the most important for RF classifiers. More importantly, the aforementioned features this section also observes via the learning algorithms are highlighted in bold above. Additionally, it can be observed that more discriminative features that improve the classification performance of RF, SVM, and graph-based classifiers, highlighted in Fig. 4.6 will list next, per class.

4.4.1 Feature Weight Analysis for Target Signals

Rockfall: the highly ranked features observed via FSLib and graph-based feature weight optimisation are: cross correlation-based features (T19-20, T25-27, T40, T47-51, T53-60, F39-48 (graph-based feature weight optimisation)), Energy of the signal filtered in $10-50\,\text{Hz}$ (T29), dominant frequency (F7 FSLib), duration (T1 FSLib) No. of peaks in the curve showing the temporal evolution of the DFTs mean/median (F30&31 FSLib), waveform amplitude attenuation patterns (T4-7, T12&13 (graph-based feature weight optimisation)). The waveforms of rockfalls are variable due to the loose material saltation and flow combined with the moving character of the source.

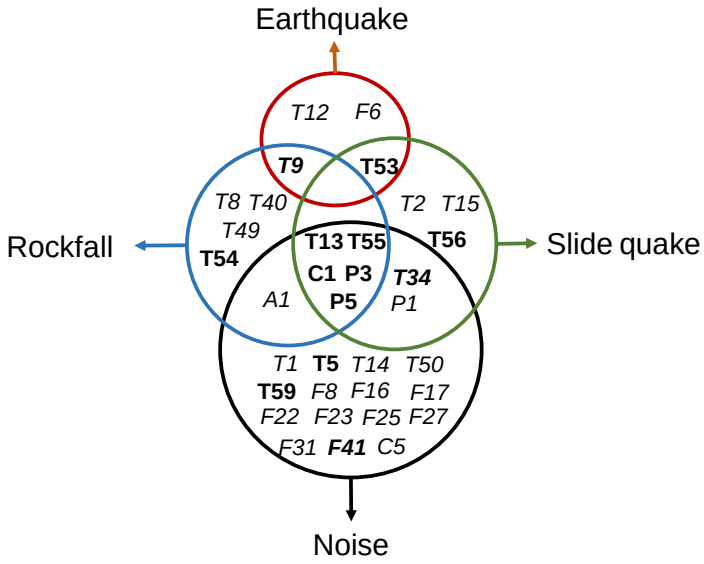


Figure 4.8: Features that best characterise rockfall, slide quake, earthquake, and environmental/anthropogenic noise. Graph-based feature weight optimisation features in bold, FSLib features in italics, and common features learnt by both graph-based feature weight optimisation and FSLib in bold italics.The notations of the features correspond to the ones in Table 4.4.

Slide quake: FSLib highlights the Optimal point of separation (T4), the max value of cross correlation with template earthquake (T16), the ratio between mean and median envelope signal (T23), and the max value of enveloped Power Density Function (F5); while graph-based feature weight optimisation provided a high rank for the max value of cross correlation with template slide quake, rockfall, earthquake, (T12&13, T16), spectral features max envelope power density function (F5) and No. of peaks of autocorrelation (F11).

Earthquake: here, both FSLib and graph-based feature weight optimisation approaches provide more focused additional importance features: Dominant frequency (F7), cross correlation-based features (T53-60 (graph-based feature weight optimisation)), No. of peaks in the curve of the temporal evolution of the DFTs central and maximum frequency (F34&35 graph-based feature weight optimisation), cross correlation-based features (F39-48 (graph-based feature weight optimisation)) to capture the typical triangular-shaped sonogram pattern for earth-

quakes.

Natural/Anthropogenic noise: due to variability of the source of noise events $|\mathcal{O}| = 53$ for FSLib, with abundant complex spectral features such as F17-44, while graph-based feature weight optimisation results in high ranking temporal features T1-10 and cross correlation-based features (T47-60).

Additionally, the following features are highly ranked for all four classes: Cepstral features C1 (standard deviation) and C3 (kurtosis); Acoustic feature A1 (No. of peaks Linear prediction filter coefficient); Polarity features P2-5 (incidence angle, polarization azimuth, degree of linear polarization, degree of plane polarization).

The most distinguishing features are summarised, defined as those with normalised feature weight > 0.8 (from Fig. 4.6), in Fig. 4.8.

4.5 Summary

This chapter addressed the challenges in automatically and accurately detecting and classifying slope failure endogenous events rockfall and slide quake (seismic sources related to landslide processes) and externally sourced earthquake and Natural/Anthropogenic noise. An automated workflow is proposed, Fig. 6.2, to analyse large amounts of seismometer data from multiple channels/sensors. This chapter also presents a detection scheme that identifies potential events and the representative signal segments from various channels in Alg. 4.1 to tackle continuous data containing interfering signals and low SNR slide quake events, without missing many catalogued events and detecting fewer uncatalogued events than processing a single channel would. Next, the graph-based classification with graph-based feature weight optimisation for characterising all four events, Alg. 4.2 has been proposed and evaluated, which are robust to lack of sufficient balanced data for training. Finally, after comprehensive experiments to demonstrate the

impact of feature engineering on classifiers depending on handcrafted features, this chapter provides a detailed list of critical features to consider for each of the four types of events. Additionally, the proposed workflow is evaluated on the continuous raw data recorded at the selected period (24-28/November/2014), manually labelled by an expert, where detection and classification with graph-based feature weight optimisation obtained promising results.

However, certain limitations in the proposed system could be addressed in future work. First, the effectiveness of the used thresholding strategy depends on the consistency of the background noise distribution; hence certain low amplitude events are nevertheless missed by our detection scheme. Furthermore, due to their complicated and changing generating mechanism, the Natural/Anthropogenic noise categorisation findings need to be enhanced. Future research will therefore focus on adaptive background noise removal to increase SNR, identifying concealed microseismic activity, and using post-processing to improve the classification outcome, particularly for natural/anthropogenic noise.

In addition, Chapter 3 and 4 construct the handcrafted features, which concluded with expert knowledge and investigated the significance of the features for various seismic signals. In comparison, these handcrafted features might be less optimal for signal representation. Thus, Chapter 6 explores the effect of deep learning-based signal representation learning on (micro)seismic classification. Before that, Chapter 5 briefly introduces the interpretability of the proposed graph-based feature weight optimisation and classification model.

Chapter 5

Graph-Based Model Interpretability: Case Study of (Micro)seismic Signal Classification

Chapter 4 proposes a graph-based feature weight optimisation and classification approach, and the performance is evaluated with the signal recorded at the on-going landslide. The experimental results demonstrate the competitiveness of the proposed system w.r.t state-of-the-art in seismic signal classification. This chapter explores the interpretability of the proposed model in Chapter 4 with the commonly used techniques for model interpretability: Shapley Values, Local Interpretable Model-Agnostic Explanation (LIME), and Cross-correction Feature Selection (CFS).

5.1 Introduction

Numerous scientific and industrial applications, such as the investigation of (micro)seismic occurrences, now have access to robust predictive models owing to machine learning [126]. Gaining model confidence without sacrificing predictive

capability has become increasingly important as the models are used for high-stakes autonomous assessments [126].

Data science that can be interpreted and explained refers to techniques that enable humans to comprehend the behavior and predictions of models. Research, business, and governments are all interested in making machine learning models more understandable (see, [127–130]).

Data visualisation, which provides insights into the underlying structure of the datasets, is frequently the first step in interpretability research. Various dimensionality reduction techniques are given to project the high-dimensional feature space into 2 or 3 dimensions without losing the high-dimensional link between data points since high-dimensional data are challenging to visualise immediately. By converting the distances between each pair of samples in the high-dimensional space into probabilities that the pair should be closely placed in the projected low-dimensional space, the widely used local dimensionality reduction method known as t-distributed Stochastic Neighbour Embedding (t-SNE) preserves the local structure of the dataset [131]. Additionally, Shapley values and LIME are thought to be two of the most often utilised methods for model interpolation [132, 133]. LIME and Shapley values are surrogate models using the black-box machine learning models, which are model agnostic. Both approaches experiment with variations in prediction with a modest input adjustment. This adjustment must be minimal (or in the local region) to maintain a close relationship with the original data point. LIME and Shapley values models are substitutes that simulate modifications to the forecast (on the changes in the input).

This chapter explores the interpretability of the proposed graph-based feature weight optimisation and classification (see Chapter 4) utilising the obtained feature weight and the commonly used Interpretability Methods LIME and Shapley values.

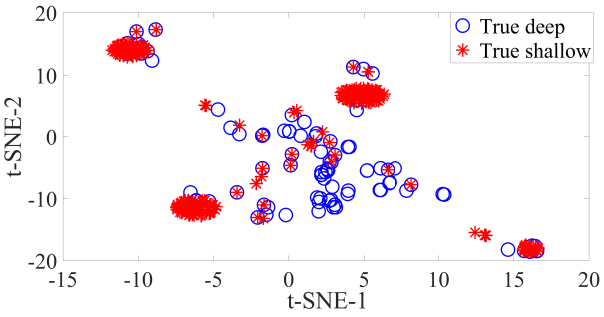
5.2 Methodology

LIME. Mathematically, local surrogate models with interpretability constraint defined as explanation$(x) = \arg\min_{\hat{m}\in\hat{M}} f\left(\Theta, \hat{m}, \pi_x\right) + \Omega(\hat{m})$, where the model $\hat{m}$ (e.g., linear regression model) is the explanation model, for instance, x, that minimises loss function $f()$ (e.g., mean squared error). The loss function measures how close the explanation is to the prediction of the original model Θ (e.g., an xgboost model), while the model complexity $\Omega(\hat{m})$ is kept low (e.g., prefer fewer features). $\hat{M}$ is the family of possible explanations, for example, all possible linear regression models. The proximity measures π_x defines how large the neighborhood around the considered instance x is.

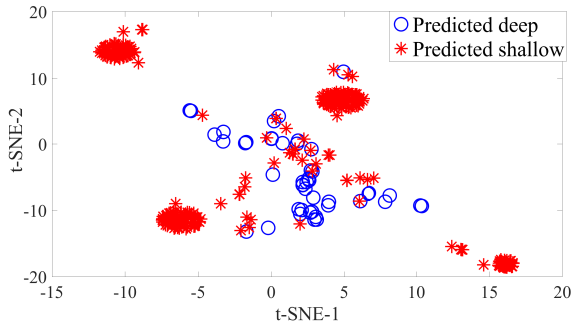
Shapley value. The Shapley value of a feature value j is its contribution to the payout, weighted and summed over all possible feature value combinations: $\phi_j(val) = \sum_{\mathcal{O}\subseteq\{1,\ldots,K\}\setminus\{j\}} \frac{|\mathcal{O}|!(K-|\mathcal{O}|-1)!}{K!}(\text{val}(\mathcal{O}\cup\{j\}) - \text{val}(\mathcal{O}))$, where ! denotes factorial, $\mathcal{O}$ is a subset of the features used in the model, and K the number of features. val$(\mathcal{O})$ is the prediction for feature values in set $\mathcal{O}$ that are marginalised over features that are not included in set $\mathcal{O}$: $\text{val}_{\mathbf{f}(i)}(\mathcal{O}) = \int \hat{f}\left(f_1(i), \ldots, f_K(i)\right) d\mathbb{P}_{\mathbf{f}(i)\notin\mathcal{O}} - E_{\mathcal{F}}(\hat{f}(\mathcal{F}))$. $\mathcal{F}$ is the set of features, $\mathbf{f}(i)$ is the vector of feature values of the instance i to be explained, $\hat{f}(\mathbf{f}(i))$ is the prediction, and $E_{\mathcal{F}}(\hat{f}(\mathcal{F}))$ is the average of the prediction.

5.3 Results and Discussion

To assess the interpretability, the proposed graph-based feature weight optimisation and classification model is evaluated with the seismic dataset of [1], which contains a catalogue of microearthquakes induced by an underground cavern collapse recorded after October 2013 in south Louisiana. Compared with the datasets described in Section 1.3.1, this dataset contains only two classes with sufficient events. The dataset comprises events characterised by 40 features con-



(a) t-SNE feature projection with true labels



(b) t-SNE feature projection with graph learning

Figure 5.1: Seismic dataset visualisation with t-SNE.

structed in time, frequency, and time-frequency domain: (i) 143 deep events (microearthquake in the salt body) located at a depth between $1.0-2.0$ km; (ii) 297 shallow events (microseismic events in the cap rock) at depths between $40-400$ m.

