

Seismic signal classification and detection based on deep  
learning

PhD Thesis

Jiaxin Jiang

Research Group

Electronic and Electrical Engineering

University of Strathclyde, Glasgow

May 30, 2025

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.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

# Abstract

Passive seismic monitoring is important for understanding subsurface processes such as landslides, mining activities, and geothermal systems, enabling the prediction and mitigation of their effects. However, continuous seismic monitoring produces vast datasets containing various sources of seismicity that require accurate classification. Manual detection and labeling of these events is both time-consuming and prone to inconsistency, even when performed by the same expert. To address these challenges, this thesis first proposes an automated joint detection and classification method for characterizing seismic events using Convolutional Neural Networks (CNNs).

Despite their effectiveness and high accuracy, deep learning models, such as CNNs, face two significant limitations: the lack of interpretability due to their “black-box” nature and the large amount of manually labeled data required for training. Interpretability is particularly important in seismic applications where reliable detection and classification of earthquakes are essential for infrastructure safety. To ensure both accuracy and explainability, the second contribution of this thesis is a novel methodology for data labeling, verification, and re-labeling through CNNs enhanced by Layer-wise Relevance Propagation (LRP), a popular explainable AI tool. This approach aims to provide transparency in seismic event detection, improving trustworthiness in AI-driven decisions.

Manual labeling of seismic events is often inefficient and contradictory to the goal of automated detection. To overcome the time and resource inefficiency of manual labeling, the third contribution of the thesis is a self-supervised learning (SSL) model that reduces the dependency on large amounts of annotated data while maintaining high detection accuracy. This model significantly reduces the manual effort involved

in labeling seismic events, thereby improving the efficiency and scalability of seismic monitoring systems.

The proposed CNN-based models achieve approximately 90% accuracy, effectively distinguishing between seismic sources such as rockfalls and earthquakes. Furthermore, the LRP-based method enhances interpretability of seismic classification by generating explainable relevance maps, which visually highlight the most influential parts of the seismic signal that contributed to the model’s decision. These maps help experts understand model reasoning, validate whether the model is focusing on geophysically meaningful features, and identify potential mislabeling or overlooked patterns in the data. The SSL model achieves accuracy comparable to state-of-the-art fully supervised methods while requiring only 5% to 30% of the labeled data. Specifically, the lower end (around 5%) is sufficient for distinguishing well-separated classes like rockfalls, while more complex scenarios with overlapping signal characteristics, such as those between quakes and a subset of earthquakes, may require up to 20% labeled data. This flexibility significantly reduces the manual labeling burden without sacrificing detection precision. Together, these contributions offer an accurate, reliable, efficient, and explainable deep learning-based framework for seismic event detection and classification, advancing the state of seismic signal monitoring and analysing.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xi</b>
<b>Acronyms</b>	<b>xvi</b>
<b>Acknowledgements</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Research Aim and Objectives . . . . .	4
1.2 Contributions of the thesis . . . . .	5
1.3 Organisation of the thesis . . . . .	7
1.4 Publications . . . . .	7
<b>2 Background</b>	<b>9</b>
2.1 (Micro)seismic Event Analysis . . . . .	9
2.1.1 System overview . . . . .	10
2.1.2 Data Acquisition . . . . .	11
2.1.3 Detection . . . . .	11
2.1.4 Feature Engineering . . . . .	13
2.1.5 Classification . . . . .	14
2.2 Deep Learning Architectures . . . . .	15
2.2.1 Performance measures . . . . .	17

2.3	Dataset . . . . .	18
<b>3</b>	<b>Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks</b>	<b>22</b>
3.1	Introduction . . . . .	22
3.2	Related work . . . . .	24
3.2.1	Binary classification . . . . .	24
3.2.2	Classification of more than one event type . . . . .	26
3.2.3	Overview of the literature of DL-based seismic signal classification	28
3.3	Methodology . . . . .	29
3.3.1	Data processing . . . . .	29
3.3.2	Neural Network Architecture . . . . .	30
3.3.3	Data and Training Strategy . . . . .	32
3.3.4	Sliding window-based detection . . . . .	33
3.4	Results and discussion . . . . .	34
3.4.1	Classification Results . . . . .	34
3.4.2	Interpretation of Misclassified Events . . . . .	36
3.4.3	Comparison with state of the art DL-based methods . . . . .	38
3.4.4	Continuous detection results . . . . .	39
3.4.5	Transferability analysis . . . . .	40
3.4.6	Internal workings of proposed architecture via feature maps visualization . . . . .	44
3.4.7	Complexity Analysis . . . . .	46
3.5	Summary . . . . .	47
<b>4</b>	<b>A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model</b>	<b>50</b>
4.1	Introduction . . . . .	50
4.1.1	Pertinence of Trustworthy AI to seismic analysis . . . . .	51
4.2	Related work . . . . .	53
4.3	Data gathering and context . . . . .	54

Contents

4.3.1	Labelling . . . . .	55
4.4	Methodology . . . . .	56
4.4.1	Proposed CNN-based architecture . . . . .	56
4.4.2	Training and testing strategy . . . . .	57
4.4.3	Continuous detection and post-processing . . . . .	58
4.4.4	Explainability-informed re-labelling . . . . .	60
4.5	Results and discussion . . . . .	61
4.5.1	Analysis of classifier output . . . . .	61
4.5.2	Explainability . . . . .	63
4.5.3	Explaining origin of misclassification . . . . .	64
4.5.4	Re-labelling results . . . . .	67
4.6	Summary . . . . .	72
<b>5</b>	<b>Generative Self-Supervised Learning for Seismic Event Classification</b>	<b>74</b>
5.1	Introduction . . . . .	74
5.2	Related work . . . . .	75
5.3	Methodology . . . . .	78
5.3.1	Data Pre-processing . . . . .	78
5.3.2	Generative-based SSL using auto-encoder . . . . .	79
5.3.3	Downstream task . . . . .	79
5.4	Results and discussion . . . . .	81
5.4.1	Simulation setup . . . . .	81
5.4.2	Unsupervised learning results and feature visualization . . . . .	82
5.4.3	Semi-supervised ANN results . . . . .	84
5.4.4	SOM-based method results . . . . .	86
5.4.5	Labelling Effort vs. Accuracy Trade-off . . . . .	87
5.5	Summary . . . . .	89
<b>6</b>	<b>Conclusion and Future Work</b>	<b>91</b>
6.1	Summary . . . . .	91
6.2	Future Work . . . . .	93

Contents

<b>Bibliography</b>	<b>95</b>
<b>A Catalogue of Larisa data</b>	<b>112</b>
<b>B Detailed Results and Confusion Matrices for Self-Supervised Learning</b>	
<b>Downstream Tasks</b>	<b>116</b>
B.1 K-means clustering . . . . .	116
B.2 ANN-based semi-supervised method . . . . .	119
B.3 SOM models . . . . .	119

# List of Figures

2.1	Block diagram of the traditional system overview for (micro)seismic event analysis. . . . .	11
2.2	Geographic location of the SZC sensor station (a) global distribution and (b) detailed topographic positioning. . . . .	19
2.3	Examples of seismic events: (a) Earthquake (b) Quake (c) Rockfall. Three different events are shown for each class, including high, medium and low SNR cases. In all cases, we show three waveforms, representing, respectively, the East, North and vertical direction of the three-component sensors from top to bottom. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023. . . . .	21
3.1	Examples of seismic signals for 3 classes (earthquake, quake and rockfall) with 3 different input formats (temporal waveform, STFT and CWT maps) in Résif dataset. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023. . . . .	30

List of Figures

3.2 The network structure for three different inputs.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 32

3.3 Waveforms of two quake events in all six channels. The quake event which was correctly detected and classified (first column) and the quake event which was misclassified as noise (second column) with the same ordinate axis ranges from  $-1.5 \times 10^{-6}$ m/s to  $1.5 \times 10^{-6}$ m/s. The misclassified quake event with the scaled zoomed (third column), its ordinate axis ranges from  $-1.5 \times 10^{-7}$ m/s to  $1.5 \times 10^{-7}$ m/s.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 37

3.4 Continuous detection results for an earthquake event.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 40

3.5 Six selected monitoring stations in Larissa, Greece. A1 configuration, based on an almost equilateral triangular array, comprises TRY1, TRY3, TRY6, and TRYN. A2 configuration, random geometry, is based on stations TRY1, TRY2, TRY3, and TRY4 that are further apart.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 43

List of Figures

3.6 Feature maps of the first, second, and fourth convolutional layer of the three CNNs with seismic events as input.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 47

3.7 Feature maps of the second fully connected layer of the three CNNs for four classes.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 48

4.1 STFT-based CNN for seismic classification.  $K_r$  denotes the number of kernels, and 'Flatten' function transforms the input data into a 1D array.  
 © 2024 IEEE. Reprinted, with permission, from J. Jiang, V. Stankovic, L. Stankovic, D. Murray, and S. Pytharouli, "Explainable AI for Transparent Seismic Signal Classification," IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024. . . . . 57

4.2 Flowchart of the proposed human-on-the-loop process. . . . . 62

4.3 Correctly classified examples of earthquake, quake and rockfall: The first column shows the time-series signal, middle column the STFT, and the right column is the LRP relevance heatmap. . . . . 65

4.4 Misclassified examples. . . . . 68

4.5 Waveform (left) and STFT map (right) of the unfiltered low-frequency earthquake. . . . . 69

4.6 Waveform (left), STFT map (middle) and the LRP map (right) of the filtered low-frequency earthquake. . . . . 69

4.7 Three examples of events with labels corrected. . . . . 70

List of Figures

5.1	Generative-based SSL using auto-encoder. 'Kr' stands for kernel. The left part presents the encoder structure, while the decoder, removed after training, is shown on the right. . . . .	80
5.2	2-D Feature visualization by t-SNE: (a) shows the features labelled by groundtruth, and (b) shows the features labelled by K-means clustering results. . . . .	84
5.3	Visualization of the $17 \times 17$ SOM Clustering Results. (a) Distribution of seismic events on the SOM, and (b) Proportions of seismic events per SOM node. . . . .	88
5.4	Accuracy Comparison of k-means, ANN, and SOM Methods with Varying Expert Labelling Workload Percentages . . . . .	89

# List of Tables

2.1	The number of labelled events in Résif catalogue. . . . .	20
3.1	Summary of prior work on using deep learning models for seismic signal classification. UUSS stands for University of Utah Seismic Stations, UUEB for Unconstrained Utah Event Bulletin, NCEDC stands Northern California Earthquake Data Center. NECIS stands for National Earthquake Comprehensive Information System, IRIC for Incorporated Research Institutions for Seismology, STEAD for Stanford Earthquake Dataset and KNMI for Royal Netherlands Meteorological Institute. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023. . . . .	27
3.2	The classification performance results for the three six-channel models. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023. . . . .	34

List of Tables

3.3 The classification performance results for the three single-channel models.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 35

3.4 The confusion matrix for the three six-channel models.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 36

3.5 The confusion matrix for the two 'deepquake' models [1].  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 39

3.6 Larissa results from proposed CNN on one hour data using readings from TYR 1, 3, 6, and N stations (A1): Comparison between automatic and manual detection.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 44

3.7 Larissa results from proposed CNN on one hour data using readings from TYR 1, 2, 3 and 4 stations (A2): Comparison between automatic and manual detection.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 45

List of Tables

3.8 The execution time for each model in ms.  
 © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023. . . . . 48

4.1 The number of labeled events chosen from the original Résif catalogue . 58

4.2 Confusion Matrix - Proposed CNN-based network with post-processing against expert labels (the numbers in brackets indicate recall rates). . . 63

4.3 The confusion matrix after label correction. The numbers in the brackets show the recall. . . . . 71

4.4 The classification performance after label correction. . . . . 71

5.1 Comparison of K-means Classification Performance Across Different Dimensionality Reduction Methods. . . . . 83

5.2 K-means Classification Performance Using 2D t-SNE Features . . . . . 84

5.3 K-means Confusion Matrix Using 2D t-SNE Features . . . . . 84

5.4 The classification performance results for the proposed SSL-based ANN method. . . . . 85

5.5 The confusion matrix for the proposed ANN-based method. . . . . 85

5.6 The classification performance results for SOM. . . . . 86

5.7 The confusion matrix for the SOM. . . . . 87

A.1 Detection results of seismic events (Larisa Dataset, Part 1/3) . . . . . 113

A.2 Detection results of seismic events (Larisa Dataset, Part 2/3) . . . . . 114

A.3 Detection results of seismic events (Larisa Dataset, Part 3/3) . . . . . 115

B.1 t-SNE 2-D Classification Performance . . . . . 116

B.2 t-SNE 2-D Confusion Matrix . . . . . 117

B.3 t-SNE 3-D Classification Performance . . . . . 117

B.4 t-SNE 3-D Confusion Matrix . . . . . 117

B.5 Original Dimension (256) Classification Performance . . . . . 117

List of Tables

B.6 Original Dimension (256) Confusion Matrix . . . . . 117

B.7 PCA 2-D Classification Performance . . . . . 117

B.8 PCA 2-D Confusion Matrix . . . . . 118

B.9 PCA 3-D (Explained variance: 79.59%) Classification Performance . . . 118

B.10 PCA 3-D (Explained variance: 79.59%) Confusion Matrix . . . . . 118

B.11 PCA 10-D (Explained variance: 95.3%) Classification Performance . . . 118

B.12 PCA 10-D (Explained variance: 95.3%) Confusion Matrix . . . . . 118

B.13 33.3% (3-fold) Classification Performance . . . . . 119

B.14 33.3% (3-fold) Confusion Matrix . . . . . 119

B.15 20% (5-fold) Classification Performance . . . . . 119

B.16 20% (5-fold) Confusion Matrix . . . . . 119

B.17 10% (10-fold) Classification Performance . . . . . 120

B.18 10% (10-fold) Confusion Matrix . . . . . 120

B.19 5% (20-fold) Classification Performance . . . . . 120

B.20 5% (20-fold) Confusion Matrix . . . . . 120

B.21 Classification performance results for SOM with 8x8 nodes (4.6%). . . . 120

B.22 Confusion matrix for SOM with 8x8 nodes (4.6%). . . . . 120

B.23 Classification performance results for SOM with 12x12 nodes (10.4%). . . 121

B.24 Confusion matrix for SOM with 12x12 nodes (10.4%). . . . . 121

B.25 Classification performance results for SOM with 17x17 nodes (21.0%). . . 121

B.26 Confusion matrix for SOM with 17x17 nodes (21.0%). . . . . 121

B.27 Classification performance results for SOM with 21x21 nodes (32.1%). . . 121

B.28 Confusion matrix for SOM with 21x21 nodes (32.1%). . . . . 121

# Acronyms

<b>3C</b>	Three-Component
<b>AI</b>	Artificial Intelligence
<b>ANN</b>	Artificial Neural Network
<b>BPF</b>	BandPass Filtering
<b>CNN</b>	Convolutional Neural Network
<b>CWT</b>	Continuous Wavelet Transform
<b>CV</b>	Computer Vision
<b>DCT</b>	Discrete Cosine Transform
<b>DL</b>	Deep Learning
<b>FC</b>	Fully Connected
<i>FN</i>	False Negative
<i>FP</i>	False Positive
<b>GAN</b>	Generative Adversarial Network
<b>GPUs</b>	Graphic Processing Units
<b>HMM</b>	Hidden Markov Model
<b>KNN</b>	K-Nearest Neighbor
<b>LDA</b>	Linear Discriminant Analysis
<b>LRP</b>	Layer-wise relevance propagation
<b>LSTM</b>	Long-Short-Term Memory

Chapter 0. Acronyms

<b>MLP</b>	Multilayer Perceptron Neural Network
<b>NB</b>	Naive Bayes
<b>NLP</b>	Natural Language Processing
<b>PCA</b>	Principal Component Analysis
<b>ReLU</b>	Rectified Linear Unit
<b>RF</b>	Random Forest
<b>SDG</b>	Sustainable Development Goal
<b>SimCLR</b>	Simple Framework for Contrastive Learning of Representations
<b>SSL</b>	Self-Supervised Learning
<b>SNR</b>	Signal-to-Noise Ratio
<b>SOM</b>	Self-Organizing Map
<b>STA/LTA</b>	Short-Term-Average/Long-Term-Average
<b>STFT</b>	Short-Time Fourier Transform
<b>SVM</b>	Support Vector Machine
<b>t-SNE</b>	t-distributed Stochastic Neighbour Embedding
<b>XAI</b>	Explainable Artificial Intelligence

# Acknowledgements

I would like to start by thanking my supervisors, Prof. Vladimir Stankovic, Prof. Lina Stankovic, and Dr. Stella Pytharouli, for their constant support and inspiration throughout my Ph.D. studies. Without their help and encouragement, this work would not have been possible. I am also very grateful to Dr. Jiangfeng Li, Dr. David Murray, Dr. Clement Hibert, Dr. Jean-Philippe Malet, and Dr. Emmanouil Parastatidis for their valuable advice, support, and feedback. I also want to thank all the members of the CIDCOM group, especially Jun Dong, Apostolos Parasyris, Tamara Sobot, Apostolos Vavouris, and Dorde Batic, for their assistance during my research. It has been a real pleasure and honor to work with them and learn from them.