This section explores the feature importance to explain some (miss)classifications by applying the graph spectral classifier normGLR on handcrafted features and tabular data (40 created features). The feature importance assignment can then be evaluated using expert knowledge to determine its meaning and whether it may have resulted in poor decisions.

Fig. 5.1(a) shows the distribution of the samples per class when projected to two dimensions using t-SNE. Each dot represents one data sample from the testing set, and two colours represent two different classes. In this case, feature vectors of length 40 are projected to two feature values and plotted on a 2D graph. Successful feature selection results in well-separated class samples with minimum overlap. Fig. 5.1(b) shows the same t-SNE feature projection after re-labelling the

samples based on the graph learning (graph-based feature weight optimisation and classification) decisions. The used training ratio is at 10%, in Alg. 4.2 graph kernel $\sum_k \sigma_k$ upper bound is $\gamma = 100$, and the initial feature importance $\frac{1}{2{\sigma_k}^2}$ is set to as 1. By comparing the two figures, one can explain the reason for some misclassified samples. Semi-supervised learning via graph spectral signal processing estimates the classification labels via low-pass filtering, i.e., by smoothing the graph signal concerning the underlying graph, ensuring that neighbouring samples (connected by high-weighted edges) have the same labels. Thus, closely located sample points in the figure correspond to the graph neighbors (assuming that the 2-dimensional t-SNE plot captures well the high-dimensional data structure). In both figures, the classifier misclassified some deep sample points (see top left corner) since they were outliers in the given feature space. Indeed, these sample points are connected strongly (with high-weight edges) with surrounding shallow samples. Similarly, clear clusters of shallow samples were formed at the bottom right and bottom left of the figure by smoothing the graph signal causing classification errors for these outliers. By comparing the two figures, it can be concluded that some data samples were misclassified because they are outliers in the feature space. Expert knowledge can now be used to go back to these samples and assess if they were labelled wrong or if the 40-dimensional feature space is not discriminative enough.

To investigate further how different features affect the predictions, Fig. 5.2 shows feature importance obtained by four methods: graph-based feature weight optimisation and classification (Chapter 4); CFS method used in [1], LIME, and Shapley values.

Fig. 5.2 illustrates the normalised feature weights (by dividing each feature weight value by the maximum weight values, ensuring that all values are between 0 and 1 for all 40 constructed features grouped as time, frequency, and time-frequency features. The Shapley values for each feature are obtained by averaging the absolute Shapley values of each sample. One can see that LIME, Shapley

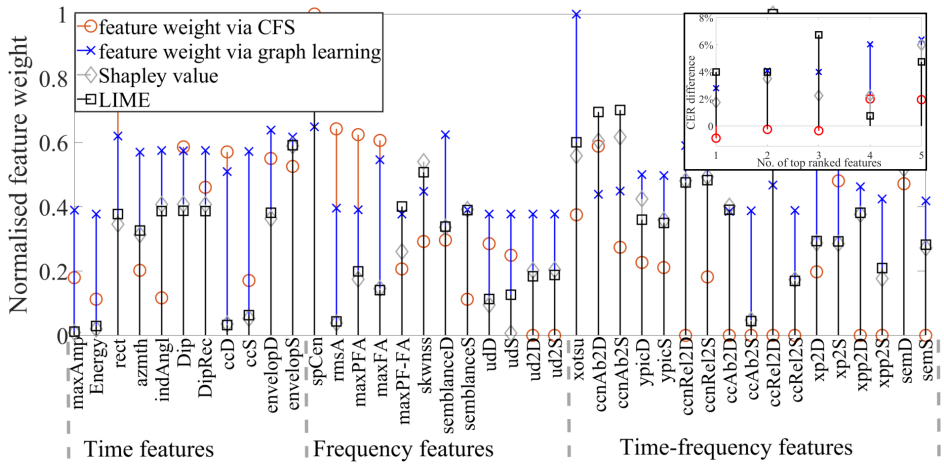


Figure 5.2: Feature Weights. The full name of the features are can be found in Table 5 in [1], with the exception of ccAb2D/S that stands for 2D-CWT cross-correlation for deep/shallow templates; ccRel2D/S represents 2D-CWT cross-correlation between the real part of deep/shallow templates; xpp2D/S stands for 2D-DWT cross-correlation for deep/shallow templates, ud2D/S is the 2D Spectral distance for the shallow template, ccnRel2D stands for Mean coefficient of normalised 2D-CWT cross-correlation between real parts for the deep template, and semS represents Spectral coherency for the shallow template.

values, and CFS methods marked many features as unimportant, discarding some (zero-weighted features). On the other hand, the graph-based method has a smaller range of feature importance values keeping all features with at least 0.4 normalised feature weight. The essential feature weight shave normalised as one, which is different for the four methods.

Explainability of feature importance is application specific. For example, events whose epicenter is farther away from the received sensor are usually observed with higher spectral energy at higher frequency [134]. Thus, from Fig. 5.2, it can be observed that spCen (spectral centroid) is one of the highly ranked features for all four feature ranking approaches; similarly, polarisation features, like Dip (polarisation), DipRec (Dip angle times rectilinearity) *etc.*, are shown to be highly correlated with event depths, and hence are suitable to discriminate shallow/deep event classes.

Regarding seismic physic characteristics, the spectral centroid is a measure of the center of gravity of a signal's frequency spectrum. The spectral centroid can be utilised in the context of seismic signals to reveal details about the dominating frequency content of the signal. The spectral centroid can be used in seismic applications to determine the dominant frequency content of the seismic signal, which can reveal details about the subsurface geology and structure [14, 134]. A shallow subsurface with small-scale structures, for instance, may be indicated by high-frequency content, whereas a deep subsurface with larger features may be indicated by low-frequency content [1].

Polarisation is a property of seismic waves that describes the direction of particle motion. Seismic waves can be decomposed into two orthogonal components, typically referred to as the horizontal and vertical components [4, 14]. The polarization direction is the direction of particle motion in the horizontal plane, perpendicular to the direction of wave propagation [4]. DipRec, or dip angle times rectilinearity, is another measure that can be used to analyse seismic signals. Dip angle is a measure of the steepness of subsurface layers, while rectilinearity is a measure of how straight the seismic wavefronts are [4]. Both polarisation analysis and DipRec can be useful tools for analysing seismic signals and interpreting subsurface structures [4]. Thus because the seismic signal triggered with different depths propagated through different subsurface structures, resulting in these distinct polarisation characteristics.

Whilst explainability is application-specific, interpretability of tabular data can be provided via permutation feature importance, which presents the increase in the classification error rate after all sample values of one feature are randomly permuted. This way, the link between the feature and the class is broken so that the model is correctly learned. If the feature is more important, then the increase in the classification error rate will be higher. The result is shown in the top right corner of Fig. 5.2 for the top 5 features given by the four feature weighting

methods when the same normGLR classifier is used. The result is obtained by averaging over 5 runs and showing a decrease in Classification Error Rate after feature permutation. The x-axis indicates the number of top features that are randomly permuted. Firstly, all sample values of the highest ranked feature for each feature weighting method are randomly permuted, then the process repeats for the top two features, and so on. From the figure, it can be seen that by permuting the top few highly ranked features obtained by the graph-based feature weight optimisation and classification, the Classification Error Rate significantly increased (from 3% when one feature is changed to 6% when top 5 feature values are changed). This shows that the proposed graph-based method identified the most discriminative features for the underlying classification algorithm [80]. Similarly, LIME and Shapley values are good indicators of feature importance, as shown by the decrease in the classification error rate by up to 7% and 4%, respectively.

5.4 Summary

This chapter investigates the interpretability of the proposed graph-based feature weight optimisation and classification model with geophysics research by distinguishing deep and shallow seismic events. The relevance of characteristics can be estimated using explainability approaches, which can also create different heatmaps to highlight the most crucial portions of the data. Although these techniques might be helpful, more is needed to explain the results. In summary, this chapter adopts t-SNE, LIME, and Shapley values techniques to demonstrate the model interpretability via illustration as Fig. 5.1, and the feature weight ranking (see Fig. 5.2).

As aforementioned, Chapter 3 and 4 build the handcrafted features created with expert knowledge and examine the importance of the features for various

seismic signals, which may not be the best for signal representation. Motivated by the success of the recently published papers J3 and C1, in which deep learning models outperformed time-series signal representation learning. Chapter 6 explores the deep learning-based model on (micro)seismic events classification.

Chapter 6

Domain Knowledge Informed Multitask Learning for Landslide Induced (Micro)seismic Classification

Chapter 3 introduced an autonomous end-to-end (micro)seismic analysis platform, including denoising, detection, and classification. Followed by Chapter 4, where a graph-based feature weight optimisation and classification approach is used to classify the actual field recorded (micro)seismic events with the handcrafted features. Subsequently, Chapter 5 explored the interpretability of the proposed graph-based model of Chapter 4, with the weighted features. Motivated by the success of deep learning approaches in geophysics research, this chapter proposes a novel CNN-based domain knowledge-informed multitask learning scheme for improving the accuracy of distinguishing (micro)seismic signals recorded at an ongoing landslide. The proposed multitask learning scheme is posed as a reconstruction problem to fine-tune the (micro)seismic events representation based on the temporal and spatial characteristics of the signal. This chapter is based on the material in journal papers (J1 and J3) and conference paper (C1).

6.1 Introduction

As reviewed in Section 1.2.3, machine learning emerges the superiority in processing extensive volume data with little expert knowledge, especially for (micro)seismic events analysis (e.g., classification). Prior work on (micro)seismic events classification concentrates on conventional approaches, such as SVM and RF suffer from several problems. One of the main issues with these approaches is their reliance on feature selection methods. However, numerous comparison studies reveal that no global feature selection method is effective with all kinds of conventional machine learning models [135]. Even though, as demonstrated in Chapter 4, the proposed graph-based feature weight optimisation outperforms state-of-the-art feature selection for seismic classification, the optimised feature set is obtained with the assistance of a graph-based kernel method classifier, which may be less optimal for other classifiers.

Contrarily, deep learning-based (micro)seismic analysis approaches are considered robust and efficient; mainly, conclude CNN, Recurrent Neural Network (RNN), and Generative Adversarial Networks architectures input of the time series recorded signal or 2D time-frequency spectrum maps [9,59,136–138]. However, they have only shown effectiveness for well-defined signals and require a large amount of training data, especially for highly variable landslide-induced microseismic and tremors.

Deep learning models develop high-level representations of data using stacked layers of neurons and multiple nonlinear transformations to learn complex relationships, find multiscale structures and patterns, and address ill-posed inverse problems [72]. Additionally, deep learning-based approaches can be superior in learning a compact representation of the target signal domain without using expert-designed handcrafted features [53], making the inference process autonomous, practical, and fast. However, the main issue with the current models

is their lack of interpretability [9], which limits their application due to not only trust issues but also difficulties in debugging, improving performance, making fair comparisons across approaches and datasets, and using the results to improve general understanding of underlying seismic processes.

Motivated by two recent studies [139], [140], this chapter proposes a domain-knowledge informed multitask learning scheme comprising signal representation learning and classification, which utilises the seismic wave equation and the P-wave Velocity (Vp) model. The first motivating study is [139], which classifies seismic events into three classes - volcanic tremor, earthquake, and ambient noise by evaluating the spatial coherence of recorded signal, concluding that the spectral differences of the seismic signals with high spatial coherence could enhance the classification performance. Instead, this chapter exploits the high spatial coherency among signal recordings at different sensors to denoise the signal using the second derivative of the seismic wave propagation in the spatial domain, estimated from the recordings of the deployed sensors.

The second motivation stems from the work of [140], where it is shown that with a limited number of sensors, the seismic wave equation can be an alternative pathway to extract the spatial wave representations (second derivative of the seismic wave propagation in the spatial domain) from the acquired signals in the time domain. Rather than numerically solving seismic wave equations, a neural network architecture is employed with a few sensor recordings to conduct the propagation velocity inversion, and displacement prediction [140]. As in [140], the signal representation learning task proposed in this chapter exploits the target signal seismic wave propagation characteristics within a neural network. However, while the seismic wave equation solved with a small amount of time and frequency domain data samples triggered by artificial shots [140], this chapter utilises a CNN architecture to estimate the propagation of temporal characteristics of the real-field seismic recordings. Additionally, this chapter leverage on the CNN

architecture's ability to operate on patch-based input, which allows it to consider local seismic reflection patterns when defining and learning features of target structures [53], to identify and exclude random or coherent seismic noise and processing artifacts of distinct patterns, resulting in reduced misclassification.

The proposed multitask learning scheme includes three key contributions: 1) a novel signal representation learning/reconstruction with two CNN architectures to estimate the signal propagation temporal and spatial characteristics from the real-field seismic recordings; 2) classify the recorded landslide-induced seismic signals with a CNN architecture; 3) experimental results that demonstrate the competitiveness of the proposed system w.r.t state-of-the-art in seismic signal classification.

The remaining part of this chapter is organised as follows. Section 6.2 introduces the state-of-the-art classification methods followed by the related preliminaries (seismic spectral attribute, seismic wave equation, and Vp model) described in Section 6.3. Section 6.4 describes the proposed multitask learning scheme and the implementation details. Section 6.5 presents the experimental results of the proposed method compared with state-of-the-art methods. The last Section 6.7 summarises the work of this chapter.

6.2 Related work

Deep learning is likely to become crucial to the future of seismology due to the development of ever-more-affordable sensors and emerging ground-motion sensing technologies, such as fiber optic cable and accelerometers in smart devices [72]. Thus, the classification of (micro)seismic signals based on deep learning has been extensively developed recently. This section first provides an exhaustive review of recent works on distinguishing (micro)seismic events with deep learning models (see Table 6.1) regards the research objective, architecture, input data format,

and the utilised dataset.

Table 6.1: Recent research on (micro)seismic signal classification

Research objective	Architecture	Input data format	Dataset
Seismic classification [66]	RF, SVM, CNN, Residual Neural Network (ResNN)	32 statistical features (RF, SVM) Bitmaps (CNN, ResNN)	Coal mine
Earthquake magnitude classification [62]	CNN, RNN	Log-Mel STFT (10 s)	STanford EArthquake Dataset (STEAD)
Earthquakes and mining Blasts classification [141]	Amplitude Ratio (AR) + Spectrogram CNN	Pg-to-Sg phase ARs and Rg-to-Sg spectral ARs	Coal mine
Earthquakes, noise and seismo-acoustics classification [63]	Arch-time + CNN Arch-spect + CNN	Continuous waveform, spectrogram	Dutch network
Microseismic, noise, electricity, whistles classification [64]	AlexNet, VGG16, ResNet 18 and ensemble model	Waveform and spectrogram	Microseismic
Earthquake detection and localisation [53]	CNN	10 s 3C waveform	Oklahoma Geological Survey
Earthquakes detection and characteristic estimation [55]	CNN	Single-station, 50 s, 3C waveform	International Federation of Digital Seismograph Networks
Earthquake, tornado, tsunamis and volcanic-seismic classification [142]	CNN, Long Short-Term Memory	Mel-Frequency Cepstral Coefficients	Freesound web
Earthquake classification [67]	Attention-Based CNN	Seismic waveform	South Korea earthquake and environment noise
Microseismic classification [143]	Genetic algorithm-optimised CFS-based CNN	CNN, DT, LDA, NB SVM, KNN, Ensemble	Underground coal mine
Microseismic classification [65]	Capsule Neural Network (CapsNet), CNN, DT, KNN	Handcrafted features	Underground mine
Microseismic classification and detection [138]	CWT+CNN	2D time-frequency map	Synthetic and field recorded data
Seismic phase picking and classification [58]	CapsNet	3C waveform	Global seismics
Earthquake classification [144]	Attention based CNN	Time series waveform	Korean earthquake
Earthquake classification [145]	Transfer CNN	Spectrogram, recurrence plot, and Stockwell transform	Korean earthquake
Seismic classification [146]	Deep canonical correlation analysis	Sonogram feature matrix	Earthquake
Microseismic classification [147]	DCNN , SVM	Multi-channel waveform	Copper Mine
Microseismic classification [148]	CNN and Spatial Pyramid Pooling	Multi-channel microseismic waveform	Recorded at mines
Earthquake classification [149]	CNN, MLP	3C waveform	STEAD

Next, this section highlights the recent research on (micro)seismic events classification, which are also considered benchmarks for the experiment. Deep learning-based classifiers relay little intimate knowledge and have become the popular topic of seismic signal analysis, such as CNN is utilised to distinguish

background noise and earthquake as in [53]. However, the effectiveness of using 3C spectrograms as input for CNN can be improved, particularly with signals of lower magnitude like tectonic earthquakes, mining-induced events, and mining blasts in [141], and quarry blasts or earthquakes in [150]. Furthermore, by comparing the performance of two CNN architectures with time and spectrum as input to discriminate noise, earthquake, and other acoustic signals, it can be concluded that the performances of these two models are roughly the same [63]. The global weighted average pooling structure is utilised as the attention module to distinguish various earthquake events [67]. Specifically, in [67], two attention modules ($GWAP_1$ and $GWAP_2$) are proposed with different weights assignments: 1) weights that represent relative importance as normalised feature values in the channels as Eq. 5 in [67]; 2) weights are calculated based on a contrast stretching method see Eq. 6 in [67]. For multi-station recorded signal, a CNN and GCN-based structure is proposed to classify seismic events with information exchange [68]. In conclusion, deep learning-based approaches can be considered a synthesis of nonlinear local filters, which is superior in learning a compact representation of the target signal without handcrafted features [53]. The main issue with the current deep learning models is the lack of interpretability, which undermines its applicability [9].

6.3 Preliminaries

This section presents the related preliminaries of seismic spectral attribute and wave propagation characteristic (seismic wave equation and Vp model).

6.3.1 Seismic Spectral Attribute

Landslide-induced (micro)seismic events with distinctive spectral properties are a crucial criterion for slope deformation monitoring. In [14], Vouillamoz pro-

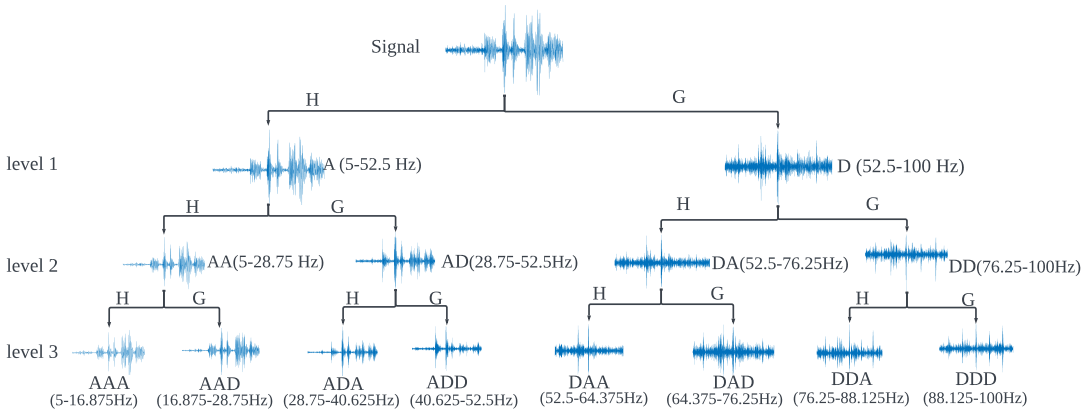


Figure 6.1: Procedure for signal decomposition using WPD

vided an exhaustive analysis of spectrum distribution for microseismic signals at slow-moving clay-rich debris slides and concluded that the dominant frequencies of earthquake and quake respectively concentrate on $1-20$ Hz, and $5-50$ Hz, while rockfall and natural/anthropogenic noise signals have broader spectrum $5-100$ Hz. Feature selection and classification techniques are utilised to highlight the most discriminative features for the signals recorded at the ongoing landslide, which mostly are the spectrum attributes such as the energies on the frequency bands $0-31.25$ Hz, $31.25-62.5$ Hz [4]. Besides, the central frequency of the first quartile is highlighted as the essential parameter for distinguishing the slope failures-induced seismic signals in [28]. Most recently, the work in Chapter 4 indicates that the kurtosis of the signal filtered with $5-9$ Hz, $13-17$ Hz could be used to identify earthquake, slide quake, and rockfall [15, 125].

While discrete wavelet transform provides flexible time–frequency resolution, it suffers from a relatively low resolution in the high-frequency region. This deficiency leads to difficulty in differentiating transient high-frequency components. In comparison, the Wavelet Packet Decomposition(WPD) further decomposes the detailed information of the signal in the high-frequency region, thereby overcoming this limitation. Fig. 6.1 schematically illustrates a WPD-based signal decom-

position process, where a three-level WPD produces a total of 8 subbands, with each subband covering one-sixteenth of the signal frequency spectrum. Due to the spectral characteristics of sensors, the raw recorded signal is bandpassed with $5-100\,\text{Hz}$. The enhanced signal decomposition capability makes WPD an attractive tool for detecting and differentiating transient elements with high-frequency characteristics.

6.3.2 Seismic Wave Equation

In general, seismic signals involve the superimposition of several arbitrarily shaped pulses traveling at speeds dictated by the elastic characteristics and density of the medium [151]. The propagation of these signals through an isotropic and homogeneous medium is usually modeled by a 1-dimensional time-dependent seismic wave equation given by [152]:

$$\frac{1}{V^2}\frac{\partial^2 u(x,t)}{\partial t^2} = \frac{\partial^2 u(x,t)}{\partial x^2}, \tag{6.1}$$

where $u(x,t)$ is the displacement in point x at time t. The solution to this differential equation can be expressed as $u(x,t) = \overline{A}\cos[2\pi(t - x/V) - \varphi]$, where $\overline{A}$ and φ are the amplitude and phase of the wave, respectively, V is the P/S-wave velocity; for typical sensor arrays with vertical channels (as the datasets utilised in this thesis) where P-waves are dominant, V is the P-wave velocity, Vp, determined by the elastic properties and density of the medium along the wave propagation pathway.

6.3.3 *Vp* Model

The location of the signal source is considered to be one of the sufficient features to distinguish seismic events occurring in the different critical zone of the monitored slope deformation [153]. For instance, rockfalls are usually triggered by falling

blocks, whose source locations concentrate on the perimeter of the slopes. Slide quakes are identified as shallow-sourced local seismic events occurring near the deployed sensors [153]. Generally, the precise signal localisation requires a three-dimensional (3D) *Vp* model of the crust, determined by the density and elastic properties of the material along the seismic wave pathway [154]. Specifically, each cube of this 3D *Vp* model represents the seismic *Vp* when the wave propagates through the material inside the cube area. The *Vp* of the crust varies from region to region and is altered by the geological composition (e.g., rock, or clay *etc.*); thus, it can be utilised to distinguish the seismic signals triggered at different locations [153].

6.4 Methodology

The main idea of the proposed method is to use the recorded signals and *Vp* models to estimate the propagation characteristic of the seismic events via domain-knowledge informed multitask learning and then used this information to classify the signal based on the fact that different classes of seismic events propagate differently through the medium. In the following, an overview of the proposed method is provided, and then the implementation details are provided.

6.4.1 Overview

The flowchart of the proposed domain-knowledge informed multitask learning scheme is illustrated as Fig. 6.2. First, the recorded seismic waveform is decomposed with a level three WPD (Block 1) to identify its frequency distribution over time with variable-width windows using a bank of low-pass and high-pass filters, trading off time and frequency resolution [155]. This results in eight signal representations, each corresponding to a frequency band.