I would like to acknowledge the financial support I received from the University of Strathclyde Faculty of Engineering Scholarship (2020-2023), without which this work would not have been possible. This work was supported by EPSRC Prosperity Partnership research and innovation programme EP/S005560/1 and in part by EPSRC New Horizons research programme EP/X01777X/1. The contextual data interpretation and labelling work by experts on the SZ dataset was supported by RSE Saltire International Collaboration Awards.

Finally, I would like to express my sincere gratitude to my internal examiner, Dr. Javier Cardona, and my external examiner, Dr. Brian Baptie, for their insightful feedback and constructive suggestions during the examination process. Their comments greatly helped to improve the final version of this thesis.

Chapter 0. Acknowledgements

– May 30, 2025 –

# Chapter 1

## Introduction

Passive seismic monitoring enables the detection of a wide range of seismic events generated by natural and anthropogenic processes. These events can include tectonic earthquakes, microseismic activity, and signals induced by landslides [2–4] or geothermal activities. One of the most common applications of passive seismic monitoring is landslide monitoring, where seismic activity such as rockfalls, tremor-like signals, and quakes are specifically triggered by landslide processes, including the formation of surface fissures. Studying such microseismic events helps improve our understanding of subsurface processes, with practical applications in areas such as landslide prediction, mining, and geothermal exploration. However, detecting these events from seismic recordings is challenging due to their low magnitudes and high attenuation. Manual identification and labeling are labour-intensive and prone to subjectivity, highlighting the need for automated methods to detect and classify seismic events accurately.

It is particularly important to discriminate between different types of seismic events associated with landslides, such as precursory rockfall signals, shallow surface slips, or deeper slope deformations, because they can provide critical early warnings and insights into ongoing slope instability. Real-time monitoring and classification of these signals enhance our ability to observe behavioral changes in a slope over time, contributing to more informed hazard assessments. This is especially crucial given the increased frequency and intensity of landslides globally, driven by factors such as climate change, deforestation, and extreme weather events [5]. Improved understanding of landslide dy-

namics through seismic analysis directly supports risk mitigation efforts, infrastructure protection, and public safety planning in vulnerable regions [6].

With higher availability in seismic recordings and advances in Artificial Intelligence (AI), seismic signal analysis has become a very much data-driven field and has spread well beyond seismology and geoscience, as it is now of interest to much broader research communities [7]. Due to the availability of many well-maintained datasets, such as the Southern California Seismic Network and the Stanford Earthquake Dataset [8], the number of data-driven deep learning approaches used in seismology has sky-rocketed in recent years [7]. These models are often trained on enormous volumes of seismic data, sometimes consisting of millions of waveform segments. While not all of this data is labeled, several large-scale datasets, such as Stanford Earthquake Dataset, provide hundreds of thousands of labeled seismic events, which are essential for supervised learning tasks. In the context of landslide-related seismic monitoring, where data are often noisy and manually analyzing signals is extremely labour-intensive, machine learning techniques offer a powerful alternative. ML models are particularly well-suited to this domain because of their ability to handle complex, high-dimensional, and noisy data, making it possible to automate detection and classification of subtle or overlapping event types that would be difficult to distinguish manually. Deep learning-based approaches dominate recent literature, including seismic event labelling [9], estimating seismic events magnitude [10] and event localisation [11]. Various deep learning architectures have been proposed, and a detailed review of deep learning architectures for seismic signal classification can be found in Chapter 3.

Deep learning has become the state of the art for detecting and classifying (micro)seismic events. However, it also presents some disadvantages. First, deep learning models are often considered “black boxes” because of their lack of interpretability. Even though these models can achieve high accuracy, it is difficult to understand how they arrive at their predictions or decisions. The complexity of deep neural networks, especially with many layers and parameters, makes it challenging to explain their inner workings in human-understandable terms. This lack of transparency is particularly problematic in sensitive areas like seismic event analysis, where interpretability is im-

portant for trust and accountability [12]. Second, deep learning models require vast amounts of labeled data to train effectively. In supervised learning tasks, the quality and quantity of labeled data directly impact the model's performance. However, obtaining large, high-quality labeled datasets can be both time-consuming and expensive. This data dependency limits the applicability of deep learning in domains like seismic event analysis where labeled data is scarce or difficult to acquire [13].

To address these challenges, this thesis not only proposes efficient deep learning models for microseismic event detection and classification but also explores the Explainable AI (XAI) tools and self-supervised learning techniques. These approaches aim to enhance the interpretability of deep learning models and significantly reduce the amount of labeled data required for training, making the models more transparent, and hence trustworthy, and accessible in data-limited domains like (micro)seismic event analysis.

## 1.1 Research Aim and Objectives

The aim of this thesis is to develop deep learning-based algorithms for microseismic detection and classification that achieve high accuracy, scalability, and transferability, while requiring minimal computational resources. Additionally, it aims to improve trustworthiness of the models, by exploring techniques that enhance the interpretability of deep learning models and addressing the challenge of large amounts of annotated data required for training. In particular, the following three objectives are set.

- Develop an automatic deep-learning based microseismic event detection and classification algorithm using Convolutional Neural Network (CNN), and evaluate its performance, transferability, and computational efficiency.
- Use advanced Explainable Artificial Intelligence (XAI) tools to enhance the interpretability of deep learning algorithms for classifying microseismic events. Explain model decisions, identify reasons for misclassifications, and improve the accuracy of manual annotation by experts through these explanations.

- Develop self-supervised algorithms to reduce the amount of labeled data required for training deep learning models to below 20%, while maintaining high accuracy. Additionally, develop algorithms that significantly improve the efficiency of manual annotation.

## 1.2 Contributions of the thesis

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

1. This research proposes three CNN models aimed at classifying four types of events: earthquakes, rockfalls, quakes, and anthropogenic noise. To capture a range of time-domain and frequency features at various scales, the models take not only filtered raw time-series waveforms as input, but also Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) coefficients. This approach results in three distinct architectures, each tailored to a specific input type. Additionally, the transferability of the pre-trained model, initially trained on seismic data from the Super-Sauze landslide in Southeast France, is evaluated using a geologically distinct dataset from Larissa region in mainland Greece. (See Chapter 3.)
  - (a) Three CNN-based multi-classifier models for three different inputs (time series, STFT maps and CWT maps) for classification of three different micro-seismicity types plus anthropogenic noise on continuous recordings.
  - (b) Detailed evaluation and analysis of classification performance of the three models, including reliability of the results, and analysis of correctly and incorrectly classified examples to shed light into the most important features of the input signal and reasons for mis-classification.
  - (c) Evaluation of transferability of the proposed model by testing the CNN model pre-trained on Résif dataset on a geographically-distinct dataset and analysis of how different array typologies affect seismic signal classification

- (d) Release of validated labelled dataset from the Larissa region.
2. This research proposes a human-in-the-loop microseismic classification system that leverages state-of-the-art XAI tools to interpret deep learning models for detecting and classifying micro-seismic signals. The explanations help improve model design and clarify both correct and incorrect predictions. In addition, in cases of discrepancies between predicted and expert-labeled classes, the data is reviewed, and experts use the model’s explanations as references to assess the basis of the misclassification. (See Chapter 4.)
    - (a) Interpretation of outcomes of the CNN model for micro-seismic signal classification through the Layer-wise relevance propagation (LRP) heatmap.
    - (b) A detailed discussion of origins of misclassifications shedding light on how trained deep learning models perceive seismic signals in frequency domain.
    - (c) The human-in-the-loop system assists experts in identifying and correcting mislabeling within datasets, thereby enhancing the datasets’ overall reliability and accuracy.
  3. This research proposes a microseismic signal classification method based on Self-Supervised Learning (SSL), designed to achieve good classification performance with minimal or no labeled data. SSL is employed for representation learning, extracting features from raw waveform recordings without the need for labels. These extracted features are used in three downstream classification tasks: unsupervised clustering, semi-supervised classification, and post-labeling. (See Chapter 5.)
    - (a) An automatic feature extraction approach for landslide micro-seismic events based on SSL
    - (b) Two new classification models, one fully unsupervised and the other semi-supervised, based on SSL
    - (c) Visualisation of the SSL model’s features for different seismic classes, explaining misclassifications in clustering

- (d) A new method for enhancing the efficiency of manual labeling using Self-Organizing Map (SOM).

### 1.3 Organisation of the thesis

The remainder of the thesis is organised as follows:

Chapter 2 reviews the background, dataset, and general definition of deep learning.

Chapter 3 provides a detailed literature review about deep learning based (micro)seismic event detection and classification. Following this, three convolutional neural network (CNN) models are proposed to classify four event types: earthquakes, rockfalls, seismic sources related to landslide processes (referred to as “quakes”), and anthropogenic noise.

Chapter 4 first introduces XAI and its current applications in (micro)seismic event classification, then proposes a human-in-the-loop system designed to provide comprehensive explanations of the key features learned by a deep neural network in multi-class classification tasks.

Chapter 5 proposes a (micro)seismic event feature extraction technique based on self-supervised learning, followed by downstream tasks for unsupervised clustering and semi-supervised classification. Additionally, an algorithm is introduced to significantly enhance the efficiency of expert manual labeling.

Chapter 6 concludes this thesis and discuss the remaining challenges in (micro)seismic event detection and classification.

### 1.4 Publications

#### Journal Articles

- J1 Jiang. J, Stankovic. V, Stankovic. L, Parastatidis. E , and Pytharouli. S, Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-14, 2023.

Contribution: Literature review, Experiments design, Algorithm implementation,

## Chapter 1. Introduction

Results analysis, Formal analysis, Draft writing.

The publication relates to Chapter 3

- J2 Jiang. J, Murray. D, Stankovic. V, Stankovic. L, Hibert. C, Pytharouli. S, and Malet. J.-P, A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model. *Science of Remote Sensing*, vol. 11, pp.100189, 2025..

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

The publication relates to Chapter 4

- J3 Jiang. J, Stankovic. V, Stankovic. L, Murray. D, and Pytharouli. S, Generative self-supervised learning for seismic event classification. *Under Review in Engineering Applications of Artificial Intelligence*, 2025.

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

The publication relates to Chapter 5

### Conference Papers

- C1 Jiang. J, Stankovic. V, Stankovic. L, and Pytharouli. S, Automatic detection and classification of microseismic events from super-sauze landslide using convolutional neural networks. *2020 AGU Fall Meeting Abstracts*, 2020.

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

The publication relates to Chapter 3

- C2 Jiang. J, Stankovic. V, Stankovic. L, Murray, D , and Pytharouli. S, Explainable AI for transparent seismic signal classification. *2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2024.

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

The publication relates to Chapter 4

## Chapter 2

# Background

### 2.1 (Micro)seismic Event Analysis

Seismic events can be intuitively described as sudden brittle failures due to natural causes or artificially triggered explosions, leading to a rapid release of energy in a localised area that generates waves propagating through the surrounding medium [14]. For example, in volcanic seismology, seismic signals are generated by physical processes within a volcano, such as the movement of gas and fluids (e.g., water, magma) and their interaction with solid rock. These signals include volcanic-tectonic events, long-period signals, tremors, and quakes [15]. In processes such as hydraulic fracturing and carbon capture, seismic sources are linked to activities like string shots, perforation shots, plug settings, ball drops, and sleeve openings, with source mechanisms often categorised as volumetric or shear double-couple components [14, 16]. Additionally, seismic events can occur from unstable slopes, such as in open-pit mining or landslides, due to elastic strain accumulation, rupture, friction, and shear between soil particles [17].

Seismic monitoring of slope instabilities emerged in the 1960s and has been significantly advanced in recent decades through the development of microseismic monitoring techniques [17]. These advances are attributed to improvements in seismometer technology (higher sensitivity compared to other slope monitoring methods), simpler installation (lower costs and reduced power requirements), and increased network density [18]. The primary goal of analysing (micro)seismic events is to support geological

hazard assessments, enhance the understanding of underlying processes, and prevent future failures to reduce the social and economic impacts of destructive events [18].

In some instances, (micro)seismic signals can act as precursors to landslides, exhibiting low dominant frequencies and nonstationary characteristics. Thus, building a solid understanding of various seismic signals is essential for effectively predicting major geological disturbances and minimizing the risks of fatalities and infrastructure damage [17, 19, 20]. Traditional analysis methods typically involve three steps: 1) applying detection techniques to identify the event time period; 2) extracting meaningful features from raw signal; 3) classifying the extracted features using classification algorithms to determine the classes of (micro)seismic event.