Block 2 estimates the second derivative w.r.t time, of the displacement vec-

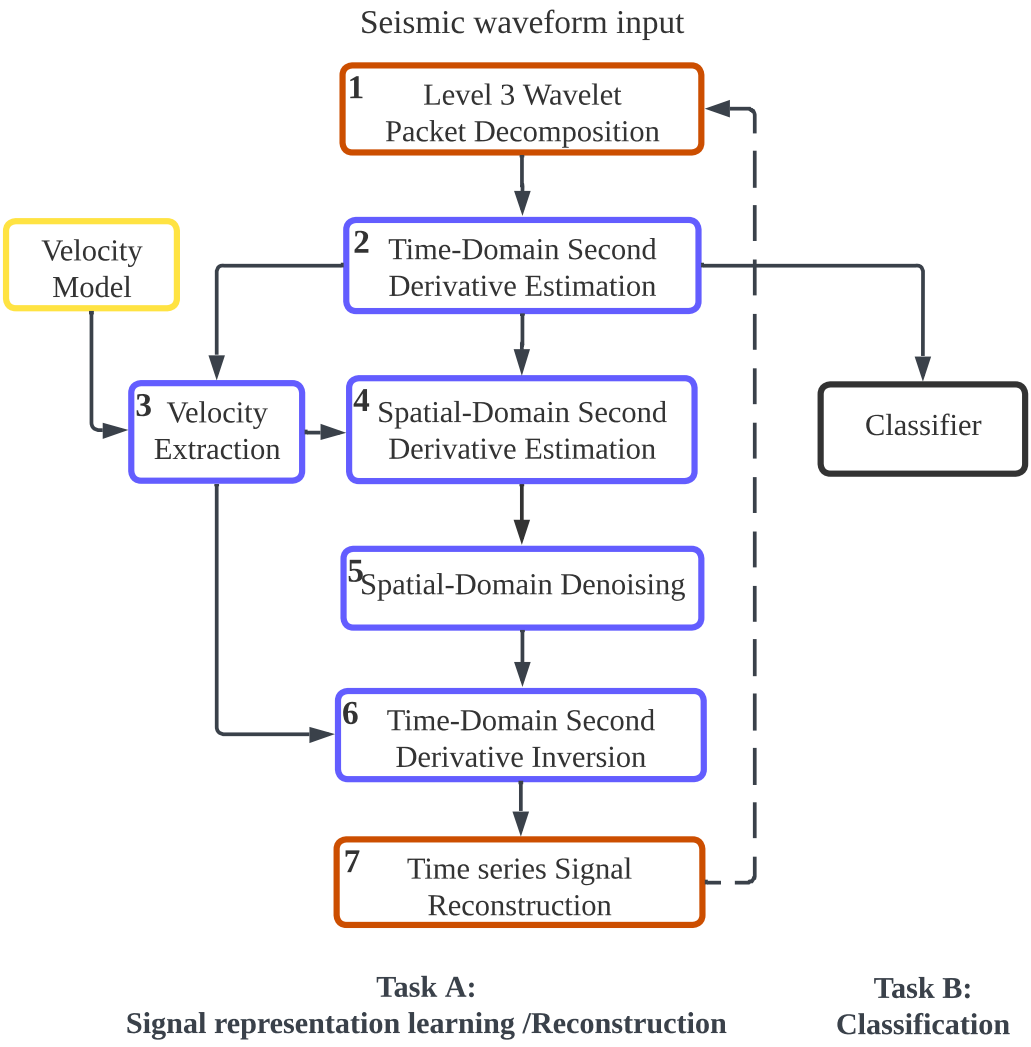
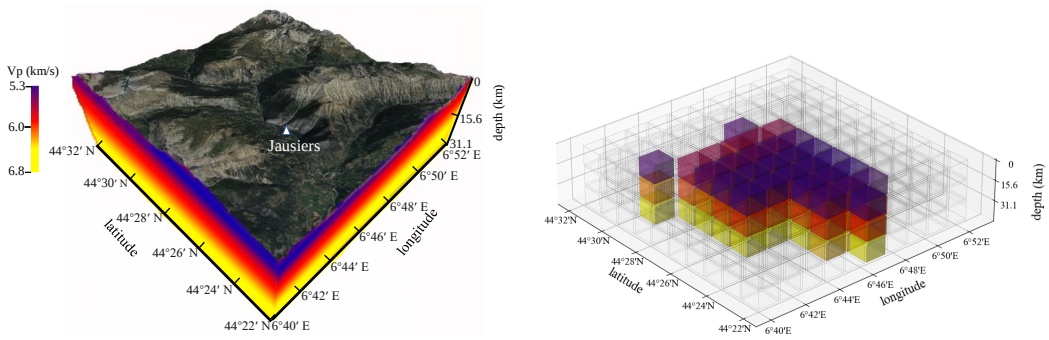


Figure 6.2: Proposed domain-knowledge informed multitask learning scheme

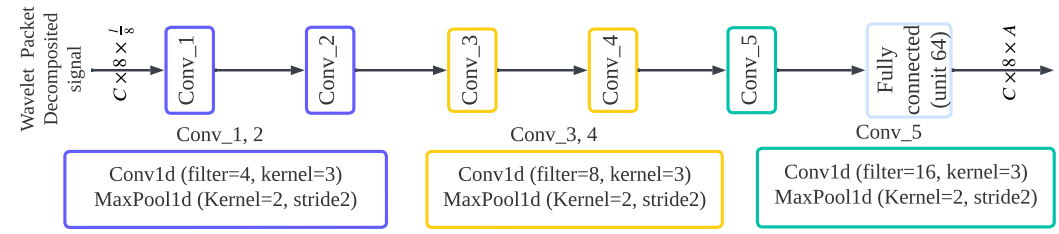
tor, $\frac{\partial^2 u(x,t)}{\partial t^2}$, from the collected wavelet packet-decomposed signals, which can be regarded as the latent feature for classification, since, the second derivative w.r.t. time carries information on wave curvature and depth, discriminating seismic events, particularly those originating from distinct locations.



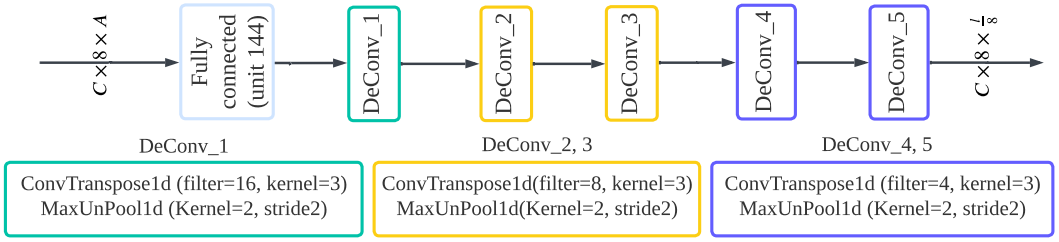
(a) Earthquake topographic Velocity model (b) Model mapped to grid

Figure 6.3: The modified velocity model for earthquake signal [2, 3]

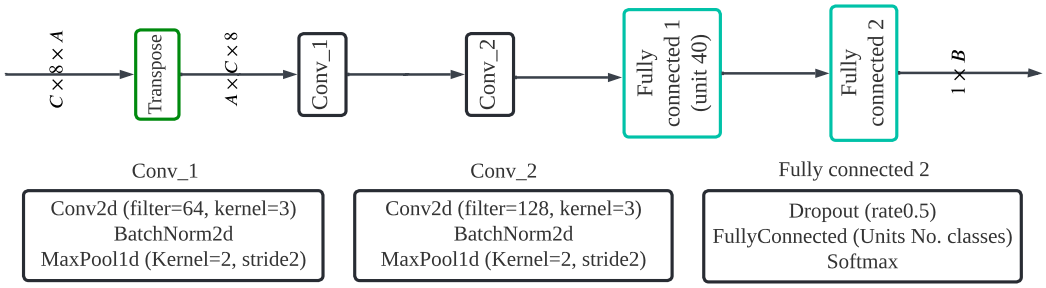
Block 3 is introduced to extract one Vp value, for each event, from each of the 3D Vp models, where the size of the grid cells can vary depending on the resolution required, trading off the resolution of the seismic velocity model and the complexity of the multi-tasking learning model. Here, the grid size of $10 \times 10 \times 3$ velocity values was chosen to balance resolution and complexity. Block 3 comprises two fully connected layers (FC-2) to estimate the sampling probabilities in the Vp model, which represent the likelihood of a velocity value being selected as the most likely propagation velocity of corresponding events, that in turn weighs the contribution of each velocity value to the overall proposed multi-task learning model fitting process. The Vp model is usually computed over a particular area with inversion techniques and interpolating the seismic signals recorded by a dense network. Fig. 6.3 (a) shows a seismic topographical area of interest using, as an example, recorded earthquakes that originated in South Alps near Jausiers. Using the Vp model in [156], as shown in Fig. 6.3 (b), the Vp value is extracted. To simplify the model, this chapter only considers the



(a) EstT_CNN (CNN for Block (2) in Fig. 6.2)



(b) Inv_EstT_CNN (CNN for Block (7) in Fig. 6.2)



(c) Class_CNN (CNN for classification)

Figure 6.4: CNN architecture (C: No. of channels, B: No. of classes, l: window length, A: Dimension of Second derivatives component)

areas (cubes) most likely to contain seismic sources (limited to 10 cubes in each direction), where the blue, red, and yellow cubes, respectively, represent the Vp around 5.3 ,6.0 ,6.7 km/s, in Fig. 6.3 (b).

Seismic waves in earth materials are subject to attenuation and dispersion in a broad range of frequencies; in addition, the field-observed seismic signals are usually obscured by a significant amount of noise, which affects the precision of estimating the second derivative w.r.t time using the seismic wave equation Eq. 6.1 [157]. Hence, the proposed workflow performs signal enhancement on derivatives w.r.t the spatial coordinate. To do that, Block 4 uses the seismic wave

equation Eq. 6.1 to convert the second derivative w.r.t time $\frac{\partial^2 u(x,t)}{\partial t^2}$ (Block 2) to the second derivative w.r.t the spatial coordinate $\frac{\partial^2 u(x,t)}{\partial x^2}$, by performing element-wise division by the squared estimated Vp value (Block 3). $\frac{\partial^2 u(x,t)}{\partial x^2} = \frac{\partial^2 u(x,t)}{\partial t^2} \oslash V^2$, where $\oslash$ stands for element-wise division

Motivated by the high spatial coherency among the seismic recordings as demonstrated in [139], Block 5 denoises the spatial domain second derivative by replacing the three highest-energy spectral components with their mean value. Block 6 performs the inverse operation to Block 4 on the denoised signal (Block 5), recovering the second derivative w.r.t time with Eq. 6.1. Finally, Block 7 reconstructs the decomposed time series signal using the recovered second derivative w.r.t time $\frac{\partial^2 u(x,t)}{\partial t^2}$ obtained by Block 6 after spatial-domain denoising in Block 5. The output of Block 7 updates the loss function of the reconstruction task by measuring the difference between the output of Block 1 (dash line in Fig. 6.2), as explained in the following section. The CNN classifier takes the labelled second derivative w.r.t time (output of Block 2) as input to predict the event label.

6.4.2 Implementation Details

This section provides the implementation details for individual blocks in Fig. 6.2, and the learning forward of the proposed multitask learning scheme is presented as Alg. 6.3. Fig. 6.4 shows CNN architectures used where B is the No. of classes, C is the No. of data channels, l is the window length, and A represents the length of the second derivative of the displacement vector $\mathbf{u}$ w.r.t time. Specifically, the purpose of the convolutional layers in EstT_CNN is to extract the time domain second derivative of the wave equation from the decomposed signal via WPD; while the ConvTranspose1d layers in Inv_EstT_CNN are utilised to reconstruct the time domain signal for computing the loss of the reconstruction task. Regarding the classification CNN architecture Class_CNN, the Conv2d layers are used for computing the category vector of the input signal with two FullyConnected

layers. Since Blocks 4 and 6 provide a simple application of the seismic wave equation Eq. 6.1 as steps 8 and 10 in Alg. 6.3, the focus is on Block 2, 3, 5, 7, and the classification block, which is described next one by one. The proposed multi-task (signal representation learning and classification) scheme is an end-to-end learning process. CNN models, Figs. 6.4 (a) and (b), used for feature learning, perform signal reconstruction, taking as input the wavelet decomposed signal and the estimated time domain acceleration, respectively. The output of Fig. 6.4 (a) is the learnt latent feature set, fed to Fig. 6.4 (c) for classification. Additionally, the loss function of the proposed scheme contains two terms, reconstruction and classification loss, which are utilised to train the whole network. Thus, CNN models, Figs. 6.4 (a) and (b), are embedded into the signal representation learning task, whose weights are updated synchronously minimising Eq. 6.3.