### 2.1.1 System overview

To comprehensively analyse (micro)seismic events using traditional methods, the process can be broken down into four primary stages: signal acquisition, detection, feature engineering, and classification. A block diagram illustrating this workflow is provided in Figure 2.1. Each stage is briefly outlined in this section and further detailed in the subsequent sections. (1) Signal Acquisition: The process begins with acquiring seismic or microseismic signals using a network of sensors such as seismometers or accelerometers. These signals may originate from natural phenomena (e.g., volcanic activity, landslides) or anthropogenic activities (e.g., hydraulic fracturing, mining). (2) Detection: Next, algorithms identify segments of the continuous signal that correspond to potential (micro)seismic events, differentiating them from ambient noise or irrelevant signals. (3) Feature Engineering: In this step, the detected signals are processed to extract meaningful features, which are critical for effective event classification. This involves a combination of feature construction, extraction, and selection. (4) Classification: Finally, extracted features are fed into classification algorithms to categorise seismic sources based on predefined classes such as earthquakes, rockfalls, or anthropogenic noise.



Figure 2.1: Block diagram of the traditional system overview for (micro)seismic event analysis.

### 2.1.2 Data Acquisition

The acquisition of (micro)seismic data is the foundational step in seismic event analysis, directly influencing the quality and interpretability of downstream processing tasks such as detection, feature engineering, and classification. Data acquisition typically involves deploying networks of geophysical sensors, including seismometers, geophones, and accelerometers, across the area of interest. These sensors capture ground motion caused by both natural and anthropogenic sources, converting analog vibrations into digital signals that can be processed and analyzed. Depending on the application, acquisition systems can be installed in various configurations: surface arrays, borehole installations, or hybrid setups combining both. Surface sensors are easier and cheaper to deploy but may suffer from high ambient noise levels and lower signal fidelity. Borehole sensors, though more expensive to install, provide superior signal-to-noise ratios (SNR) due to their placement within the subsurface, closer to potential seismic sources [16, 17]. Sensor placement, orientation, and calibration are crucial for accurate waveform capture. The quality of acquired data can be affected by environmental noise (e.g., wind, rainfall, human activity) and instrumental limitations (e.g., sensor drift, gain inconsistencies). Therefore, pre-processing steps such as filtering, baseline correction, and sensor response removal are typically applied before detection and feature extraction [21].

### 2.1.3 Detection

In seismic analysis, efficient and highly accurate signal detection is important and has a wide range of applications. For example, in earthquake analysis, the identification of distinct wave phases, such as primary (P-waves) and secondary (S-waves), is particularly important. P-waves, the fastest seismic waves, are typically the first to arrive at a seismic station. Their precise arrival time, determined through a process known

as P-wave picking, is a crucial step in earthquake analysis. Detection, on the other hand, refers more broadly, to recognising that an earthquake has occurred and may involve signals from P-waves, S-waves, or other seismic phases. Manually identifying signals of interest from continuous seismic recordings, using expert knowledge, is often time-consuming, labor-intensive, and subjective.

In recent years, there has been a growing shift toward the development of algorithms that automatically detect seismic signals. The most widely used method is the Short-Term Average to Long-Term Average ratio (STA/LTA), which applies sliding windows of predefined lengths to the time/frequency domain amplitude, envelope, or higher-order statistical features (e.g., skewness and kurtosis) of the recorded signal [22–27]. Events are detected by setting trigger and dettrigger thresholds based on the ratio of the short- and long-window average values. STA/LTA and its variants have been commonly used in detecting seismic events from earthquakes [28], volcanic activity [29], and landslides [25–27]. However, STA/LTA has some significant limitations, such as 1) improper parameter initialization leading to false alarms [30]; 2) high sensitivity to ambient noise [19, 31, 32]; and 3) time-consuming and inefficient processes for selecting window lengths and thresholds [33, 34]. Based on STA/LTA, a modified energy ratio detection method has been introduced, utilizing equal-length pre- and post-sample windows [35]. Later, three enhanced detection methods using seismic attributes—energy ratio, fractal dimension, and entropy—were developed [36]. The Akaike Information Criterion (AIC) detection algorithm proposed by [37] relies on the idea that nonstationary seismic signals can be segmented into locally stationary parts, each approximated as an autoregressive process. Another approach, known as matched filtering, which is based on template matching, requires prior knowledge of representative parent waveforms [38, 39].

Low signal-to-noise ratio (SNR) and varying levels of ambient noise present significant challenges for detecting (micro)seismic events [34, 40]. Traditional detection methods work well for known seismic events but struggle to detect unknown or unexpected low SNR microseismic events, such as landslide-induced quakes or rockfalls [15, 17, 34].

### 2.1.4 Feature Engineering

In traditional (micro)seismic event analysis, feature engineering (including feature construction, extraction, and selection) plays a critical role in efficient signal classification. A large set of features with redundant information can increase processing time and lead to issues such as classifier overfitting, multicollinearity, and suboptimal feature ranking during the selection process [41]. A comprehensive review of feature construction for (micro)seismic events is provided in [19], where temporal, spectral, and cepstral features, along with their combinations, are derived from raw, denoised data.

Feature extraction and selection are frequently employed to reduce the dimensionality of the feature space, thereby minimizing storage requirements and reducing the testing and training time of classifiers. Among these, Principal Component Analysis (PCA) is the most widely used feature extraction and dimensionality reduction technique. It has consistently proven effective for a variety of (micro)seismic event analysis tasks [42, 43]. Besides PCA, many other dimensionality reduction methods have been used for (micro)-seismic signal analysis, including, t-distributed Stochastic Neighbor Embedding (t-SNE) and Linear Discriminant Analysis (LDA) [44].

Selecting the most discriminative features for a given task is not straightforward. Many feature selection methods have been proposed, typically classified into several types: filter-based (the most common in (micro)seismic analysis), wrapper-based, embedded, hybrid, and ensemble approaches. Filter-based methods, which assess and select features using various statistical tests, are model-agnostic, meaning they can be applied to any learning algorithm to eliminate irrelevant and redundant features, and they have lower computational complexity [45]. These feature selection methods have been widely explored in various (micro)seismic analyses.

Feature engineering in traditional machine learning for seismic event classification is a labor-intensive process, requiring substantial manual effort to design, construct, and select relevant features. This approach heavily relies on domain expertise, introducing potential biases and subjectivity, which may result in incomplete or suboptimal feature sets. Manually designed features can fail to capture the full complexity of the raw data, limiting the model's ability to recognize intricate patterns in seismic events. Moreover,

the challenge of selecting the right features can lead to issues like overfitting, multicollinearity, or reduced model performance. Predefined feature sets further constrain the model’s adaptability to new or unknown seismic events, making it less effective in dynamic environments compared to deep learning models, which automatically learn relevant features directly from raw data.

### 2.1.5 Classification

In the context of (micro)seismic signal analysis, classification refers to the process of categorizing seismic signals into predefined classes based on their characteristics, such as frequency range, amplitude, signal shape, and duration. The primary goal is to distinguish events of interest (e.g., (micro-)earthquakes, rockfalls) from noise or other signals, enabling better understanding and monitoring of seismic activity. The main challenges in (micro)seismic classification include: (1) a lack of openly accessible annotated datasets [19]; (2) an imbalanced catalog of labeled events due to the scarcity of events of interest [19]; and (3) the high degree of similarity between unknown natural and anthropogenic “interfering” signals and the events of interest in either the time or frequency domain [26]. As a result, manual classification remains prevalent, where experts analyze signals based on their characteristics to assign categories such as (micro-)earthquakes, block falls, rockfalls, quarry blasts, and multiple events [46].

Automatic classification methods, such as Support Vector Machine (SVM) and Random Forest (RF), are typically applied alongside various feature extraction or selection techniques, as previously discussed. SVMs are effective for addressing high-dimensional nonlinear classification challenges with a limited number of training samples [47], and they have been used to differentiate between long-period events, tremors, and volcanic tectonics in [15, 48], as well as between earthquakes and non-earthquake events in [49], among others. RF is parallelizable, performs well with high-dimensional signals, is fast in both training and prediction, robust to outliers and non-linear data, can manage imbalanced datasets, and has low bias [50]. It has been utilized to classify landslide (micro)seismic events, including rockfalls, slide quakes, earthquakes, and both natural and anthropogenic noise [18, 26]. Other classification models, such as

Multilayer Perceptron Neural Network (MLP), Linear Discriminant Analysis (LDA), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Feed-Forward Back-Propagation Neural Networks, have also been explored for distinguishing volcanic-seismic events [51, 52]. Conventional (micro)seismic event classification heavily relies on handcrafted feature construction, extraction, and selection, which typically necessitates substantial domain knowledge for the physical characterization of events.

Traditional classifiers, in addition to requiring manual feature engineering, often struggle to capture complex, nonlinear relationships in data. These models rely on predefined features and lack the flexibility to adapt to new data types or variations. They may overfit on small datasets, especially when feature selection is not optimal, leading to poor generalization. Furthermore, traditional classifiers face scalability issues with larger datasets or high-dimensional data, and their fixed feature sets make them less responsive to emerging patterns or anomalies. In contrast, deep learning models handle complex relationships more effectively, scale efficiently with larger datasets, and offer better performance, particularly in tasks involving high-dimensional or complex data, such as seismic event classification.

## 2.2 Deep Learning Architectures

Deep learning is a subset of machine learning that has gained prominence in recent years due to its effectiveness in tasks involving large amounts of data. Unlike traditional machine learning methods that rely heavily on feature engineering, deep learning models automatically learn hierarchical features from raw data, which makes them particularly useful for tasks like image recognition, speech processing, and natural language understanding. Deep learning is built on artificial neural networks, with multiple layers of interconnected units (neurons) mimicking the structure and function of the human brain. These multi-layered networks, often referred to as DNN, allow models to learn complex patterns and representations from data. The backbone of deep learning is the neural network architecture, which consists of an input layer, several hidden layers, and an output layer. Each layer comprises neurons, or nodes, which perform simple

calculations and pass the output to the next layer. The network’s learning process relies on minimizing a loss function using an optimization algorithm, commonly gradient descent, through a process known as backpropagation. In recent years, the availability of large datasets, advanced hardware, especially Graphic Processing Units (GPUs), and improved training algorithms have made deep learning models more practical and scalable for a wide range of applications [53].

One of the most successful deep learning architectures is the CNN [54]. CNNs are specifically designed for processing data with a grid-like topology, such as images, where spatial relationships between pixels are critical. A CNN typically consists of three key types of layers: 1) Convolutional Layers, in this layer, filters (or kernels) slide across the input image to extract local patterns, such as edges, textures, or more complex features. These filters allow CNNs to capture spatial hierarchies in images by learning increasingly abstract features at deeper layers; 2) Pooling layers are used to reduce the spatial dimensions of feature maps, which decreases computational load and minimizes the risk of overfitting. The most common pooling method is max pooling, which selects the maximum value from each region covered by the filter, thus preserving the most prominent features; 3) Fully Connected Layers, after several convolutional and pooling layers, the feature maps are flattened and passed through fully connected layers. These layers combine the extracted features to produce the final output, such as classifying an image into a specific category.

The proposal of CNNs has demonstrated impressive performance in various applications. The architecture is highly effective at tasks such as image classification and object detection. Some of the most famous generic CNN models include LeNet-5, one of the earliest architectures designed for digit recognition [55]; AlexNet, which significantly advanced image classification by utilizing ReLU activation and dropout [54]; and ResNet, which introduced skip connections to enable the training of very deep networks [56]. The VGG16 model [57] is more recent, and currently one of the best CNN architecture, renowned for its simplicity and uniformity in design. The key innovation in VGG16 lies in its use of smaller  $3 \times 3$  convolution filters throughout the network, stacked in increasing depth, allowing the model to capture fine-grained patterns in the

data while maintaining computational efficiency. By using smaller filters but increasing the number of convolutional layers, VGG16 is able to extract more complex features progressively, which improves its performance in tasks like classification. The deep learning models proposed in this thesis are mostly inspired by VGG16.

Though demonstrating competitive performance, deep learning presents some disadvantages. First, deep learning models are often considered “black boxes” because of their lack of interpretability. Even though these models can achieve high accuracy, it is difficult to understand how they arrive at their predictions or decisions. The complexity of deep neural networks, especially with many layers and parameters, makes it challenging to explain their inner workings in human-understandable terms. This lack of transparency is particularly problematic in sensitive areas like seismic event analysis, where interpretability is important for trust and accountability [12]. Second, deep learning models require vast amounts of labeled data to train effectively. In supervised learning tasks, the quality and quantity of labeled data directly impact the model’s performance. However, obtaining large, high-quality labeled datasets can be both time-consuming and expensive. This data dependency limits the applicability of deep learning in domains like seismic event analysis where labeled data is scarce or difficult to acquire [13].

### 2.2.1 Performance measures

To evaluate performance of the proposed models, standard classification performance measures, namely precision, recall (also referred to as sensitivity), and the F1 score, are utilized as defined in [19]:

$$Precision = \frac{TP}{TP + FP} \quad (2.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (2.3)$$

True positive (TP) is the number of correctly detected and classified events, i.e.,

the input window contains the waveform of the target event, and the model correctly classified it. False positive (FP) is the number of false alarms, i.e., an event of the non-target class is misclassified as the target class. False negative (FN) is the number of missed events, i.e., an event of the target class is misclassified as another classes or it is not detected.

In the context of seismic event detection, recall is often considered more critical than precision or even the F1 score, particularly in applications such as early warning systems or hazard monitoring. This is because missing a true seismic event (false negative) can lead to severe consequences, including delayed response to natural hazards and failure to issue timely alerts. In contrast, false alarms (false positives), while inconvenient, are generally more tolerable in real-world applications [58].

## 2.3 Dataset

The dataset utilized in this thesis is publicly available through the Résif Seismological Data Portal and was collected by the French Landslide Observatory, OMIV (Observatoire Multi-disciplinaire des Instabilités de Versants) [59]. The data was gathered using the MT network, specifically from Super-Sauze C (SZC) stations, located on the east and west sides of the Super-Sauze landslide in Southeast France (Latitude: 44.34787, Longitude: 6.67805), as shown in Figure 2.2. Further information on the sensors and terrain can be found in [60]. The signals were recorded during three periods: from 11 October to 19 November 2013; from 10 to 30 November 2014; and from 9 June to 15 August 2015. The seismic stations were equipped with short-period seismometers (Noe-max and Sercel L4C), which have a flat frequency response in the range of 5–100 Hz, enabling detection of a broad range of seismic events. The signals were recorded with broadband seismic recorders (RefTek 130S-01) at a sample frequency of 250 Hz. The array consisted of a tripartite arrangement of 40 m layout, centered around one three-component seismometer and three vertical, one-component seismometers arranged in an equilateral triangle configuration, resulting in six channels for data acquisition [26].