Time-Domain Second Derivative Estimation (Block 2, in Fig. 6.2)

The input signal is first decomposed with a level three WPD using db4 wavelet, resulting in 8 frequency bands (step 1 in Alg. 6.3). Let $\mathbf{T} \in \mathbb{R}^{C \times 8 \times \frac{l}{8}}$ be the wavelet-decomposed signal at the output of Block 1, where 8 comes from the number of signal representations, each corresponding to one frequency band (see Fig. 6.1). Taking signal $\mathbf{T}$ as the input, the proposed CNN architecture EstT_CNN, illustrated in Fig. 6.4 (a), which contains five convolution layers, ReLU activation function, and maxpooling operations, is used to estimate the second derivative w.r.t time, $\mathbf{U} = \frac{\partial^2 u(x,t)}{\partial t^2} \in \mathbb{R}^{C \times 8 \times A}$ (step 2 in Alg. 6.3) for each input window. Heuristically, A as the length of the second derivative signal is set to 64.

Velocity Extraction (Block 3, in Fig. 6.2)

Given initial Vp models (denoted as $\mathcal{M} = \{\mathbf{M}_1, ..., \mathbf{M}_O\}$, where O is the No. of provided velocity models) from [2, 3, 153], and the estimated time-domain second derivative $\mathbf{U}$, the workflow poses the RANdom SAmple Consensus (RANSAC) as

a voting strategy to extract velocities (Vp values, denotes as $\mathbf{v} = [v_1, ..., v_O]$) that optimally describe the underlying seismic wave attributes. In general, Block 3 identifies one Vp value v_o for each given Vp model $\mathbf{M}_o \in \mathbb{R}^{10\times10\times3}$ (steps 6 and 7 in Alg. 6.3). In particular, similarly to [158], the sampling indices are generated by a set of *trainable* probability tensors $\mathcal{P} = \{\mathbf{P}_1, ..., \mathbf{P}_O\}$, and, $\mathbf{P}_o \in \mathbb{R}^{10\times10\times3}$ to select the optimal hypothesis of candidate values that optimises the objective (step 3 in Alg. 6.3). This work employs the FC-2 network with the input and output unit defined respectively as A, 128 and 128, 300 to estimate the probability tensor $\mathbf{P}_o$ for generating random h hypotheses (specifically, in the implementation set as $h = 12$) based on the multinomial distribution so that the dimension of $\mathbf{P}_o$ matches the dimensions of the initial Vp model $\mathbf{M}_o$ (see Subsection 6.4.1). The network takes as an input the output of Block 2 ($\mathbf{U}$) and predicts the probability of signal originating in each of $10 \times 10 \times 3$ cubes. At the output of Block 3, the velocity v_o that corresponds to the second derivative w.r.t time estimated in Block 2 is extracted.

Spatial-Domain Denoising (Block 5, in Fig. 6.2)

After obtaining spatial-domain second derivative $\mathbf{B} \in \mathbb{R}^{C\times8\times A}$ via step 8 in Alg. 6.3 with seismic wave equation Eq. 6.1, Block 5 performs denoising by first extract the top 3 values along the second derivative axis (last dimension) of $\mathbf{B}$ and form a tensor as $\overrightarrow{\mathbf{B}} \in \mathbb{R}^{C\times8\times3}$. Then the pairwise distance $\overrightarrow{\mathbf{B}}$ are measured and the data channel index $\hat{I}_c$ that minimise the distance for c-th data channel defined as:

$$\hat{I}_c = \arg \min_{j\in\{1,...,C\},j\neq c} \left\| \overrightarrow{\mathbf{B}_c} - \overrightarrow{\mathbf{B}_j} \right\|_2^2, \tag{6.2}$$

thus, with $\forall c \in \{1, ..., C\}, \hat{\mathbf{I}} = \{\hat{I}_1, ..., \hat{I_C}\}$ the denoised spatial-domain second derivative $\hat{\mathbf{B}} \leftarrow 0.5 \times (\mathbf{B} + \mathbf{B}(\hat{\mathbf{I}}))$.

Time-Domain Signal Reconstruction (Block 7, in Fig. 6.2)

Block 7 performs time series signal reconstruction by setting a network objective to be a reconstruction task (Task A in Fig. 6.2) that optimises the mean square error between the recovered time domain signal $\hat{\mathbf{T}} \in \mathbb{R}^{C \times 8 \times \frac{l}{8}}$ after spatial domain denoising (output of Block 6) and the input wavelet-decomposed signal $\mathbf{T}$, defined as:

$$L\_r = \frac{1}{C} \sum_{c=1}^{C} \frac{1}{8} \sum_{i=1}^{8} (\frac{1}{\frac{l}{8}} \sum_{j=1}^{\frac{l}{8}} \left(T_{c,i,j} - \hat{T}_{c,i,j} \right)^2), \tag{6.3}$$

where i and j represent the index of wavelet-decomposed frequency bands and the data samples of the event window. Specifically, a CNN architecture is proposed that contains five deconvolution layers, ReLU activation function, and maxpooling operations, as illustrated in Fig. 6.4 (b) (Inv_EstT_CNN) and performs the reverse operation to Fig. 6.4 (a) (EstT_CNN).

Classification

With the learned optimal latent feature (output of Block 2, $\mathbf{U}$), a CNN architecture is employed for classification (Task B in Fig. 6.2), containing two 2-dimension convolution layers and two fully connected layers, with the Cross-Entropy loss function defined as $L\_c = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{B} \hat{y}_{i,j} \log (q_{i,j})$, where N and B representing, respectively, the No. of events and classes, $\hat{y}_{i,j}$ is a binary indicator (0 or 1) $\hat{y}_{i,j} = 1$ if class label j is the correct classification for the event i, otherwise 0, $q_{i,j}$ is the predicted probability that event i is of class j.

6.5 Experimental Setup

This section uses, as benchmarks, state-of-the-art methods to evaluate the accuracy and precision of the proposed CNN-based multitask learning scheme to classify the seismic signals recorded at an ongoing landslide. All experiments are

Algorithm 6.3 Multi-task learning forward

Require: Input data $\mathbf{X}$; No. of hypothesis h for RANSAC; Set of velocity model $\mathcal{M} = \{\mathbf{M}_1, ..., \mathbf{M}_O\}$;

Ensure: The predicted label y;

1: Decompose $\mathbf{X}$ into $\mathbf{T}$ with level 3 WPD;
2: Estimate $\mathbf{U}$ with EstT_CNN by feeding $\mathbf{T}$;
3: Estimate $\mathbf{P}_o, \forall o \in \{1, ..., O\}$ via FC-2 by feeding $\mathbf{U}$;
4: **for** i in the range of h **do**
5: **for** o in the range of O **do**
6: Sample a cell index j using multinomial distribution based random sampling from $\mathbf{P}_o$ [159];
7: Extract v_o from $\mathbf{M}_o$ using index j;
8: Calculate the spatial-domain second derivative $\mathbf{B} \leftarrow \frac{1}{v_o^2} \times \mathbf{U}$;
9: Obtain denoised spatial-domain second derivative $\hat{\mathbf{B}} \leftarrow$ Eq. 6.2;
10: Calculate the time-domain second derivative $\hat{\mathbf{U}}_o \leftarrow \hat{\mathbf{B}} \times {v_o}^2$;
11: Smooth across all velocity models get $\hat{\mathbf{U}} \leftarrow \frac{1}{O} \sum_{o=1}^{O} \hat{\mathbf{U}}_o$;
12: Reconstructed signal $\hat{\mathbf{T}}$ using Inv_EstT_CNN by feeding $\hat{\mathbf{U}}$;
13: Classify the events input $\hat{\mathbf{U}}$ as latent feature map to Class_CNN get $\mathbf{y}_i$;
14: Determine the prediction $y = \arg\max_b \frac{1}{h} \sum_{i=1}^{h} \mathbf{Y}_{i,b}$, where $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_h \end{bmatrix}$;
15: **return** Reconstructed signal $\hat{\mathbf{T}}$; The predicted label y;

Table 6.2: Experimental Setup

No. of data	**Rockfall**	**Slide Quake**	**Earthquake**	**Noise**
Train (60%)	240	140	233	211
Validation (10%)	40	70	116	105
Test (30%)	121	24	39	35
Network Parameter	Total: 183136 (183120 (Trainable), 16 (Non-trainable))			

conducted on NVIDIA RTX 3090 with PyTorch v1.12. All CNN-based models are trained over 450 epochs using Adaptive Moment Estimation (ADAM) optimiser with $\beta_1 = 0.9$, $\beta_2 = 0.999$. Note that the overall loss function of the proposed domain knowledge-informed multitask learning scheme is defined as $Loss = L\_r + L\_c$.

This chapter utilises datasets SZ13, SZ14, and SZ15 to evaluate the proposed scheme. As preprocessing, the experiment first filter the six-channel raw signals with $5 - 100$ Hz, and then segments a 10 s window for each event from 2 s before the event arrival, to form the input as $\mathbf{X} \in \mathbb{R}^{C \times l}$, where $C = 6$, and $l = 2500$. For benchmarking, this chapter employed stratified sampling with 60% (training), 10% (validation), and 30% (testing). At the testing stage (see Table 6.2), the best model corresponds to the highest validation accuracy. To eliminate the impact of random data splitting, the experiment is repeated 50 times with a randomly split training, validation, and testing set. To prevent overfitting, the proposed model adapts early stopping and drop-out layers.

6.5.1 *Vp* Models

Table 6.3: Seismic topographic areas

Event	Latitude (N)	Longitude (E)
Rockfall	44°20′41″-44°21′8″	6°40′24″-6°40′47″
Slide Quake	44°20′51″-44°21′58″	6°40′30″-6°40′41″
Earthquake	44°24′10″-44°31′27″	6°40′44″-6°52′16″
Noise	44°20′41″-44°21′8″	6°40′24″-6°40′47″

The *Vp* models are obtained from the seismic topographic areas (Table 6.3) and the depth provided in [2, 3, 153]. Since the recorded data analysed in this chapter contains four different classes, four 3D *Vp* models, one for each class, are introduced. Specifically, the 3D *Vp* models of rockfall, slide quake, and noise events are obtained by cubic interpolation of the 2D profiles, and these 2D profiles

are computed by inverting the anthropogenic shots recorded by the sensors deployed in the preset geophones using the Quasi-Newton inversion approach [153]. The 3D *Vp* model is available under (`https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=b259c149-19dd-11ec-9603-497c92695674`) accessed on 20 March 2022, for the earthquake is obtained by inverting the dense network recorded teleseismic data in [156].

6.6 Results and Discussion

Table 6.4: Experimental results Averaged with 50 runs (mean (std))

Classifier	**Acc**	**Recall**			
		Rockfall	**Slide Quake**	**Earthquake**	**Noise**
CNN + GCN [68]	0.8576 (0.03)	0.8715 (0.04)	0.7662 (0.07)	0.8442 (0.06)	0.9176 (0.04)
DeepQuake [63]	0.8650 (0.03)	0.9041 (0.04)	0.7560 (0.06)	0.8769 (0.05)	0.8796 (0.05)
ConNetQuake [53]	0.8470 (0.03)	0.8901 (0.04)	0.7254 (0.08)	0.8494 (0.05)	0.8756 (0.05)
$GWAP_1$ [67]	0.9140 (0.01)	0.9210 (0.03)	0.8249 (0.07)	0.9379 (0.03)	0.9389 (0.02)
$GWAP_2$ [67]	0.9345 (0.02)	0.9405 (0.03)	0.8520 (0.05)	**0.9554 (0.02)**	0.9592 (0.02)
Proposed	**0.9422 (0.02)**	**0.9413 (0.03)**	**0.9083 (0.04)**	0.9474 (0.02)	**0.9602 (0.02)**
Replace Task A (Fig. 6.2)	0.9312 (0.01)	0.9393 (0.02)	0.8571 (0.05)	0.9403 (0.03)	0.9491 (0.02)

The bold content represents the highest Acc and Recall for each class.