In addition to the raw seismic data, a catalogue of labelled events is provided. The dataset contains four distinct types of seismic events: earthquakes, quakes, rock-

## Chapter 2. Background

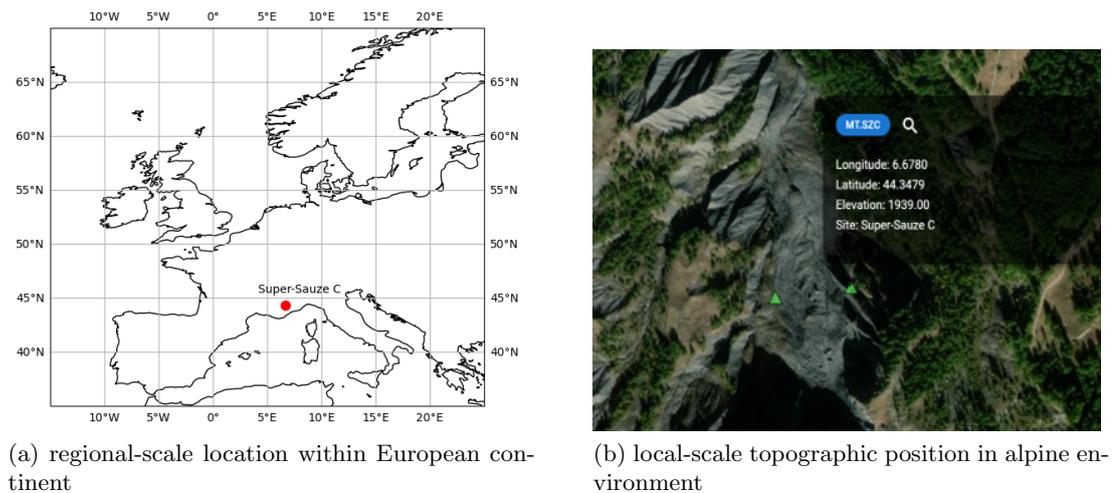


Figure 2.2: Geographic location of the SZC sensor station (a) global distribution and (b) detailed topographic positioning.

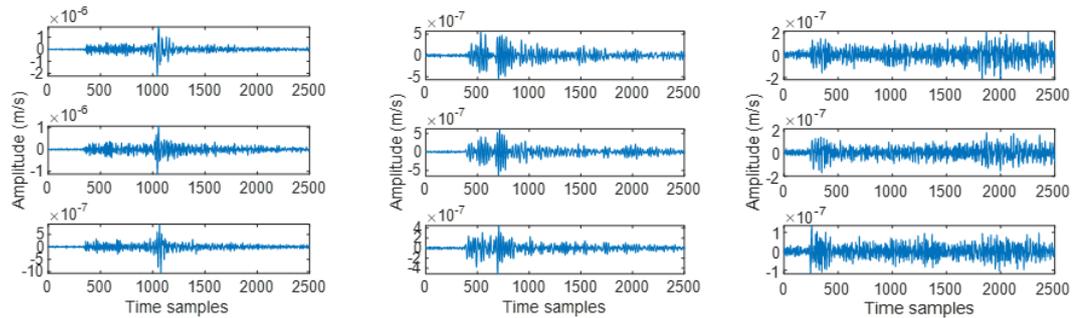
falls, and natural/anthropogenic (N/A) noise signals [26]. Table 2.1 displays the total number of events per class in the used dataset, after removing duplicates. Rockfalls predominantly occur at the main scarp of the landslide, where rigid blocks fall from a steep slope (height over 100m). Quakes are defined as low-magnitude endogenous seismic events that are hypothesized to originate from several internal mechanisms within the landslide body. These include shear deformation along basal or internal slip surfaces, fracturing or collapse of voids, and the progressive accumulation of strain leading to micro-failures. These processes generate seismic signals that are typically short in duration (less than 5 seconds) and display strong attenuation, often being recorded by only a few nearby seismometers. The location and waveform characteristics of these signals reflect their origin within the deforming landslide mass, distinguishing them from other event types. These physical characteristics enable their classification as a distinct event class in the dataset [26]. In contrast, earthquakes correspond to exogenous seismic events, including regional tectonic activity and teleseisms, with stronger signal amplitudes and broader spatial detectability. N/A noise events include various environmental and human-made noise sources. Environmental noises originate from phenomena such as wind or heavy rainfall, while anthropogenic noises come from activities like traffic, pedestrians, or helicopters. For further details on the endogenous

seismicity at the Super-Sauze landslide, refer to [61], [62], and [3]. Figure 2.3 presents examples of seismic signals, including earthquakes, quakes, and rockfalls, with varying Signal-to-Noise Ratio (SNR), demonstrating the range of noisy events in the dataset.

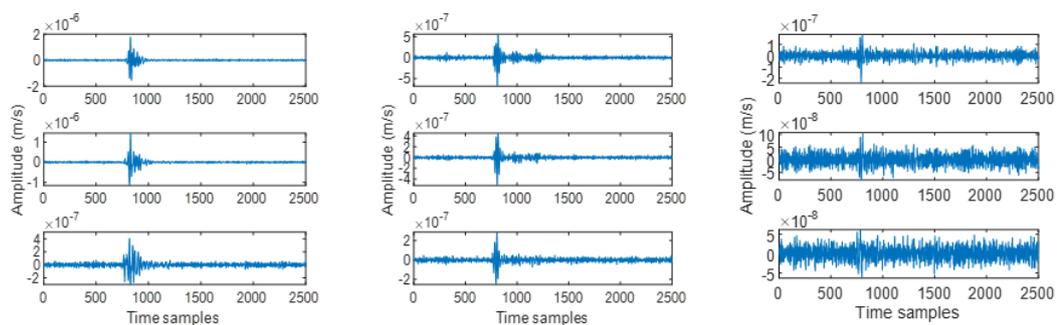
This study focuses on seismic events collected on landslides, a topic that has received comparatively less attention than earthquakes in the field of automatic seismic event detection. Most existing detection algorithms are primarily developed and evaluated on earthquake datasets. This is partly due to the availability of large, well-annotated seismic datasets in tectonic contexts, such as those from the Southern California Seismic Network or the Stanford Earthquake Dataset, which include millions of labeled events. In contrast, annotated datasets focused on landslide-related seismicity remain scarce. By investigating landslide-related seismicity, this work aims to fill this research gap and highlight the potential for improving monitoring systems in landslide-prone regions. Detecting seismic signals generated by landslides is important for understanding the internal dynamics of slope failure, identifying precursory signals, and ultimately mitigating geological hazards that threaten infrastructure and human lives [63].

Table 2.1: The number of labelled events in Résif catalogue.

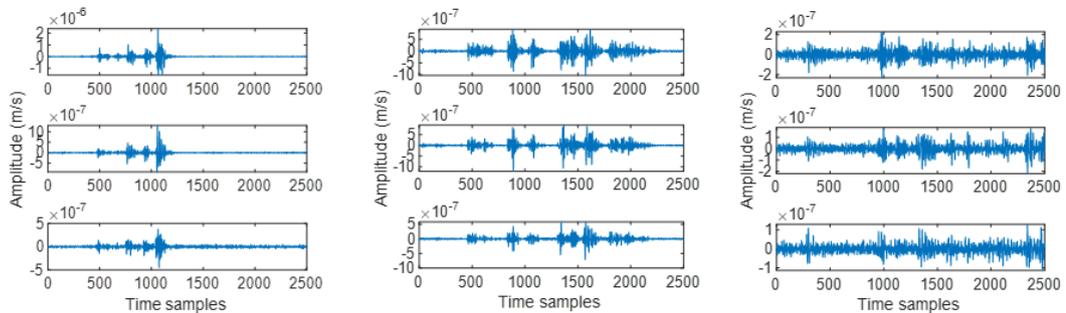
Class	Total No. events
Earthquake	388
Quake	234
Rockfall	401
Noise	351



(a) Earthquakes



(b) Quakes



(c) Rockfalls

Figure 2.3: Examples of seismic events: (a) Earthquake (b) Quake (c) Rockfall. Three different events are shown for each class, including high, medium and low SNR cases. In all cases, we show three waveforms, representing, respectively, the East, North and vertical direction of the three-component sensors from top to bottom.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

## Chapter 3

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

– May 30, 2025 –

### 3.1 Introduction

As discussed in Chapter 1, algorithms for automated detection and classification of seismic events are important but also challenging to develop. Prior work has focused on application of traditional machine learning algorithms to classification of (micro-)seismic events, supported by various signal processing tools for denoising and detection of events. Most classification approaches have been based on well-known algorithms, such as Hidden Markov Model (HMM), SVM and RF [64], [65], [66], [67].

For example, Provost et al. [26] propose a classification method using an RF supervised classifier to classify micro-seismic events on slow-moving landslides. The method uses the Short-Term-Average/Long-Term-Average (STA/LTA) algorithm for detection, then calculates 71 seismic attributes as features inputted into a supervised RF to classify each event into one of four pre-determined classes (earthquake, quake, rockfall and Natural/Anthropogenic noise). Ruano et al. [66] propose a seismic detection system

based on SVMs to improve the detection accuracy of seismic events at the station level. The system processes seismic data from Portugal’s seismographic network using features derived from power spectral density at selected frequencies. Curilem et al. [64] focuses on classifying seismic signals from Villarrica, one of the most active volcanoes in South America. The study specifically looks at three key types of seismic signals: long-period (LP), tremor (TR), and energetic tremor (ET). The authors developed a classifier that processes 30-second signal windows to classify these events. A feature extraction process was applied to identify relevant signal characteristics, and a MLP was optimized with genetic algorithms to classify. Hibert et al. [65] propose a method to detect rockfall events at Piton de la Fournaise volcano by identifying their distinct high-frequency seismic signals within continuous seismic data, filtering out non-rockfall signals using time-frequency analysis. More recently, an end-to-end automated system is proposed in [19] consisting of signal denoising, event detection via statistical Neyman-Pearson based thresholding, feature selection, and graph-based classification.

In contrast to traditional pipeline-based approaches, e.g., [64], [65], [66], [26], [19], deep learning provides an integrated approach to detection, feature representation and classification, with competitive performance under the assumption that a good representative dataset is available for training. Though there have been many attempts to use various deep learning architectures for seismic signal detection and classification (e.g., [11], [68], [69], [70], [71], [72]), classification of microseismic endogenous landslide events based on deep learning is rarely studied. Moreover, transferability of deep learning classification models to different monitoring network geometries is rarely discussed.

In this chapter, we formulate the microseismic classification problem as a multi-class classification task, and propose three CNN models. To capture a variety of time-domain and frequency features at different scale, besides using filtered raw time-series waveforms as input to the network, we also use STFT and CWT coefficients. As typical with time-series data analysis, we slide a fixed length window over the input to achieve multi-classification on continuous data. In addition, we test our pre-trained model on a geologically-distinct dataset from the region of Larissa in mainland Greece, using two

different array geometries, to evaluate its transferability.

## 3.2 Related work

We present a summary of methods for seismic signal classification utilizing Deep Learning (DL). The majority of DL-based seismic classifiers are designed as binary models, typically distinguishing between the event of interest (e.g., earthquakes) and a catch-all category for other events. There are some emerging seismic multi-classifiers which output more than two categories. We also include DL approaches whose end result is classification, although they also perform detection or phase-picking. It is worth noting that an additional detection step is not necessary if performing classification on continuous recordings, parsed into windows. For non-DL approaches, an up-to-date review can be found in [19].

### 3.2.1 Binary classification

In [69], seismic data are sampled and parsed into time windows of 20 msec duration, which are then fed into a CNN. The CNN model, consisting of one 1-D convolutional layer, one pooling layer and 3 Fully Connected (FC) layers, acts as a binary classifier and classifies the input window into earthquake event waveform or noise waveform. Perol et al. [11] propose earthquake detection techniques for multi-channel 1-D data using CNN with 3 channels and sampling rate of 100Hz. The raw recordings were split into 10sec seismic waveform windows that are fed into a trained network consisting of eight 1-D convolutional layers to extract features for earthquake detection, followed by a FC layer to perform the earthquake/noise classification and location estimation by features outputted by convolutional layers.

Besides feeding raw signals, various methods are proposed that take as inputs spectrograms [73], [74]. For example, Dokht et al. [75] propose a CNN model to classify the input spectrogram into earthquake or noise. The authors use Three-Component (3C) spectrograms of 10sec seismic data as input to a CNN architecture consisting of 4 convolutional and 2 FC layers. Each convolutional layer is followed by a max-pooling

### Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

layer. The final output layer is a two-neuron FC layer with softmax activation which outputs the probability distribution of 2 classes (earthquake and noise). Liao et al. [73] use the method of transfer learning to detect whether the CWT map contains first break of earthquake. In particular, the authors use pre-trained CNNs used for image classification tasks, namely, GoogLeNet, AlexNet and SqueezeNet, to perform transfer learning from image data to seismic signals and classify the CWT output into first-break waveform and not first-break waveform. Linville et al. [76] use Long-Short-Term Memory (LSTM) and CNN models to classify seismic events as either quarry blasts or earthquakes. The LSTM model routes input spectrograms to output classes (0 or 1) which represents quarry blasts or earthquakes through 4 bidirectional layers as a many-to-one learning scenario which takes input from many time steps to make one binary classification output. The CNN architecture with 4 convolutional, 4 max-pooling and 2 FC layers outperforms RF, SVM and residual neural network.

Mousavi et al. [77] propose a sequence-to-sequence learning model, ‘EQ-transformer’, for phase picking and earthquake classification using a multi-task structure, that outputs 3 sequences of probabilities, representing presence of earthquake, P-phase picking and S-phase picking. The deep network structure consists of an encoder that converts the raw input signal into features through 1-D convolution, max-pooling, residual convolution, and LSTM layers, and 3 separate decoders. In [78], the authors propose a vision transformer (ViT)-based system for earthquake detection and its magnitude prediction. The system consists of two separate ViT networks: the first one detects earthquake events from the picked P-wave; the second network predicts the magnitude of the detected earthquakes.

In [70], spectrograms of 30 seconds 3-component seismograms are used as input to a CNN-RNN Earthquake Detector (CRED), that consists of convolutional, recurrent and dense layers, in a residual structure. A 2D convolution layer extracts features from the input spectrograms. Then, bi-directional LSTM performs sequence learning. Finally, dense layers classify the extracted features and output a sequence of predicted probabilities, for classification of earthquakes and noise. [79] perform a 7-level CWT on 3-component 30s-long time window with 100Hz sampling frequency to construct a

CWT map as input to an encoder-decoder network with residual learning to classify earthquake signals using 2 classes - earthquake signal or not.

### 3.2.2 Classification of more than one event type

A CNN is proposed in [68] to classify seismic events into 3 categories - tectonic earthquakes, mining-induced events, and mining blasts, based on 90sec long spectrograms as input. The model consists of 4 2-D convolutional and a 3-node softmax activated dense layer.

A ‘deepquake’ CNN architecture is proposed in [1] that classifies 3-component 20sec input data into earthquake, other events and noise. Two CNNs are built for two different input types: time series and spectrograms. The two architectures consist of 6 convolution layers for feature extraction and one dense layer for classification. An attention-based CNN architecture is proposed in [72] using multi-task learning. This architecture first acts as a binary classifier and classifies the seismic waveform into earthquake or noise; then as a multi-class classifier, it classifies the seismic waveform into micro-earthquake, macro-earthquake or noise. The input data is a 10-sec raw seismic waveform with 100Hz sampling rate. 8 1-D convolutional layers (with Relu activation) with an attention module to extract features, and 2 task-specific layers with 2 FC layers (softmax activation) classify the features.

A 3D-CNN/RNN-based architecture is proposed in [10] to classify earthquake magnitudes. Each segment of 60-sec waveform is split into  $6 \times 10$  sec clips, which are then processed and transformed into a 2-D Log-Mel map. Thus, the input data of the model is the 3-D matrix of 6 Log-Mel maps stack that is first inputted to a 3-D CNN and then to RNN. FC layer is used to classify the extracted features into five categories (greater than or equal to 0.0, 1.0, 2.0, 3.0 and 4.0 on Richter scale). In [80], a CNN is used to detect and classify seismic events into microseismic event, single-phase event, and ambient noise events. The model’s input is a  $22 \times 2000$  seismogram image obtained from 22 seismometer channels with 2000 sampling points. The proposed architecture consists of 6 convolutional, 6 pooling, and 2 FC layers. The models are first trained and tested on synthetic data and then used to detect microseismic events from a field data set. It

### Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

is showed that training on synthetic data and testing on the field data leads to poor performance. The results are then improved by labelling the field data and using it for training the model and testing.

Table 3.1: Summary of prior work on using deep learning models for seismic signal classification. UUSS stands for University of Utah Seismic Stations, UUEB for Unconstrained Utah Event Bulletin, NCEDC stands Northern California Earthquake Data Center. NECIS stands for National Earthquake Comprehensive Information System, IRIC for Incorporated Research Institutions for Seismology, STEAD for Stanford Earthquake Dataset and KNMI for Royal Netherlands Meteorological Institute.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

Paper	Year	Aim	Best Architecture	Input data	Dataset & Transf. learning
[76]	2019	seismic classification (quarry blasts and earthquakes)	CNN & LSTM	Spectrograms (three channels)	UUSS
[68]	2019	seismic classification (tectonic earthquakes, mining-induced events, mining)	CNN	Spectrograms (three channels)	UUSS & UUEB [81]
[70]	2019	seismic detection (earthquakes and noise)	residual structure with convolutional, recurrent, and dense	Spectrograms (three channels)	North California [82] transfer learning to Arkansas [83]
[72]	2020	seismic detection (earthquake and noise) and classification (microearthquake, macroearthquake and noise)	attention-based CNN	Time series (three channels)	NECIS [84] & IRIS [85]
[77]	2020	seismic detection and phase picking (earthquake, P-phase and S-phase)	multi-task structure	Time series (three channels)	STEAD [8] transf learning to aftershock region of Tottori
[10]	2021	earthquake magnitude classification (greater than or equal to 0.0, 1.0, 2.0, 3.0 and 4.0 on Richter scale)	3D-CNN RNN	Log-Mel spectrogram (one channel)	STEAD
[79]	2021	Seismic detection (earthquake and noise)	U-NET	CWT map (3 channels)	Northern California transfer learning to Arkansas, Texas [86], Japan and Egypt
[80]	2021	microseismic classification (Dyke-roadway, Longwall, Low-frequency and noise)	CNN	seismogram (6 channels)	an underground coal mine
[9]	2021	Microseismic event classification (microseismic event, single-phase event, and ambient noise)	CNN	seismogram (22 channels)	synthetic data; unsuccessful transfer learning to field data
[78]	2022	seismic detection (P-wave and noise) and magnitude estimation	Vision Transformer	Time series (three channels)	STEAD
[1]	2022	seismic classification (earthquake, other events and noise)	CNN	Time series and seismogram (three channels)	KNMI [87]
This thesis	2023	microseismic classification (earthquake, quake, rockfall and noise)	CNN	Time series, STFT maps and CWT maps (6 channels)	Résif [59] Transfer learning to Larissa [88]

### 3.2.3 Overview of the literature of DL-based seismic signal classification

Table 3.1 summarises the state-of-the-art in DL-based approaches whose end result is classification. In the table, ‘transfer learning’ refers to the application of a pre-trained model to evaluate data from a previously unseen dataset collected at a different location. In [68], [76], [10], [72], [80], [9], a single type of feature is used as input, with no study on feature selection. It can be seen that the majority of the proposed architectures rely on CNNs and primarily use temporal or spectral features, often with 1 to 6 channels. The CNN model of [1] reports classification results that are among the best in the literature, without complex feature engineering. Although using more complex input features and network structures, such as [10], [70], [77], [79], [78], can lead to high performance, the complex pre-processing steps and deep models are computationally demanding. In addition, sequence-to-sequence learning requires clearly labelled start and end times of each event. Limited by the absence of such a dataset for rockfall and quakes, as is usually the case in practice, we adopt a sequence-to-point classification architecture.

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

### 3.3 Methodology

#### 3.3.1 Data processing

First, we apply band-pass filtering to remove high-frequency measurement noise and low-frequency noise from sources such as humans, vehicles, rain, and animals. Specifically, considering the frequency range of the events of interest, the raw recordings are processed using a 3-order Butterworth band-pass filter with a frequency range of 5-60Hz.

We use 3 different model inputs: raw temporal data, STFT and CWT maps. Figure 3.1 shows examples of three classes of seismic events (earthquake, quake and rock-fall). For time-series raw signals, we use 10 seconds window as input. Given the sampling frequency of 250 Hz, the length of the input window is thus 2500 samples, which is usually sufficient to capture the entire seismic event, and short enough to ensure manageable complexity. Thus, for 6-channel recordings, the input time series signal is of dimension  $2500 \times 6$ . To perform STFT, we use Hann window with heuristically set length of 128 samples with 75% overlap, generating output of dimension of  $65 \times 75 \times 6$ . For CWT, as in [79], as mother wavelet, we use the Morlet wavelet with 8 cycles. We construct scalograms using 80 scales spanning frequency range between 5Hz and 60Hz. Thus, CWT model's input dimension is  $80 \times 2500 \times 6$ .

The choice of these three inputs is motivated by their widespread use in prior literature, as summarized in Table 3.1. Time series, STFT, and CWT representations are among the most commonly employed inputs in seismic signal classification, offering a fair and meaningful basis for comparison. Moreover, in practice, when experts manually label seismic events, they mainly observe time series and STFT representations. Each input type has its advantages and disadvantages: time series preserve the complete temporal structure of the signal but may require the model to learn frequency-related features implicitly; STFT provides joint time-frequency information with fixed resolution but may lose fine details at different scales; CWT offers high resolution analysis that captures both low- and high-frequency components effectively but comes with higher computational cost and redundancy.

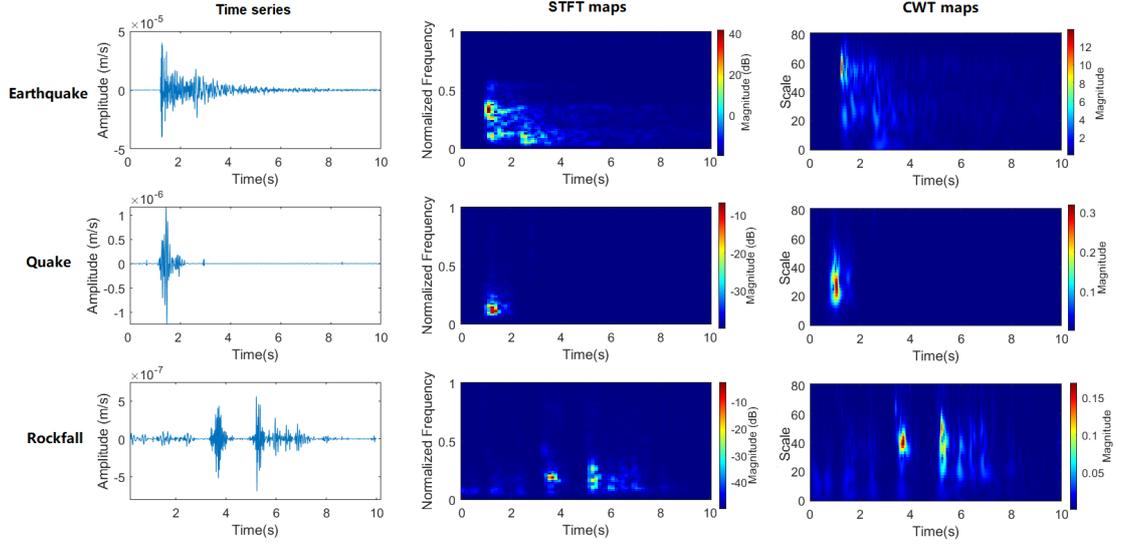


Figure 3.1: Examples of seismic signals for 3 classes (earthquake, quake and rockfall) with 3 different input formats (temporal waveform, STFT and CWT maps) in Résif dataset.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

### 3.3.2 Neural Network Architecture

The architectures of the proposed three networks, one for each type of signal input, are inspired by VGG16 which mentioned in Section 3.2 and adapted to the sampling rate, the size of the input seismic signal and its feature map. These are deep networks composed of convolutional layers (for feature representation and extraction), max pooling layers (for downsampling the extracted features to obtain the feature map of a small size) and FC layers (for classification). That is, after multiple convolutional layers, where the number of convolutional kernels increases with the layer number, and max-pooling layers, the input signals are compressed into small feature maps, that are then classified through 3 FC layers. The output layers have 4-node softmax classifiers, providing the probability distribution of 4 classes of events. The three architectures are shown in Figure 3.2.

For the raw time-series model, the input dimension is  $2500 \times 6$ , corresponding to 10-

second waveform segments sampled at 250 Hz across 6 channels. All convolutions and pooling operations are performed in 1D. The first convolutional layer contains 64 filters with kernel size 9. Subsequent convolutional layers double the number of filters up to 512 in the final layer. The max-pooling filter size is set to 4. These parameters were selected to balance temporal resolution preservation and computational efficiency, while ensuring that the receptive field grows sufficiently deep to capture temporal features over the 10-second window. Importantly, for the time-series model, the activation functions in the first two convolutional layers are set to linear instead of ReLU. This is to prevent the premature suppression of information, since waveform data contains both positive and negative values, and early ReLU activation may cause many neurons to be zeroed out, impairing learning effectiveness.

For the STFT and CWT models, the architectural design was adjusted to maintain comparable model complexity and capacity to the time-series case, facilitating a fair comparison of performance across input types. The input to the STFT model is a  $64 \times 75 \times 6$  tensor, representing spectrogram slices per channel. All convolutional layers here are 2D, with  $3 \times 3$  kernels, which is a standard choice in image-based CNNs for capturing local spatial and frequency correlations. In the CWT model, the input map is highly anisotropic ( $80 \times 2500$ ), with one dimension (time) being much longer. To manage this, stride values of  $1 \times 2$  are used in select layers to compress the time axis more aggressively, preventing overly deep architectures. Additionally, some convolutional kernels are enlarged to  $3 \times 9$  to allow a broader temporal field of view, which is beneficial for capturing long-duration waveform patterns unique to certain seismic events.

Many deployed monitoring network configurations contain less than 6 channels. For such systems, we design single-channel models that take one channel at a time as inputs. The parameters of the single-channel model are the same as in the multi-channel model except that the input shape is changed. To classify single-channel data, the input size is  $2500 \times 1$  for time series,  $65 \times 75 \times 1$  for STFT maps and  $80 \times 2500 \times 1$  for CWT maps. When these single-channel models are used with  $n$ -channel data, it will output  $n$  softmax vectors for each event. Then, to make a decision, we calculate the mean of

each class for these  $n$  vectors.

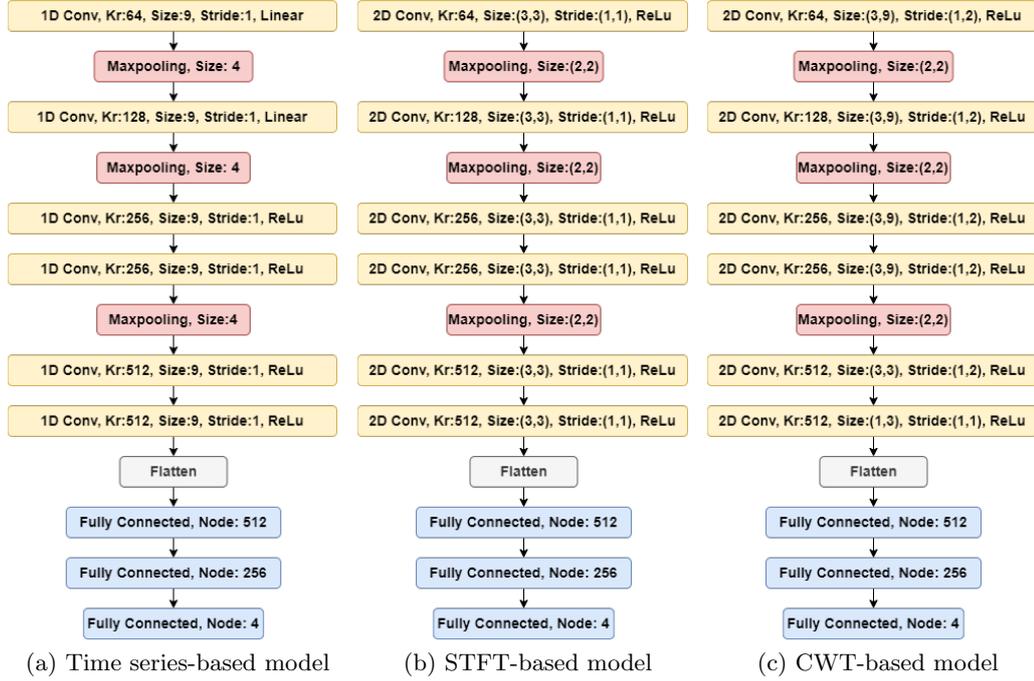


Figure 3.2: The network structure for three different inputs.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

### 3.3.3 Data and Training Strategy

The dataset used to train and test all the models is openly accessible from the Résif Seismological Data Portal, acquired by the French Landslide Observatory OMIV (Observatoire Multi-disciplinaire des Instabilités de Versants) [59]. To validate our classifiers, first we use only labelled events, that is, we removed all sections in the dataset that were not catalogued. We note that the catalogue includes natural/anthropogenic noise segments, hence the classifiers are trained to distinguish this type of noise as well as the other three micro-seismic events. We split the dataset of labelled events into training (60%), validation (10%), and testing (30%) sets according to the time of the event (earliest to latest). To increase learning efficiency, we standardise the dataset by subtracting the mean and dividing by the standard deviation after denoising. Since

the original dataset is very unbalanced (see Table 2.1), the training set was balanced to avoid training bias, by generating new events by shifting the seismic events in the window as well as adding background noise from non-catalogued eventless waveforms to the catalogued events. This results in approximately 600 training samples per class, ensuring a more uniform class distribution during training.

While alternative techniques such as loss function reweighting (e.g., assigning higher weights to minority classes in the cross-entropy loss) or resampling strategies could also address class imbalance, we chose data augmentation because it not only balances class frequencies but also improves the diversity and robustness of the training data. In particular, shifting the target events within the sliding window not only simulates realistic waveform variation but also enhances the model’s ability to detect events in continuous data streams, where event boundaries are not fixed. This makes our approach especially suitable for real-time applications involving streaming seismic signals. Importantly, the test set remains unbalanced and untouched to reflect the true event distribution and maintain the realism of performance evaluation.