To evaluate the performance of the proposed multitask learning scheme, this section compares the classification performance to the state-of-the-art in terms of Accuracy (*Acc*) and Recall (as in Section 3.4.1), as shown in Table 6.4. The utilised benchmarks are based on recent CNN architectures primarily proposed to distinguish seismic signals, such as earthquakes, from ambient noise, [63], and [53] as well as designs of [68], and [67] that are based on GCN and attention modules to enhance the classification performance. It can be seen from the table that the proposed method outperforms the benchmarks in terms of *Acc* averaged over all four classes. Furthermore, it provides the highest recall rate for all classes except earthquakes, where the schemes optimised for detecting and classifying earthquakes, such as [67], perform slightly better. However, the limitation of [67] is that the size of the training set required for good performances of seismic

classification is large.

Comparing the results across the classes, the recall of slide quakes is lower since the number of slide quakes in the training data is significantly smaller than the number of the others.

To evaluate the performance of the intermediate CNN models (Fig. 6.4 (a) and (b)), the state-of-the-art feature extraction architecture of [144] is implemented to replace task A in Fig. 6.2. The optimised latent feature maps are fed into the proposed classifier (Fig. 6.2), and the resulting classification performance is shown in the last row of Table 6.4. Comparing the last two rows which use the same proposed classifier, the proposed intermediate feature learning architectures (Fig. 6.4 (a) and (b)) outperform that of [144] for all classes. The largest improvement is observed for slide quake events.

6.7 Summary

This chapter proposes a novel domain knowledge-informed multitask learning scheme with the seismic wave equation and Vp models for ongoing landslide-induced (micro)seismic event classification. The proposed approach involves an alternative pathway for seismic signal propagation characteristic estimation. Specifically, rather than numerically solving the seismic wave equation, this chapter novelly adapts a CNN architecture to estimate the signal propagation time domain second derivative. Additionally, to migrate the impact of noise, the proposed approach employs smoothness in the spatial domain as concluded in [139] that the signal recorded channel has high spatial coherence for (micro)seismic events. The experimental result demonstrates that the proposed multitask learning scheme is effective and outperforms state-of-the-art methods. The proposed model can also be applied to other applications, such as volcano active monitoring and earthquake early warning. In future work, it is worth investigating the

effect of the 3D velocity model accuracy comparing homogeneous, 2-layer, and gradient velocity models.

Chapter 7

Conclusion

This chapter gives a summary and conclusion of this work. It also provides several insights into future research.

7.1 Summary

Recent advancements in in-situ microseismic monitoring systems have led to a rise in the use of (micro)seismic events analysis based on signal processing and machine learning. Using transforms (e.g., STFT, Wavelet, and WPD *etc.*), statistics analysis (e.g., extract the high-order statistical features: skewness, kurtosis, and *etc.*), and machine learning techniques (e.g., regression, classification, and clustering *etc.*), (micro)seismic event denoising, detection, feature engineering, classification, and localisation have all been investigated in recent years. However, recent researches on landslide-induced (micro)seismic events are generally less accurate and effective, mainly when processing continuous real-time recorded signals. This research proposed an exact and economic system for (micro)seismic event analysis based on advanced signal processing and machine learning. This thesis achieves this goal via (1) analysing the probability distribution and the coherence of the single or multi-channel recorded signal to facilitate the identi-

fication of the potential (micro)seismic events; (2) signal representation learning by proposed graph-based feature weight optimisation with handcraft features and the deep learning-based approach with little expert knowledge.

Specifically, the following methodologies are proposed to address the related challenges. Firstly, Chapter 2 introduces the basic information about GSP, containing an illustration of graph construction (see Fig. 2.1). Following this, Chapter 3 proposes an end-to-end automated (micro)seismic analysis platform, containing (1) signal denoising via ambient noise probability distribution analysis and GraphBF; (2) (micro)seismic events detection with Neyman-Pearson lemma; (3) the events are classified with GRL. Each block of the proposed system outperforms the state-of-the-art approaches (e.g., wavelet transform for denoising, STA/LTA for detection, and RF for classification); additionally, the proposed system is highly effective.

Then in Chapter 4, the (micro)seismic events detection with multi-channel recordings is expanded and established on the integration of MCM and Neyman-Pearson lemma (see Chapter 3). Additionally, a comprehensive feature engineering, which covers a comparison of state-of-the-art feature selection methods (filter, wrapper, and embedded approaches) and feature extraction via PCA, is explored to benchmark the proposed graph-based feature optimisation and classification for landslide-induced (micro)seismic events. The proposed approach is evaluated with continuously recorded signals, which is more in line with practical applications. Besides, Chapter 4 contributes a feature recommendation section that concludes the most distinctive characteristics of rockfall, slide quake, and earthquake. Chapter 5 demonstrates the interpretability of the proposed graph model in Chapter 4 with t-SNE, LIME, and Shapley values, concluding that the proposed model is fully explainable and reliable.

The chapters investigate the (micro)seismic events classification established on the handcraft features. A domain-knowledge-informed multitask learning scheme

is proposed to improve the accuracy via deep learning approaches in Chapter 6. The proposed scheme involves *Vp* model and seismic wave equation to assist the signal representation learning and classify the signals with distinct latent feature maps. The experimental result demonstrates that the proposed multitask learning scheme is effective and outperforms state-of-the-art methods.

7.2 Future Work

Emerging signal processing and machine learning techniques facilitate accurate and effective improvement for slope stability monitoring. For (micro)seismic events analysis in practice, although many challenges are addressed in this research, there remain open challenges where this section provides several suggestions for future research:

1. **(Micro)seismic events analysis enhancement with hardware instruments**

   Future work includes identifying the proposed methods' weaknesses and providing an efficient real-time implementation of those algorithms, integrating them into slope stability monitoring systems. Combining the proposed methods with hardware instruments could be cost-reduced, regards data storage and power supply, *etc.* For example, embedding denoising and detection approaches to seismometers so that they could operate only at the moments in which predefined (micro)seismic occurs.

2. **Events identification and classification with insufficient annotated samples**

   Another interesting area of research is to detect and classify the (micro)seismic events with limited annotated samples. As attempted, graph-based classifiers ((norm)GLR, and GTV) consider semi-supervised learning approaches

requiring fewer labelled events, while the demonstrated results are competitive and even superior to the supervised machine learning approaches. Regarding these graph-based approaches, future work can explore novel distance measures or postprocessing. The Euclidean distance is usually utilised for GSP-based classifiers, which might suffer from the curse of dimensionality resulting in poor performance with the high-dimension feature set.

Motivated by the success of few/one-shot learning on computer vision tasks, another research direction could be applying and modifying the state-of-the-art few/one-shot learning approaches on (micro)seismic events identification and classification, with very few annotated data samples to mitigate the impact of uncertainty from manual annotation. Transfer learning methods are also deserved to explore, which could employ the knowledge obtained from the well-studied dataset to investigate the new observations collected at different geological conditions.

3. **Performance enhancement with interpretability methods**

   As demonstrated in this thesis, machine learning, particularly deep learning algorithms, exhibit significant promise for learning signal representations while having poor interpretability. With the emergence of explainable Artificial Intelligence, this drawback has been mitigated. Thus, the future work of this research can focus on applying the interpretability methods to enhance the (micro)seismic events detection and classification performances.

Bibliography

[1] S. M. Mousavi, S. P. Horton, C. A. Langston, and B. Samei, "Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression," *Geophysical Journal International*, vol. 207, no. 1, pp. 29–46, 2016.

[2] L. Jenatton, R. Guiguet, F. Thouvenot, and N. Daix, "The 16,000-event 2003–2004 earthquake swarm in ubaye (french alps)," *Journal of Geophysical Research*, vol. 112, no. B11, 2007.

[3] A. Paul, M. Cattaneo, F. Thouvenot, D. Spallarossa, N. Béthoux, and J. Fréchet, "A three-dimensional crustal velocity model of the southwestern alps from local earthquake tomography," *Journal of Geophysical Research*, vol. 106, no. B9, pp. 19 367–19 389, 2001.

[4] F. Provost, C. Hibert, and J.-P. Malet, "Automatic classification of endogenous landslide seismicity using the random forest supervised classifier," *Geophysical Research Letters*, vol. 44, no. 1, pp. 113–120, 2017.

[5] D. J. Varnes, "Landslide types and processes," *Landslides and engineering practice*, vol. 24, pp. 20–47, 1958.

[6] ——, "Slope movement types and processes," *Special report*, vol. 176, pp. 11–33, 1978.

[7] L. Highland, P. T. Bobrowsky *et al.*, *The landslide handbook: a guide to understanding landslides.* US Geological Survey Reston, 2008.

[8] F. Guzzetti, "Forecasting natural hazards, performance of scientists, ethics, and the need for transparency," *Toxicological & Environmental Chemistry*, vol. 98, no. 9, pp. 1043–1059, 2016.

[9] Z. Ma and G. Mei, "Deep learning for geological hazards analysis: Data, models, applications, and opportunities," *Earth-Science Reviews*, vol. 223, p. 103858, 2021.

[10] M. Wenner, C. Hibert, L. Meier, and F. Walter, "Near real-time automated classification of seismic signals of slope failures with continuous random forests," *Natural Hazards and Earth System Sciences*, no. Jul 8, pp. 1–23, 2020.

[11] V. I. Grechka and W. M. Heigl, *Microseismic monitoring.* Society of Exploration Geophysicists Tulsa, Oklahoma, USA, 2017.

[12] M. Malfante, M. Dalla Mura, J.-P. Métaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Processing Magazine*, vol. 35, no. 2, pp. 20–30, 2018.

[13] F. Staněk and L. Eisner, "Seismicity induced by hydraulic fracturing in shales: A bedding plane slip model," *Journal of Geophysical Research: Solid Earth*, vol. 122, no. 10, pp. 7912–7926, 2017.

[14] N. Vouillamoz, S. Rothmund, and M. Joswig, "Characterizing the complexity of microseismic signals at slow-moving clay-rich debris slides: the super-sauze (southeastern france) and pechgrabe (upper austria) case studies." *Earth Surface Dynamics*, vol. 6, no. 2, 2018.

[15] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, "Automated platform for microseismic signal analysis: Denoising, detection, and classification in slope stability studies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7996–8006, 2020.

[16] N. Rosser, M. Lim, D. Petley, S. Dunning, and R. Allison, "Patterns of precursory rockfall prior to slope failure," *Journal of geophysical research: earth surface*, vol. 112, no. F4, 2007.

[17] R. Allen, "Automatic phase pickers: Their present use and future prospects," *Bulletin of the Seismological Society of America*, vol. 72, no. 6B, pp. S225–S242, 1982.

[18] F. Li, J. Rich, K. J. Marfurt, and H. Zhou, "Automatic event detection on noisy microseismograms," in *2014 SEG Annual Meeting*. OnePetro, 2014.

[19] J. Akram, "Automatic p-wave arrival time picking method for seismic and micro-seismic data," in *CSPG CSEG CWLS Convention*. CSEG Calgary, AB, Canada, 2011.

[20] 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," *Journal of Volcanology and Geothermal Research*, vol. 340, pp. 130–142, 2017.

[21] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, "High-accuracy real-time microseismic analysis platform: case study based on the super-sauze mud-based landslide," in *Geoconvention*, 2020.

[22] S. Kumar, R. Vig, and P. Kapur, "Development of earthquake event detection technique based on sta/lta algorithm for seismic alert system," *Journal of the Geological Society of India*, vol. 92, no. 6, pp. 679–686, 2018.

[23] F. Lara, R. León, R. Lara-Cueva, A. F. Tinoco-S, and M. Ruiz, "Detection of volcanic microearthquakes based on homomorphic deconvolution and sta/lta," *Journal of Volcanology and Geothermal Research*, vol. 421, p. 107439, 2022.

[24] O. M. Saad and Y. Chen, "Earthquake detection and p-wave arrival time picking using capsule neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6234–6243, 2020.

[25] S. M. Mousavi, C. A. Langston, and S. P. Horton, "Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform," *Geophysics*, vol. 81, no. 4, pp. V341–V355, 2016.

[26] S. M. Mousavi and C. A. Langston, "Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding," *Bulletin of the Seismological Society of America*, vol. 106, no. 4, pp. 1380–1393, 2016.