### 3.3.4 Sliding window-based detection

To illustrate the applicability of our previously trained time-series model on continuous data, we evaluate it on an unseen period (25-28 Nov 2014) from the Résif dataset, during which 18 quakes, 23 earthquakes, and 65 rockfalls were recorded. Since the input of the model is a time window, and the output is a 4-class probability vector, we slide the input window on the continuous data to achieve a continuous series of probabilities, as in [89], [30], [90]. This was necessary as the CNN was designed to process a fixed-length signal, and a continuous stream of signals is not of a fixed length. The sliding window method involves dividing the input signal into smaller windows of fixed length, with a fixed overlap between consecutive windows. The input window is set, as before, to 10 sec. The input time window was slid with 90% overlap, that is, the classification result is output each second. We set the decision threshold to 0.7: that is, a class probability greater than 0.7 will be considered as an event class of this time window.

### 3.4 Results and discussion

#### 3.4.1 Classification Results

All networks were implemented using the Keras framework and trained for 100 epochs. In preliminary experiments, training curves indicated that the validation loss typically stabilized around the 80th epoch. The Adam optimizer was used due to its effectiveness in training deep networks with sparse gradients and its adaptive learning rate capabilities. The cross-entropy loss function was chosen as the task involves multi-class classification with four target categories (earthquake, quake, rockfall, and noise). Cross-entropy is a standard choice for such tasks, as it directly measures the divergence between the predicted probability distribution and the true labels. The initial learning rate was set to 0.0007, a conservative value to ensure stable gradient descent. To promote convergence and prevent the model from getting stuck in shallow local minima, the learning rate was reduced by 10% every five epochs. This step decay strategy helps the model learn quickly in the early stages of training, while gradually refining the weights as it approaches convergence, improving final performance and generalization.

Table 3.2: The classification performance results for the three six-channel models. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

The classification performance results for the three models are shown and compared in Table 3.2. Each model was trained and tested 5 times using the same training and testing sets to ensure repeatability. The results are presented in the form of “mean±standard deviation” where the mean and standard deviation are calculated using the results obtained after 5 trainings and tests.

It can be seen from the table, that all three models provide highly accurate and similar classification performance, with the F1-score ranging from 0.85 for quakes to

Table 3.3: The classification performance results for the three single-channel models. © 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

0.98 for earthquakes. The CWT model has marginally better classification accuracy (averaged over all classes) of 90.82%, followed closely by the STFT model with 90.62%, and the time series model with 89.97%. However, the relatively lower overhead of the time-series model over the others makes it more desirable in practice.

The results using the proposed single-channel models when only one channel is available are shown in Table 3.3. Comparing with the multi-channel models, the single-channel models performs slightly worse for all classes of events. This can be explained by the fact that the channels with very poor SNR affect the output due to averaging, which is not the case with the multi-channel model.

We compare our results to traditional machine learning methods in [26], where an RF classifier is used on a subset of randomly chosen events from the same dataset, and correctly classified 94% of earthquakes (vs. 98% with the proposed method in Table 3.2), 94% of rockfalls (vs. 92% in Table 3.2), 92% of noise (vs. 86% in Table 3.2) and 93% of quakes (vs. 89% in Table 3.2).

The corresponding confusion matrices are shown in Table 3.4, for time waveform, STFT, and CWT models, respectively. From the confusion matrices, we can see that few quake events are misclassified, due to much shorter duration (<5seconds), smaller SNR and the fact that quakes are usually localised events and hence not detected by all channels, which is not the case with earthquakes. Quake classification results are better for STFT and CWT than time series inputs, with fewer misclassified events, since CWT and STFT features take into account both frequency and time duration. However, the recall (sensitivity) of quake is still slightly inferior to the 93% reported by [26].

We think there are three reasons for this. First, Provost et al. [26] use 71 constructed features including nine network geometry attributes (such as mean and std correlation lag between the stations; stations with max/min amplitudes, etc.), that assume knowledge of positions of the deployed sensors. After removing these features, the performance of [26] was reduced to 90% in average over all classes, ranging between 86% and 94%. Secondly, the number of labelled quake events is much smaller than the number of other events. For example, there are 401 labeled rockfalls and only 234 quakes, and this has negative impact on the generalisation ability of the networks. Thirdly, in [26] the channel with the highest SNR is chosen for feature extraction. In our models, all 6 channels' data were inputted, and some channels with low SNR may impact classification results.

Table 3.4: The confusion matrix for the three six-channel models.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

### 3.4.2 Interpretation of Misclassified Events

The misclassified events for the three multi-class models, time series input, STFT, and CWT, accounted for 12.65%, 11.14% and 11.12%, respectively, of the total number of events. As can be seen from the confusion matrices, the main cause of a drop in recall is quake events being misclassified as noise, which is not surprising given their low SNRs. Therefore, a close examination was conducted on cases where quake events were misclassified as noise, totaling 10 events across all three multi-class models. Out of these 10 misclassified quake events, 4 events were misclassified by 2 or 3 models. We focus on these events and show a representative example for the time-series model in Figure 3.3. From the figure, it can be seen the main reason for the wrong classification is that the energy of the seismic signal is too weak compared to background noise.

Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

Indeed, in Figure 3.3 (left column), the wave peak of the correctly classified quake event reaches roughly  $1 \times 10^{-6}$ m/s while the wave peak of the missed quake event (right column) is only around  $8 \times 10^{-8}$ m/s, that is, two magnitudes lower.

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

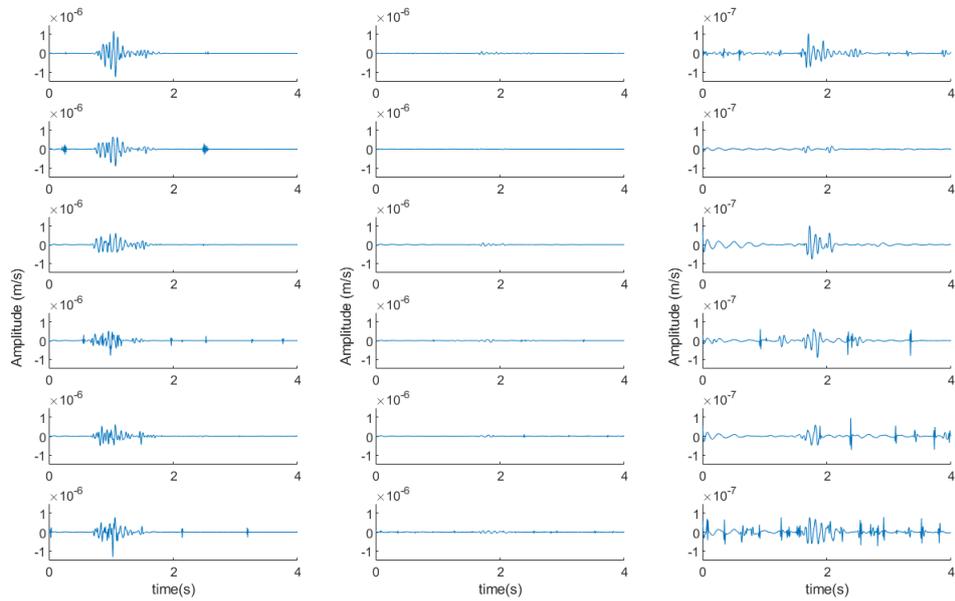


Figure 3.3: Waveforms of two quake events in all six channels. The quake event which was correctly detected and classified (first column) and the quake event which was misclassified as noise (second column) with the same ordinate axis ranges from  $-1.5 \times 10^{-6}$ m/s to  $1.5 \times 10^{-6}$ m/s. The misclassified quake event with the scaled zoomed (third column), its ordinate axis ranges from  $-1.5 \times 10^{-7}$ m/s to  $1.5 \times 10^{-7}$ m/s.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

### 3.4.3 Comparison with state of the art DL-based methods

In addition to benchmarking against traditional machine learning methods [26], we compare the performance against the state-of-the-art CNN-based ‘deepquake’ network [1] that was shown to outperform similar DL architectures [1]. The ‘deepquake’ network has two models, one for time series input and another for STFT map, namely ‘arch-time’ and ‘arch-spect’, respectively. The network uses 20 seconds windows with 100Hz sampling frequency as input. Thus, we first down-sample our data to 100Hz (from 250Hz), use the same proposed pre-processing method to normalise data and extract 20 seconds of event waveforms as input window. The ‘deepquake’ model classifies inputs into three classes: earthquakes, other events and noise. Thus, we re-label both quakes and rockfalls as ‘other events’. We use ‘deepquake’ pre-trained models as initial weights of each layers. Then, we re-train these two models using the Résif dataset for an additional 80 epochs. The results are shown in Table 3.5. From the confusion matrix, it can be seen that ‘arch-spect’ has better performance than ‘arch-time’ on Résif data, which is expected and in accordance with our results, i.e., the STFT map input model outperforms the time series input model.

Comparing these results with Table 3.4, we can see that our time series-based CNN model outperforms ‘arch-time’ for all classes. Indeed, 104 earthquake events are correctly classified by ‘arch-time’ while 112 earthquake were correctly classified by our time series-based CNN. 22 other events (rockfalls and quakes) are misclassified as noise by ‘arch-time’ while 6 rockfalls and 7 quakes (13 in total) are misclassified as noise by our model. For STFT maps as input, we can see that ‘arch-spect’ performs worse than our STFT-based CNN. Both models correctly classify 113 earthquakes. 14 other events are misclassified as noise by ‘arch-spect’ while 7 rockfalls and 4 quakes (11 in total) are misclassified as noise by our STFT-based model. This shows that our proposed model is in line with the state-of-the-art, with the advantage of additionally distinguishing endogenous landslide seismicity, including rockfalls and quakes.

Table 3.5: The confusion matrix for the two 'deepquake' models [1].

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

### 3.4.4 Continuous detection results

For results we test our network on continuous stream, time series-based CNN correctly detected and classified 91% of earthquakes, 83% of quakes and 94% of rockfalls (TP). These results are similar to the time series-based model's recall (sensitivity) in Tables 3.2 and 3.4, and therefore we conclude that the network is robust to continuous detection and classification. In Figure 3.4, we show 2 and a half minutes (i.e., 150 sec) of the continuous waveform of an earthquake event that occurred at 4:05:39 on November 28, 2014. It can be seen from the figure that the model correctly detects the start of the signal. In addition, our network detected many other events that have not been catalogued, namely, 174 earthquakes, 260 quakes and 32 rockfalls.

It is important to note that performing continuous detection and classification using a sliding window over continuous data streams introduces a much higher proportion of background noise compared to testing on curated datasets. Due to the good sensitivity of the CNN, a small portion of the background noise may be misclassified as events (earthquakes, quakes and rockfalls). Although the probability of such misclassification is low, the overwhelming amount of noise in continuous data inevitably leads to a number of false positives. This phenomenon causes the precision of seismic event detection to decrease, which in turn reduces the overall F1 score. A detailed evaluation of model performance on continuous stream, including precision, recall, and F1 score under these conditions, will be presented and discussed in Chapter 4.

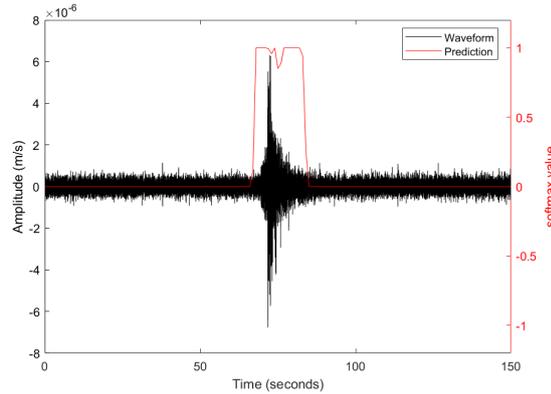


Figure 3.4: Continuous detection results for an earthquake event.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

### 3.4.5 Transferability analysis

Using the same sliding window of continuous time series recordings as above, we evaluate transferability of our trained time series input model. The models described above, trained on the Résif dataset, are used to detect events on a microseismic dataset from Larissa region in Greece [91], [88]. There are inherent differences between the two sites. The Résif dataset was collected from the Sauze catchment basin in the Alps, characterized by limestone formations and black marl [60]. The region around Larissa in Greece is a seismically active area, characterized by gneiss and schists [92]. The two sites are geologically different, with different monitoring networks: different number of sensors and deployment geometries, sampling at different rates (250Hz Résif vs. 100Hz Larissa). Such differences, especially the different geological background, alter the characteristics (amplitude, frequency content) of the signals, therefore making the transferability problem very challenging.

To transfer the model from the 6-channel Résif dataset, we selected six channels from the Larissa dataset from HT network as the array: all three components (North, East, vertical) of station TYR1, vertical (Z) components of TYR3, TYR6 and TYRN [88]. The choice of these particular stations, referred in the following as Array 1 (A1), was

based on their location and quality of recordings. The four chosen stations form an almost equilateral triangular array, with three stations at the three vertices and one inside the triangle - see Fig. 3.5. This is a commonly used geometry in microseismic monitoring surveys to maximise detection of microseismic events.

To assess the sensitivity of the results to sensor array deployment geometry, we perform the same analysis as above on another set of stations forming a more random geometry, referred to as Array 2 (A2): all 3 components of TYR1, Z-components of TYR2, TYR3 and TYR4. Compared to the previous array geometry, the sensors in this array are located further apart - see Fig. 3.5 - and hence, waveforms from the same event are likely to look different in the recordings of the different stations.

Since the sampling rate of two datasets is different, we up-sampled the data from Larissa using linear interpolation to 250Hz and denoised the signal as described in Subsection 3.3.1. The dataset from Larissa contains 86 catalogued earthquakes during the entire day of 17th March 2021, from 0:00 to 24:00. The catalogue we used can be downloaded from the Institute of Geodynamics, National Observatory of Athens [91] using as search area a circle with radius of 218km and centre at longitude 22.1777 degrees and latitude 39.6460 degrees. Our proposed models correctly classified **all** catalogued earthquakes, demonstrating good transferability.

Additional (to the aforementioned earthquake catalogued events) earthquake, quake and rockfall events were classified via our proposed approach. These findings are verified via manual detection, as well as the commercial InSite software v3.15 developed by ITASCA, which is widely used for seismic event detection. The software uses energy-based triggering algorithms to identify potential seismic events by detecting sudden increases in signal amplitude [93]. For manual verification, experienced analysts examined the multichannel time series waveforms and corresponding spectrograms to identify and validate seismic events. We randomly chose one hour within 17th March 2021, from 18:53 to 19:53 and manually searched the recordings of the same six channels to detect seismic events via visual observation. The data were filtered using a 5-100Hz band-pass filter (as per Subsection 3.3.1). An event is valid if it was visually observed on at least 2 channels (at different stations). Note that manual detection of the events

### Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

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

An additional catalogue of all events detected and classified via the proposed CNN, manual detection and InSite software is published and publicly available in [94], and also provided in Appendix A for further analysis by the research community. This catalogue, which we refer to as comprehensive catalogue from now, also includes the type or class of event and its the duration. A1 and A2 refer to the events classified by the proposed CNN multi-classifier for the first (equilateral triangular) array and second array configurations, respectively. Similarly, M1 and M2 refer to events identified via manual detection on array 1 and 2, respectively, whilst I1 and I2 denote events identified by the InSite software.

The proposed CNN multi-classification model is designed for waveform pattern recognition (not P-wave picking as in [77]), therefore it does not estimate the start time and duration precisely because the model locates the event on a subset of the 6 input channels using a 10-sec window. After manual verification of traces, we observed that there are a number of occasions where the CNN model estimated multiple adjacent earthquakes as either one event or vice versa. For distant events from the station locations, the different signal phases (e.g., P wave, S wave etc.) arrive with a distinct time difference. This results in later phases, e.g., P wave reflections, being detected by the CNN model as separate earthquake events, instead of a single event. To mitigate this effect, we perform processing as follows. In the network’s last FC layer, we set a bias towards classifying segments as non-events. In particular, we set as a decision threshold softmax value of 0.7 (instead of the default value of 0.5), which means that only when the softmax value greater than 0.7 is reached, the candidate window will be classified as an event. Next, as a post-processing step, we merge all events that start within 5-sec time interval, into one to prevent classifying different wave reflections into multiple events. Furthermore, for events that originated far away from the monitoring

### Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

stations, the time interval between P-phase and S-phase will be large ( $>10$  sec), and hence in this case, an earthquake event will appear as two separate events.

Referring to the proposed comprehensive catalogue, we compare events detected between time-series based on (1) CNN model with post-processing, (2) manual event detection and (3) automatic detection using the InSite software. It can be seen that the results of the proposed model and manual detection are very aligned. Indeed, only 8 earthquake events are detected manually by M1 and missed by A1, 4 of which were detected by A2. Similarly, 13 earthquakes detected by M2 are missed by A2, but 6 of these events were picked up by A1. These were more distant or more localised events, respectively. Each array configuration detected 62 earthquakes. 15 events were detected by A1 and missed by A2, or vice versa. InSite detected 32 and 38 of these earthquakes for A1 and A2, respectively, 22 of which are common for all 6 detection methods (2 arrays with manual, automatic detection and InSite). 9 (11) and 21 (8) rockfall (quake) events were detected by A1 and A2, respectively. Out of these, 7 and 4 rockfall events were detected by I1 and I2, respectively, 2 of which are common for all 6 detection methods.

– May 30, 2025 –



Figure 3.5: Six selected monitoring stations in Larissa, Greece. A1 configuration, based on an almost equilateral triangular array, comprises TRY1, TRY3, TRY6, and TRYN. A2 configuration, random geometry, is based on stations TRY1, TRY2, TRY3, and TRY4 that are further apart.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

Tables 3.6 and 3.7 summarise the 3 types of events detected by the proposed approach (A1 and A2), that were manually corroborated (referred to as TP), missed by the proposed approach (False Negative (*FN*)) and not confirmed manually (False Positive (*FP*)). We observe that no quake and rockfall events detected by the proposed CNN were missed. Fewer earthquakes were missed by A1 than A2, since the A2 configuration includes stations that are spread further apart. Most earthquake events are detected by both A1 and A2 configurations, with some events of distant origin picked by A2 only, and localised events by A1 only. Details of the 3 distinct events and their time of occurrence, that were detected manually, via Insite Software and proposed CNN are provided in the comprehensive catalogue<sup>1</sup>.

Table 3.6: Larissa results from proposed CNN on one hour data using readings from TYR 1, 3, 6, and N stations (A1): Comparison between automatic and manual detection.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

### 3.4.6 Internal workings of proposed architecture via feature maps vizualization

We visualize the feature maps at the output of different convolutional layers in our proposed network designs and visualize the features at the output of the second FC layer (which is the input to the output layer). The feature maps are the result of applying the filters to the input of convolution layers. Visualizing feature maps can be used to explain which input features are extracted in convolutional layers and analyse

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

Table 3.7: Larissa results from proposed CNN on one hour data using readings from TYR 1, 2, 3 and 4 stations (A2): Comparison between automatic and manual detection.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

the influence of time, frequency, wavelet domain representation on the interior of the network. A visualization example is shown in Figure 3.6. The same earthquake event forms an input to all three proposed CNNs. The figure shows the first 9 feature maps output by the first, the second, and the fourth convolutional layer after max-pooling, for each of the three proposed models. For time series, we can see the input signal being transformed into many earthquake-like signals by the first convolution layer. Then in the second convolution layer, the feature map extracts more detailed features from the input signals. The key features in earthquake signals (i.e., P-waves and S-waves) are preserved and enhanced in the fourth convolutional layer, where we can see clearly extracted peaks of events to be classified. In the STFT network, the first convolutional layer highlights the frequency band and the time step where the event occurred. From the second convolutional layer, it can be seen that the STFT-based model has learned the frequency features of the event, as the feature maps highlight the high-frequency and low-frequency range of the event. The fourth convolutional layer transforms these feature maps of different highlighted frequencies into more abstract representations. The CWT-based model works similarly to the STFT-based model. In the first convolutional layer, the feature maps highlight where, in time and scale, the event occurs. Compared with the feature maps of the first layer, the feature maps of the second layer have larger highlighted areas of the event, as the model is extracting

more detailed wavelet features. Then, the fourth convolutional layers transform the features into abstract representations. In summary, through the visualization of feature maps, we explain the internal process of CNN in classifying seismic signals. The first convolutional layer is often used to find the position of the event in the input window. Then, CNN starts to extract more detailed features of the event. With the deepening of the convolutional layers, the extracted features are gradually transformed into abstract representations for subsequent classification. After the feature maps output by the convolutional layers are flattened, these one-dimensional features are classified by the FC layers. There are 256 nodes in the second FC, so there are 256 features. Figure 3.7 shows the features extracted by the second FC layer (which is the input of the last layer) for the four classes by three different models. For earthquake, the max value of the feature tends to be large (greater than 20), while quake and rockfall events have relatively small max features (between 10 and 20), and the max features of noise are the smallest (less than 10). As deep learning is a black-box algorithm, it is difficult to determine what attributes of the event these 252 features represent. However, we can see that different events have different feature distributions, and seismic events (earthquakes, quakes and rockfalls) have larger feature values than noise. This means that the CNN can extract different features for the seismic event waveforms, but does not extract many features from noise signals.

### 3.4.7 Complexity Analysis

Table 3.8 presents the execution time required to process a 10-second input window with 6 channels (equivalent to 15,000 samples), covering denoising, transformation, and testing for each CNN model. CNN models were designed and tested using Python 3 and the Keras framework. The denoising and transform steps are also programmed in Python 3. All experiments were performed on an i5-10310U CPU. The time series-based model has the lowest complexity requiring 24ms to output result for a 10-sec window, followed by the STFT-based (37ms), and CWT-based model (808ms). This is expected since the time series- and STFT-based models have 5,684,036 trainable parameters, while the CWT-based model has 6,035,012 parameters. Note that the

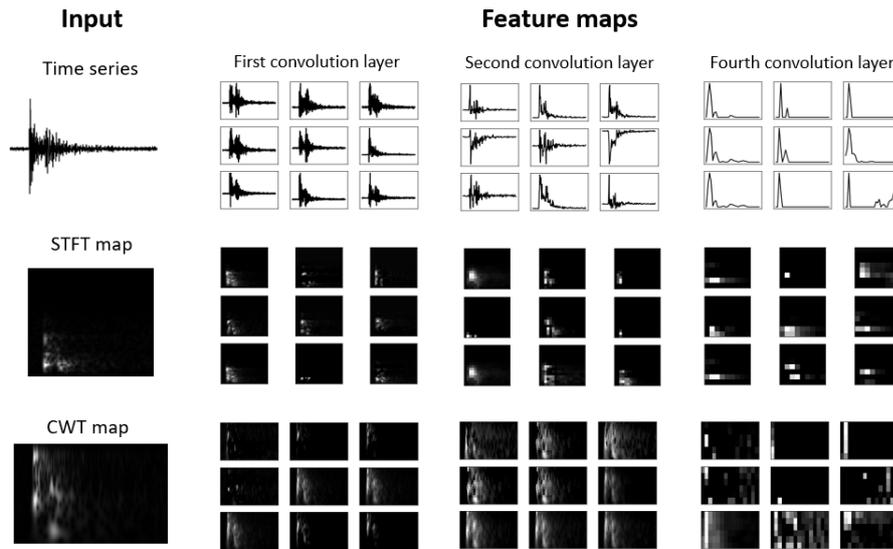


Figure 3.6: Feature maps of the first, second, and fourth convolutional layer of the three CNNs with seismic events as input.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

– May 30, 2025 –

time series-based and STFT-based models require much less than the 1sec window time shift applied on continuous data traces, and hence these two models can process the data in real time.

### 3.5 Summary

This chapter proposes microseismic classification on continuous recordings (no additional detection step needed) via a CNN, exploiting the inherent feature engineering ability of deep learning. Three CNN models were developed for three types of data inputs: temporal waveform, STFT and CWT maps. These proposed models were trained on the labelled Résif dataset, in order to detect/classify three types of events, namely earthquakes, quakes and rockfalls. During testing on an unseen portion of the Résif dataset, the time-series-, STFT- and CWT-based models all had similar performance for the three microseismic events and anthropogenic noise classes. Additionally, the time series-based model was observed to be the fastest during complexity analysis,

Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

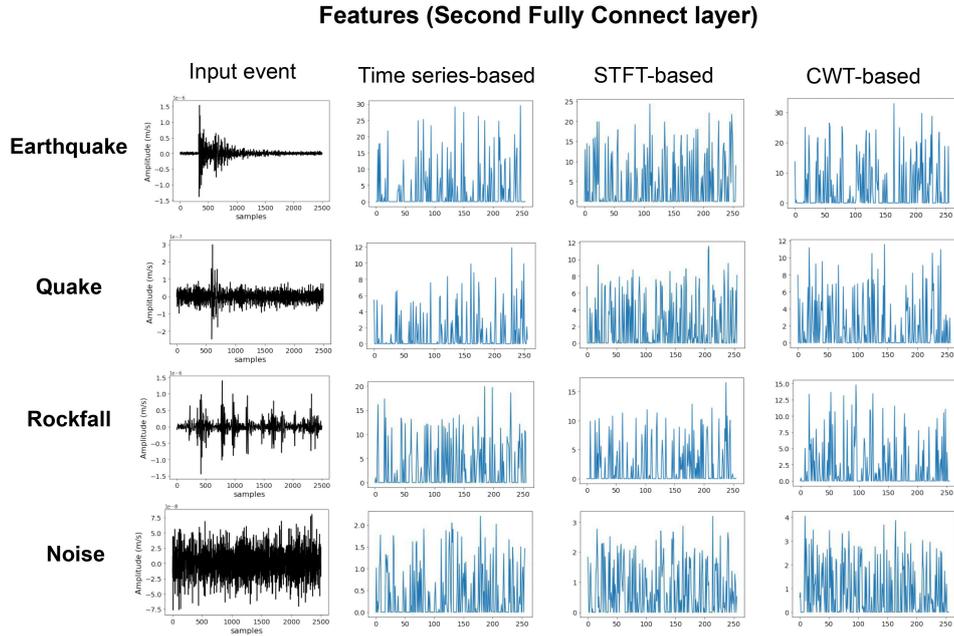


Figure 3.7: Feature maps of the second fully connected layer of the three CNNs for four classes.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

Table 3.8: The execution time for each model in ms.

© 2023 IEEE. Reprinted, with permission, from Jiang et al., "Microseismic Event Classification With Time-, Frequency-, and Wavelet-Domain Convolutional Neural Networks," \*IEEE Transactions on Geoscience and Remote Sensing\*, vol. 61, pp. 1–14, 2023.

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

### Chapter 3. Microseismic event classification with time, frequency and wavelet-domain Convolutional Neural Networks

demonstrating near real-time performance. The ability of the proposed pre-trained (on the Résif dataset) model to classify events from continuous recordings in a geologically distinct site was demonstrated via transferability to a 24 hours dataset from the region of Larissa, Greece. All 86 catalogued earthquakes made available to us during that time period were correctly detected and classified despite major differences in monitoring layout used at the two sites and geological terrain. Using a less favourable deployment geometry, results were still within a satisfactory range. The following chapter will address the 'black box' problem in CNN models by utilizing advanced XAI tools to enhance the interpretability of the proposed CNN model. It will demonstrate how deep learning classifies seismic events and provide insights into why the model may occasionally misclassify them.

## Chapter 4

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

### 4.1 Introduction

Deep learning models achieve state-of-the-art performance in detecting and classifying seismic signals avoiding cumbersome manual feature generation, selection and extraction process, with their ability to automatically learn most discriminative features from raw recordings as discussed in Chapter 3. However, this also means that these models are limited by the used training set, and may learn specifically spurious correlations with the prediction target [95], [96]. Furthermore, the fact that the feature engineering task is taken away from the designer, makes deep learning models opaque, and hence often referred to as “black box”, which limits their use. Indeed, geoscientists are still reluctant to use them and rather rely on less complex interpretable methods based

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

on hand-crafted features [19] that ensure that relevant physical features are used for detection and classification (see, e.g., Table I in [19] and Table A1 from [97]).

Explainable AI (XAI) [98], [99] is a research direction that provides human-interpretable explanations that can potentially enhance training process, correct manual data annotation, improve models, and contribute towards building trust in AI-generated outputs [100], [101]. XAI tools have been extensively used in computer vision (e.g., [102]) and time-series signal analysis problems (e.g., [103]); however, the work on explaining the output of deep learning models for seismic signal analysis, and using these explanations to improve confidence in data labelling, model training and building trust in inferred outputs, is still in its infancy.

This chapter provides comprehensive explanations to identify key features learnt by a deep neural network for multi-class classification, and demonstrates that these features are in agreement with the physical properties of seismic signals and common hand-crafted features used in the literature [19]. The generated explanations are then used to explain instances of misclassifications and correct errors in manual labelling, jointly with a geoscientist, who verified the corrected labels of the classified events and the features associated with these events. This builds trust in the models confirming that the learnt feature representations agree with expert knowledge.

We use state-of-the-art XAI tools to explain deep learning models for detection and classification of micro-seismic signals and show how these explanations can be used to improve the designs and explain correct and wrong predictions. In particular, we use CNN-based architectures for detection and classification of seismic signals into four classes: earthquake, micro-earthquake referred to as quake, rockfall and noise. These are the same classes as used in Chapter 3.

#### 4.1.1 Pertinence of Trustworthy AI to seismic analysis

In order to pave the way towards a regulatory framework for ensuring trust in AI, the European Commission has published seven principles of Trustworthy AI [104], which include Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination and Fairness, Societal

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model and Environmental Well-Being and Accountability.