[27] J. Akram and D. Eaton, "Adaptive microseismic event detection and automatic time picking," in *2012 CSEG Annual Convention*, vol. 15, 2012.

[28] M. Wenner, C. Hibert, A. van Herwijnen, L. Meier, and F. Walter, "Near-real-time automated classification of seismic signals of slope failures with continuous random forests," *Natural Hazards and Earth System Sciences*, vol. 21, no. 1, pp. 339–361, 2021.

[29] L. Han, J. Wong, J. Bancroft *et al.*, "Time picking and random noise reduction on microseismic data," *CREWES Research Report*, vol. 21, pp. 1–13, 2009.

[30] J. I. Sabbione and D. Velis, "Automatic first-breaks picking: New strategies and algorithms," *Geophysics*, vol. 75, no. 4, pp. V67–V76, 2010.

[31] R. Sleeman and T. Van Eck, "Robust automatic p-phase picking: an on-line implementation in the analysis of broadband seismogram recordings," *Physics of the earth and planetary interiors*, vol. 113, no. 1-4, pp. 265–275, 1999.

[32] E. Caffagni, D. W. Eaton, J. P. Jones, and M. van der Baan, "Detection and analysis of microseismic events using a matched filtering algorithm (mfa)," *Geophysical Journal International*, vol. 206, no. 1, pp. 644–658, 2016.

[33] L. Feng, V. Pazzi, E. Intrieri, T. Gracchi, and G. Gigli, "Joint detection and classification of rockfalls in a microseismic monitoring network," *Geophysical Journal International*, 2020.

[34] J. Akram and D. W. Eaton, "A review and appraisal of arrival-time picking methods for downhole microseismic dataarrival-time picking methods," *Geophysics*, vol. 81, no. 2, pp. KS71–KS91, 2016.

[35] J. Saari, "Automated phase picker and source location algorithm for local distances using a single three-component seismic station," *Tectonophysics*, vol. 189, no. 1-4, pp. 307–315, 1991.

[36] V. Oye and M. Roth, "Automated seismic event location for hydrocarbon reservoirs," *Computers & Geosciences*, vol. 29, no. 7, pp. 851–863, 2003.

[37] P. Venegas, N. Pérez, D. Benítez, R. Lara-Cueva, and M. Ruiz, "Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from cotopaxi volcano," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1991–2003, 2019.

[38] X. Shang, X. Li, A. Morales-Esteban, and G. Chen, "Improving microseismic event and quarry blast classification using artificial neural networks

based on principal component analysis," *Soil Dynamics and Earthquake Engineering*, vol. 99, pp. 142–149, 2017.

[39] R. Ghosh and H. Sardana, "Multi-feature optimization strategies for target classification using seismic and acoustic signatures (conference presentation)," in *Automatic Target Recognition XXX*, vol. 11394. Proc SPIE Int Soc Opt Eng, 2020, p. 113940I.

[40] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli, "Infinite latent feature selection: A probabilistic latent graph-based ranking approach," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1398–1406.

[41] W. D. Fisher, T. K. Camp, and V. V. Krzhizhanovskaya, "Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection," *Journal of Information and Computing Science*, vol. 20, pp. 143–153, 2017.

[42] R. Lara-Cueva, D. Benítez, E. Carrera, M. Ruiz, and J. Rojo-Álvarez, "Feature selection of seismic waveforms for long period event detection at cotopaxi volcano," *Journal of Volcanology and Geothermal Research*, vol. 316, pp. 34–49, 2016.

[43] F. Martínez-Álvarez, J. Reyes, A. Morales-Esteban, and C. Rubio-Escudero, "Determining the best set of seismicity indicators to predict earthquakes. two case studies: Chile and the iberian peninsula," *Knowledge-Based Systems*, vol. 50, pp. 198–210, 2013.

[44] G. Cortés, M. C. Benitez, L. García, I. Álvarez, and J. M. Ibanez, "A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 253–263, 2015.

[45] R. Soto, F. Huenupan, P. Meza, M. Curilem, and L. Franco, "Spectro-temporal features applied to the automatic classification of volcanic seismic events," *Journal of Volcanology and Geothermal Research*, vol. 358, pp. 194–206, 2018.

[46] C. Levy, D. Jongmans, and L. Baillet, "Analysis of seismic signals recorded on a prone-to-fall rock column (vercors massif, french alps)," *Geophysical Journal International*, vol. 186, no. 1, pp. 296–310, 2011.

[47] T. Joachims, "Making large-scale svm learning practical," Technical Report, Tech. Rep., 1998.

[48] M. Curilem, A. Cuevas, R. Soto, F. Huenupan, C. Martin, M. S. Khan, F. Gil, C. Cardona, and L. Franco, "Classification of volcanic seismic events: An expert knowledge analysis for feature selection," in *International Conference of Pattern Recognition Systems (ICPRS)*. IET, 2017, pp. 2–6.

[49] J. Kortström, M. Uski, and T. Tiira, "Automatic classification of seismic events within a regional seismograph network," *Computers & Geosciences*, vol. 87, pp. 22–30, 2016.

[50] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[51] P. Venegas, N. Pèrez, D. S. Benítez, R. Lara-Cueva, and M. Ruiz, "Building machine learning models for long-period and volcano-tectonic event classification," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. IEEE, 2019, pp. 1–6.

[52] P. E. E. Lara, C. A. R. Fernandes, A. Inza, J. I. Mars, J.-P. Métaxian, M. Dalla Mura, and M. Malfante, "Automatic multichannel volcano-seismic classification using machine learning and emd," *IEEE Journal of Selected*

Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1322–1331, 2020.

[53] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances*, vol. 4, no. 2, p. e1700578, 2018.

[54] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9476–9488, 2019.

[55] A. Lomax, A. Michelini, and D. Jozinović, "An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network," *Seismological Research Letters*, vol. 90, no. 2A, pp. 517–529, 2019.

[56] S. Lapins, B. Goitom, J.-M. Kendall, M. J. Werner, K. V. Cashman, and J. O. Hammond, "A little data goes a long way: automating seismic phase arrival picking at nabro volcano with transfer learning," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 7, p. e2021JB021910, 2021.

[57] A. Shaheen, U. b. Waheed, M. Fehler, L. Sokol, and S. Hanafy, "Groningennet: Deep learning for low-magnitude earthquake detection on a multi-level sensor network," *Sensors*, vol. 21, no. 23, p. 8080, 2021.

[58] O. M. Saad and Y. Chen, "Capsphase: Capsule neural network for seismic phase classification and picking," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[59] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection," *Scientific Reports*, vol. 9, no. 1, pp. 1–14, 2019.

[60] Z. He, P. Peng, L. Wang, and Y. Jiang, "Enhancing seismic p-wave arrival picking by target-oriented detection of the local windows using faster-rcnn," *IEEE Access*, vol. 8, pp. 141 733–141 747, 2020.

[61] T. Wang, D. Trugman, and Y. Lin, "Seismogen: Seismic waveform synthesis using gan with application to seismic data augmentation," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 4, p. e2020JB020077, 2021.

[62] M. Shakeel, K. Itoyama, K. Nishida, and K. Nakadai, "Emc: Earthquake magnitudes classification on seismic signals via convolutional recurrent networks," in *2021 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2021, pp. 388–393.

[63] L. Trani, G. A. Pagani, J. P. P. Zanetti, C. Chapeland, and L. Evers, "Deepquake—an application of cnn for seismo-acoustic event classification in the netherlands," *Computers & Geosciences*, vol. 159, p. 104980, 2022.

[64] J. Li, S. Tang, K. Li, S. Zhang, L. Tang, L. Cao, and F. Ji, "Automatic recognition and classification of microseismic waveforms based on computer vision," *Tunnelling and Underground Space Technology*, vol. 121, p. 104327, 2022.

[65] P. Peng, Z. He, L. Wang, and Y. Jiang, "Microseismic records classification using capsule network with limited training samples in underground mining," *Scientific Reports*, vol. 10, no. 1, pp. 1–16, 2020.

[66] Y. Duan, Y. Shen, I. Canbulat, X. Luo, and G. Si, "Classification of clustered microseismic events in a coal mine using machine learning," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 13, no. 6, pp. 1256–1273, 2021.

[67] B. Ku, G. Kim, J.-K. Ahn, J. Lee, and H. Ko, "Attention-based convolutional neural network for earthquake event classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 12, pp. 2057–2061, 2020.

[68] G. Kim, B. Ku, J.-K. Ahn, and H. Ko, "Graph convolution networks for seismic events classification using raw waveform data from multiple stations," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[69] N. Ackerley, M. Beer, I. Kougioumtzoglou, E. Patelli, and S. Au, "Principles of broadband seismometry," *Encyclopedia of Earthquake Engineering,(Springer, Berlin)*, 2014.

[70] "Sensors." [Online]. Available: https://www.passcal.nmt.edu/content/instrumentation/sensors

[71] B. Biescas, F. Dufour, G. Furdada, G. Khazaradze, and E. Suriñach, "Frequency content evolution of snow avalanche seismic signals," *Surveys in Geophysics*, vol. 24, no. 5-6, pp. 447–464, 2003.

[72] S. M. Mousavi and G. C. Beroza, "Deep-learning seismology," *Science*, vol. 377, no. 6607, p. eabm4470, 2022.

[73] H. R. Pourghasemi, B. Pradhan, C. Gokceoglu, M. Mohammadi, and H. R. Moradi, "Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at haraz watershed, iran," *Arabian Journal of Geosciences*, vol. 6, no. 7, pp. 2351–2365, 2013.

[74] X. Dong, D. Thanou, L. Toni, M. Bronstein, and P. Frossard, "Graph signal processing for machine learning: A review and new perspectives," *IEEE Signal processing magazine*, vol. 37, no. 6, pp. 117–127, 2020.

[75] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-

dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 50, no. 3, pp. 83–98, Apr. 2013.

[76] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[77] W. Hu, J. Pang, X. Liu, D. Tian, C.-W. Lin, and A. Vetro, "Graph signal processing for geometric data and beyond: Theory and applications," *IEEE Transactions on Multimedia*, 2021.

[78] A. Ortega, *Introduction to graph signal processing*. Cambridge University Press, 2022.

[79] B. Zhao, K. He, L. Stankovic, and V. Stankovic, "Improving event-based non-intrusive load monitoring using graph signal processing," *IEEE Access*, vol. 6, pp. 53 944–53 959, 2018.

[80] C. Yang, G. Cheung, and V. Stankovic, "Alternating binary classifier and graph learning from partial labels," in *Asia-Pacific Signal and Information Processing Association (APSIPA)*. IEEE, 2018, pp. 1137–1140.

[81] M. Ye, V. Stankovic, L. Stankovic, and G. Cheung, "Robust deep graph based learning for binary classification," *IEEE Transactions on Signal and Information Processing over Networks*, 2020.

[82] K. He, C. Yang, V. Stankovic, and L. Stankovic, "Graph-based clustering for identifying region of interest in eye tracker data analysis," in *International Workshop on Multimedia Signal Processing*. IEEE, 2017, pp. 1–6.

[83] M. Ye, C. Yang, V. Stankovic, L. Stankovic, and S. Cheng, "Distinct feature extraction for video-based gait phase classification," *IEEE Transactions on Multimedia*, vol. 22, no. 5, pp. 1113–1125, 2019.

[84] K. He, D. Jakovetic, B. Zhao, V. Stankovic, L. Stankovic, and S. Cheng, "A generic optimisation-based approach for improving non-intrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6472–6480, 2019.

[85] C. Yang, G. Cheung, and W. Hu, "Graph metric learning via gershgorin disc alignment," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5530–5534.

[86] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[87] J. Pang and G. Cheung, "Graph laplacian regularization for image denoising: Analysis in the continuous domain," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1770–1785, 2017.

[88] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.

[89] B. Motz, G. Cheung, A. Ortega, and P. Frossard, "Re-sampling and interpolation of dibr-synthesized images using graph-signal smoothness prior," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2015, pp. 1077–1084.

[90] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[91] H. Attia, S. Gaya, A. Alamoudi, F. M. Alshehri, A. Al-Suhaimi, N. Alsulaim, A. M. Al Naser, M. A. J. Eddin, A. M. Alqahtani, J. P. Rojas *et al.*,

"Wireless geophone sensing system for real-time seismic data acquisition," *IEEE Access*, vol. 8, pp. 81 116–81 128, 2020.

[92] C. Huynh, C. Hibert, C. Jestin, J.-P. Malet, P. Clément, and V. Lanticq, "Real-time classification of anthropogenic seismic sources from distributed acoustic sensing data: Application for pipeline monitoring," *Seismological Research Letters*, 2022.

[93] D. Kumar and I. Ahmed, "Seismic noise," *Encyclopedia of Solid Earth Geophysics*, pp. 1–6, 2020.

[94] G. Liu, S. Fomel, L. Jin, and X. Chen, "Stacking seismic data using local correlation," *Geophysics*, vol. 74, no. 3, pp. V43–V48, 2009.

[95] A. Ansari, A. Noorzad, H. Zafarani, and H. Vahidifard, "Correction of highly noisy strong motion records using a modified wavelet de-noising method," *Soil Dynamics and Earthquake Engineering*, vol. 30, no. 11, pp. 1168–1181, 2010.

[96] M. Beenamol, S. Prabavathy, and J. Mohanalin, "Wavelet based seismic signal de-noising using shannon and tsallis entropy," *Computers & Mathematics with Applications*, vol. 64, no. 11, pp. 3580–3593, 2012.

[97] S. M. Mousavi and C. A. Langston, "Adaptive noise estimation and suppression for improving microseismic event detection," *Journal of Applied Geophysics*, vol. 132, pp. 116–124, 2016.

[98] Y. Chen, "Non-stationary least-squares complex decomposition for microseismic noise attenuation," *Geophysical Journal International*, vol. 213, no. 3, pp. 1572–1585, 2018.

[99] J. Han and M. van der Baan, "Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding," *Geophysics*, vol. 80, no. 6, pp. KS69–KS80, 2015.

[100] Y. Chen, "Automatic microseismic event picking via unsupervised machine learning," *Geophysical Journal International*, vol. 212, no. 1, pp. 88–102, 2018.

[101] R. Kimiaefar, H. Siahkoohi, A. Hajian, and A. Kalhor, "Seismic random noise attenuation using artificial neural network and wavelet packet analysis," *Arabian Journal of Geosciences*, vol. 9, no. 3, p. 234, 2016.

[102] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2012.

[103] S. Taheri and J. Behboodian, "Neyman-pearson lemma for fuzzy hypotheses testing," *Metrika*, vol. 49, no. 1, pp. 3–17, 1999.

[104] A. Trnkoczy, "Understanding and parameter setting of sta/lta trigger algorithm," in *New Manual of Seismological Observatory Practice (NMSOP)*. Deutsches GeoForschungsZentrum GFZ, 2009, pp. 1–20.

[105] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, "Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms," *Journal of Volcanology and Geothermal Research*, vol. 180, no. 1, pp. 1–8, 2009.

[106] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 20, no. 2, pp. 189–201, 2009.

[107] G. Roffo, "Feature selection library (matlab toolbox)," *arXiv preprint arXiv:1607.01327*, 2016.

[108] A. Tonnellier, A. Helmstetter, J.-P. Malet, J. Schmittbuhl, A. Corsini, and M. Joswig, "Seismic monitoring of soft-rock landslides: the super-sauze and valoria case studies," *Geophysical Journal International*, vol. 193, no. 3, pp. 1515–1536, 2013.

[109] 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 Research Letters*, vol. 88, no. 3, pp. 878–891, 2017.

[110] S. M. Mousavi, S. P. Horton, C. A. Langston, and B. Samei, "Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression," *Geophysical Journal International*, vol. 207, no. 1, pp. 29–46, 2016.

[111] S. Qu, Z. Guan, E. Verschuur, and Y. Chen, "Automatic high-resolution microseismic event detection via supervised machine learning," *Geophysical Journal International*, vol. 218, no. 3, pp. 2106–2121, 2019.

[112] S. Esmaili, S. Krishnan, and K. Raahemifar, "Content based audio classification and retrieval using joint time-frequency analysis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5. IEEE, 2004, pp. V–665.

[113] S. Zaugg, M. Van Der Schaar, L. Houégnigan, C. Gervaise, and M. André, "Real-time acoustic classification of sperm whale clicks and shipping impulses from deep-sea observatories," *Applied Acoustics*, vol. 71, no. 11, pp. 1011–1019, 2010.

[114] F. Provost, J.-P. Malet, C. Hibert, C. Abanco Martínez de Arenzana, and M. Hurlimann Ziegler, "Towards a standard typology of endogenous landslide seismic sources," *Earth Surface Dynamics*, vol. 6, no. 4, pp. 1059–1088, 2018.

[115] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1739–1747, 2018.

[116] A. Gadde, S. K. Narang, and A. Ortega, "Bilateral filter: Graph spectral interpretation and extensions," in *IEEE International Conference on Image Information Processing*. IEEE, 2013, pp. 1222–1226.

[117] S. M. Mousavi and C. A. Langston, "Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data," *Geophysics*, vol. 82, no. 4, pp. V211–V227, 2017.

[118] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[119] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, 2018.

[120] P. Shi, D. Angus, S. Rost, A. Nowacki, and S. Yuan, "Automated seismic waveform location using multichannel coherency migration (mcm)–i: theory," *Geophysical Journal International*, vol. 216, no. 3, pp. 1842–1866, 2019.

[121] M. F. A. Hady and F. Schwenker, "Semi-supervised learning," *Handbook on Neural Information Processing*, pp. 215–239, 2013.

[122] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu, "Graph learning: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 109–127, 2021.

[123] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *International conference on Machine learning (ICML-03)*, 2003, pp. 912–919.

[124] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, "Graph laplacian tomography from unknown random projections," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1891–1899, 2008.

[125] J. Li, C. Yang, V. Stankovic, L. Stankovic, and S. Pytharouli, "Graph-based micro-seismic signal classification with an optimised feature space," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020.

[126] S. Letzgus, P. Wagner, J. Lederer, W. Samek, K.-R. Müller, and G. Montavon, "Toward explainable artificial intelligence for regression models: A methodological perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 4, pp. 40–58, 2022.

[127] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," *arXiv*, 2 2017. [Online]. Available: http://arxiv.org/abs/1702.08608

[128] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible

AI," *Information Fusion*, vol. 58, no. October 2019, pp. 82–115, 6 2020. [Online]. Available: https://doi.org/10.1016/j.inffus.2019.12.012https://linkinghub.elsevier.com/retrieve/pii/S1566253519308103

[129] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, vol. 8, no. 8, p. 832, 7 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/8/832

[130] Z. C. Lipton, "The Mythos of Model Interpretability," *Queue*, vol. 16, no. 3, pp. 31–57, 6 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3236386.3241340

[131] L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, 2008.

[132] E. Štrumbelj and I. Kononenko, "An efficient explanation of individual classifications using game theory," *Journal of Machine Learning Research*, vol. 11, 2010.

[133] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?"," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Augu. New York, NY, USA: ACM, 8 2016, pp. 1135–1144.

[134] D. Fagan, K. Van wijk, and J. Rutledge, "Clustering revisited: A spectral analysis of microseismic events," *Geophysics*, vol. 78, no. 2, 2013.

[135] P. Zhou, X. Hu, P. Li, and X. Wu, "Online feature selection for high-dimensional class-imbalanced data," *Knowledge-Based Systems*, vol. 136, pp. 187–199, 2017.

[136] Y. Wu, Y. Lin, Z. Zhou, D. C. Bolton, J. Liu, and P. Johnson, "Deepdetect: A cascaded region-based densely connected network for seismic event detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 62–75, 2018.

[137] W. Zhu and G. C. Beroza, "Phasenet: a deep-neural-network-based seismic arrival-time picking method," *Geophysical Journal International*, vol. 216, no. 1, pp. 261–273, 2019.

[138] G. Zhang, C. Lin, and Y. Chen, "Convolutional neural networks for microseismic waveform classification and arrival picking," *Geophysics*, vol. 85, no. 4, pp. WA227–WA240, 2020.

[139] T. Permana, T. Nishimura, H. Nakahara, and N. Shapiro, "Classification of volcanic tremors and earthquakes based on seismic correlation: application at sakurajima volcano, japan," *Geophysical Journal International*, vol. 229, no. 2, pp. 1077–1097, 2022.

[140] S. Karimpouli and P. Tahmasebi, "Physics informed machine learning: Seismic wave equation," *Geoscience Frontiers*, vol. 11, no. 6, pp. 1993–2001, 2020.

[141] R. Tibi, L. Linville, C. Young, and R. Brogan, "Classification of local seismic events in the utah region: A comparison of amplitude ratio methods with a spectrogram-based machine learning approachclassification of local seismic events in the utah region," *Bulletin of the Seismological Society of America*, vol. 109, no. 6, pp. 2532–2544, 2019.

[142] A. O. Ekpezu, I. Wiafe, F. Katsriku, and W. Yaokumah, "Using deep learning for acoustic event classification: The case of natural disasters," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2926–2935, 2021.

[143] P. Peng, Z. He, L. Wang, and Y. Jiang, "Automatic classification of microseismic records in underground mining: a deep learning approach," *IEEE Access*, vol. 8, pp. 17 863–17 876, 2020.

[144] B. Ku, J. Min, J.-K. Ahn, J. Lee, and H. Ko, "Earthquake event classification using multitasking deep learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 7, pp. 1149–1153, 2020.

[145] G. Kim, B. Ku, and H. Ko, "Multifeature fusion-based earthquake event classification using transfer learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 6, pp. 974–978, 2020.

[146] O. Lindenbaum, N. Rabin, Y. Bregman, and A. Averbuch, "Seismic event discrimination using deep cca," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, pp. 1856–1860, 2019.

[147] B. Lin, X. Wei, and Z. Junjie, "Automatic recognition and classification of multi-channel microseismic waveform based on dcnn and svm," *Computers & Geosciences*, vol. 123, pp. 111–120, 2019.

[148] B. Lin, X. Wei, Z. Junjie, and Z. Hui, "Automatic classification of multi-channel microseismic waveform based on dcnn-spp," *Journal of Applied Geophysics*, vol. 159, pp. 446–452, 2018.

[149] S. Zhang, B. Ku, and H. Ko, "Learnable maximum amplitude structure for earthquake event classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[150] L. Linville, R. C. Brogan, C. Young, and K. A. Aur, "Global-and local-scale high-resolution event catalogs for algorithm testing," *Seismological Research Letters*, vol. 90, no. 5, pp. 1987–1993, 2019.

[151] S. Stein and M. Wysession, *An introduction to seismology, earthquakes, and earth structure*. John Wiley & Sons, 2009.

[152] R. E. Sheriff and L. P. Geldart, *Exploration seismology*. Cambridge university press, 1995.

[153] F. Provost, J. Malet, J. Gance, A. Helmstetter, and C. Doubre, "Automatic approach for increasing the location accuracy of slow-moving landslide endogenous seismicity: the apoloc method," *Geophysical Journal International*, vol. 215, no. 2, pp. 1455–1473, 2018.

[154] A. AlAli and F. Anifowose, "Seismic velocity modeling in the digital transformation era: a review of the role of machine learning," *Journal of Petroleum Exploration and Production Technology*, pp. 1–14, 2021.

[155] A. Chakraborty and D. Okaya, "Frequency-time decomposition of seismic data using wavelet-based methods," *Geophysics*, vol. 60, no. 6, pp. 1906–1916, 1995.

[156] M. Paffrath, W. Friederich, S. M. Schmid, M. R. Handy, AlpArray, A.-S. D. W. Group *et al.*, "Imaging structure and geometry of slabs in the greater alpine area–a p-wave travel-time tomography using alparray seismic network data," *Solid Earth*, vol. 12, no. 11, pp. 2671–2702, 2021.

[157] T. M. Müller, B. Gurevich, and M. Lebedev, "Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks—a review," *Geophysics*, vol. 75, no. 5, pp. 75A147–75A164, 2010.

[158] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother, "Dsac-differentiable ransac for camera localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6684–6692.

[159] E. L. Lehmann and E. Lehmann, *Testing statistical hypotheses*. Springer, 1986, vol. 2.