Depending on how the AI-based seismic analysis will be used, from understanding the subsurface processes and mechanics to hazard and disaster management, the AI systems can be seen as minimal risk to high risk, and therefore subject to strict oversight before they can be used to ensure infrastructure and human safety. Therefore, the following principles are of relevance to seismic analysis. First, AI systems should empower decision makers when it comes to hazard assessment or infrastructure planning, allowing them to make informed decisions from the AI system outputs. The principle of Human Agency and Oversight caters for proper oversight mechanisms that need to be ensured, which can be achieved through human-on-the-loop and human-in-command approaches. Second, the principle of technical robustness and safety, in part states that AI systems need to be accurate, reliable and reproducible to ensure unintentional harm can be minimised and prevented. Accuracy refers to the ability to correct predictions based on AI models and can be implemented via rigorous evaluation and indication of likelihood of potential errors. Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions. A reliable AI system is one that works properly with a range of inputs and in a range of situations. Third, the principle of privacy and data governance enables users to trust the data gathering process and that it does not contain inaccuracies, errors or mistakes, especially with respect to labelling or cataloguing by expert geoscientists. Fourth, the principle of transparency states that the data and AI system should be transparent through traceability mechanisms in the form of documentation of datasets and processes that yielded in decision, including data gathering, data labelling and algorithms used. Furthermore, transparency also includes explainability, that is, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. This includes XAI. Fifth, transparency also states that humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations. Finally, the social and environmental well-being principle state that the AI systems should be sustainable and environmentally friendly - this can be through taking into considering the resource usage and energy consumption for

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

training the models. Moreover, they should consider the societal impact. Monitoring, understanding, modelling and predicting landslide processes due to climate change, especially rainfall, tackle United Nations (UN) Sustainable Development Goal (SDG) 13 on Climate Action [105]. As explained in [106], shearing and friction between the soil grains results in release of seismic energy within the landslide body. Therefore, passive seismic monitoring is a good approach to monitor and mitigate slope instabilities, as it provides high temporal resolution data in near real time that relate to the dynamics of the landslide. This means that the transition (and rapid transformation) of the landslide from slow rate sliding into a rapid slope failure may be detected and therefore mitigate associated hazards.

## 4.2 Related work

To ensure trust and expert’s control of the decision process, machine learning-based seismic signal analysis has been performed either in a semi-automated manner [107] using continuous expert oversight and monitoring (human-on-the-loop), using interpretable models [19], or using non-interpretable models (such as Random Forests) but with numerous hand-crafted features [26] to ensure that the inference is made on signal characteristics identified by experts as important. In [97] a detailed study of feature importance is presented where 119 features are constructed based on seismic signal literature and their importance tested using four different feature importance methods and different classifiers based on Support Vector Machine, Random Forest, and three graph signal processing based semi-supervised approaches. The features are experimentally ranked showing time-, frequency-, cepstrum and polarity features that are of highest importance in inference making per studied class. The results show that out of 119 constructed features only a subset contributed significantly to the decision. Note that this study was based on quantifying the importance of hand-crafted features in accurately classifying multiple event classes from continuous data, thus deep learning networks were not considered.

In [1], CNNs are used to classify isolated catalogued seismic events into noise, earthquake and other events. The authors developed a heatmap-based visualisation tool to

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

explain model outputs via the outputs of activation functions of each filter in the convolutional layers and then overlapping the result with the raw input signal. However, this study has several weaknesses when it comes to gaining trust in model outputs. Firstly, it is not clear how explanations are formed by fusing outputs of the activation functions from different layers. Secondly, only binary classification is considered, i.e., identifying relatively well-defined earthquakes from other signals. Thirdly, the approach does not exploit advanced XAI methods, and it is not used to explain any false predictions.

In [108], the authors proposed a Dual-Channel CNN Module where one channel contains raw time-domain waveforms, and the other channel contains frequency-domain information by Discrete Cosine Transform (DCT) to classify input seismic waveform into rock fracturing and noise, together with an explanation module, EUG-CAM (Elaborate Upsampling-based Gradient-weighted Class Activation Mapping). It builds upon the principles of the gradient weighted class activation mapping (GradCAM) [109], harnessing the influence of feature map values and gradients to elucidate the importance of diverse features in the last convolutional layer. Recognizing the discrepancy between feature map sizes and input data dimensions, EUG-CAM uses a strategic amalgamation of transposed convolution, unpooling, and interpolation, to generate feature mappings from a coarse localization map. This results in an explanation feature map that effectively encapsulates class activation, learning insights, and network architecture considerations. However, the model’s limitation is in classifying only two classes (rock fracturing vs. noise) and its confinement to binary classification. Furthermore, the reliance on a 1-D CNN model facilitates explanations primarily within the time domain, possibly neglecting the benefits of frequency-domain insights garnered from the DCT. Additionally, the visualization maps cannot show the adverse input signal influence (negative contribution) on classification results, hampering a comprehensive and well-rounded comprehension of the model’s decision-making process.

### 4.3 Data gathering and context

The dataset we used is Résif which has been introduced in Chapter 2. In this study, we used data from the 3C sensor. This choice aligns with common practices in seismic

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

waveform classification, where a 3-channel input is standard, such as EQ-transformer [77] and DeepQuake [1]. Additionally, it facilitates transfer learning, as many seismometers employ three-component sensors, ensuring compatibility with various seismic datasets and applications. Using 3 channels also reduces the number of false positives which can occur with arrival mismatches and reduces the computational demand.

### 4.3.1 Labelling

The number of labels for each class in the Résif catalogue, referred to as the original catalogue, is detailed in Chapter 2 Section 2.3. Events were classified from continuous recordings using both manually designed feature-based classifiers and deep-learning-based classifiers with automated feature extraction, as discussed in Chapter 3. Since detection and classification were performed on a continuous data stream, additional events not present in the original catalogue were identified by both the Normalised Graph Laplacian Regularisation (normGLR)-based classifier [97] and the CNN-based classifier. These methods detected and classified hundreds of previously uncatalogued events, particularly between November 25th and 28th, 2014—a period characterized by heightened activity on the SZ slope [110].

As reported in [97], all four types of events are present in this 4-day time period, and in addition to the 120 events (65 rockfalls, 18 quakes, 23 earthquakes and 14 noise) labelled in the original catalogue, 17 quakes, 89 earthquakes and 92 rockfalls events were detected and classified by the normGLR classifier whereas an additional 260 quakes, 174 earthquakes and 32 rockfalls were detected and classified with the CNN approach proposed in Chapter 3. These algorithms only missed 1 earthquake, 1 rockfall and 2 noise events that were present in the original catalogue.

All events detected by the normGLR classifier, the CNN classifier and an additional classifier based on Siamese networks [111] were reviewed by an expert for labelling following the methodology used to build the original catalogue, which is based on the seismic signal waveform and spectrogram features. The final outcome of the expert reviews for this 4-day period were 69 quakes, 29 earthquakes and 126 rockfalls. Note that the normGLR classifier was too sensitive, overestimating the number of earth-

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

quakes [97]. The CNN-based 6-channel input multi-classifier in Chapter 3 was too sensitive for quakes and earthquakes but missed a number of rockfalls.

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

## 4.4 Methodology

### 4.4.1 Proposed CNN-based architecture

In this chapter, an STFT-based CNN model adapted from Chapter 3 utilising STFT maps as inputs, is used. These frequency-domain inputs have been shown to provide better results on average compared to directly feeding time-series signals. Additionally, human experts traditionally detect and classify seismic events based on their spectral characteristics, reinforcing our choice of the STFT model. This spectral analysis aligns well with expert practices, making STFT a natural fit for our human-on-the-loop objectives.

The architecture of the model is composed of convolutional layers, max pooling layers and FC layers, adapted to the input shapes and output categories, as shown in Figure 4.1. Convolutional layers perform feature representation and extraction, followed by max-pooling layers that downsample the extracted feature into a feature map with smaller size.

Compared to the model in Chapter 3, several modifications have been made to improve the model’s robustness and suitability for long-duration seismic event detection in continuous data streams. Specifically, the input window size has been extended from 10 to 15 seconds. Additionally, the number of convolutional kernels and neurons in each layer has been reduced. These adjustments are motivated by findings in Chapter 3, which reveal that CNN models tend to exhibit excessive sensitivity. Through extensive

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

experimentation, we found that reducing certain parameters helps to mitigate this sensitivity, lowers model complexity, and achieves a better balance between performance and complexity. The detailed network configuration is presented in Figure 4.1, while the classification performance of the revised model is discussed in Section 4.5.1. Moreover, recognising the prevalence of waveforms captured by three-component sensors, the input to the network is 3-channel input data, in contrast to 6-channel used in Chapter 3, which significantly expands the model’s applicability across a wider range of scenarios.



Figure 4.1: STFT-based CNN for seismic classification. Kr denotes the number of kernels, and ‘Flatten’ function transforms the input data into a 1D array. © 2024 IEEE. Reprinted, with permission, from J. Jiang, V. Stankovic, L. Stankovic, D. Murray, and S. Pytharouli, ”Explainable AI for Transparent Seismic Signal Classification,” IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024.

#### 4.4.2 Training and testing strategy

The inputs to the model for both training and testing comprise STFT maps generated from the raw recordings as discussed in the previous subsection. Chapter 3 demonstrate

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

that CNN models tend to be overly sensitive. To address this, we refine the sensitivity of our CNN by only using the high-quality events to train the model. Specifically, we visually inspected and chose events from the original catalogue to ensure that the set used for training comprised only high-quality events based on signal clarity and high-SNR for earthquake, quake and rockfall classes. This selection process was conducted by two experts with experience in seismic data analysis, who manually reviewed both the multichannel time series waveforms and their corresponding spectrograms. The number of events in each class is shown in Table 4.1. All noise events originate from the original catalogue. In addition to the manually selected events, we utilise the labelled events from the 25th November 2014 (one day) to train the model further. This additional data allows us to augment the training set with events that are not included in the high-quality subset of the original catalogue and help to improve precision and recall.

Table 4.1: The number of labeled events chosen from the original Résif catalogue

Class	Total No. events
Earthquake	340
Quake	207
Rockfall	378
Noise	351

#### 4.4.3 Continuous detection and post-processing

Raw signals recorded by 3-channel (North, East and vertical direction) seismic recorders are used. Since the classes of interest are 5-60Hz bandwidth, we first use a BandPass Filtering (BPF) to remove low frequency noise (denoising) as in Chapter 3. To allow prediction on a continuous stream of signals, a sliding window method as discussed in Chapter 3 is used to segment the continuous stream into smaller windows.

In this study, a window size of 3750 samples (i.e., 15 seconds) is used. The overlap between consecutive windows is set to 93% of window size (3500 samples (14 seconds)), which corresponds to a shift by 1 sec, allowing the CNN model to capture the temporal dynamics of the events of interest. Furthermore, since the peak amplitude of signals belonging to different classes is large, to improve the learning ability of the models, we perform normalization of the filtered recordings. In particular, in order to enable the

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

model to focus on classifying the input signals and facilitate the subsequent explanation of the classification results, we normalise each 15-second window by subtracting mean and dividing by the maximum of the absolute value of each input window. For the STFT map input, in order to get good time and frequency resolution, ‘Boxcar’ window with length of 128 samples with 70% overlap is used. We perform STFT on denoised and normalized time series input window. Thus, the input shape for the STFT-based model is  $65 \times 95 \times 3$  samples.

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

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

us to choose the class of the event with the longest duration, as it indicates a higher confidence level in the classification result.

#### 4.4.4 Explainability-informed re-labelling

Unlike classifiers such as RF, SVM and (norm)GLR-based classifiers that take hand-crafted features as inputs and where feature importance was studied in detail in [97], the CNN multi-classifier is essentially a “black box” since we do not know what features were deemed important. We therefore utilise LRP to understand feature importance for the deep-learning CNN multi-classifier.

LRP [112] is a state-of-the-art XAI method, that shows the contribution of each sample in the input data to the classification results and can be implemented in the pre-trained model [113]. In this chapter, LRP is used to help identify which parts of the seismic signal are most important in making the final classification decision. This helps understanding which features of the seismic signal are most relevant for seismic detection, and identify any potential biases in the model. In addition, LRP can provide interpretable and detailed explanations of the model’s decision-making process, which can be useful for communicating the model’s results to human experts.

The LRP method starts from the output of the model, sets the output value before activation function as relevance, and gradually back propagates the relevance, iteratively, layer by layer, to the input nodes. In the backpropagation, the relevance follows the conservation law, that is, a neuron’s relevance equals to the sum of the relevance it flows out toward all other neurons. Various propagation rules have been proposed, such as LRP- $\gamma$ , LRP- $\epsilon$ , LRP-0 rule [101]. In this paper, we used LRP- $\epsilon$  rule which is suitable for convolutional layers and max pooling layers [114], and is defined as:

$$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_{0,j} a_j w_{jk}} R_k, \quad (4.1)$$

where  $R_j$  represents the relevance score assigned to neuron  $j$ ,  $a_j$  denotes an input activation,  $w_{jk}$  is the weight connecting neuron  $j$  to neuron  $k$  in the layer above,  $\sum_{0,j}$  denotes that we sum over all neurons  $j$  in the lower layer plus a bias term  $w_{0k}$  with

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

$a_0 = 1$ .  $\epsilon$  is a regularisation term, i.e., a small value that prevents the denominator from being 0.

In our implementation, The LRP algorithm was implemented using the iNNvestigate toolbox, default parameters of the LRP- $\epsilon$  rule are used. All other parameters, such as the layer-wise activations and weights, were taken directly from the pre-trained CNN model without modification.

In practical expert usage, LRP maps are generated for all instances where the CNN’s predicted class differs from the expert-provided label (as outlined in Subsection 4.3.1). The post-processing techniques described earlier play an important role in making the CNN’s outputs on continuous streams smoother, thus enabling experts to more easily observe the model’s predicted class transitions during continuous detection. The LRP maps are presented alongside the corresponding STFT spectrograms and raw time-series signals. This multi-modal visualization enables experts to assess whether the CNN’s attention (i.e., relevance) focuses on physically meaningful seismic features (e.g., P-wave or S-wave arrivals, spectral characteristics of tremor). If the highlighted regions align with seismological expectations, it may suggest that the CNN correctly identified the event despite an incorrect label, prompting expert re-evaluation. Thus, LRP maps act as a bridge between deep learning outputs and seismological expertise, promoting transparency, improving model validation, and supporting the generation of more accurate labeled datasets. If, upon review, the expert determines that the original label was incorrect, the event is re-labelled accordingly. These corrected labels, together with their STFT spectrograms and LRP maps, are made publicly available in [115]. The overall expert-driven label verification process is illustrated in Figure 4.2.

## 4.5 Results and discussion

### 4.5.1 Analysis of classifier output

Network models are implemented in Keras framework. Since the activation function of the output layer is softmax, we use categorical cross entropy as loss function. The used optimiser is Adam with an initial learning rate of 0.0007. Adaptive learning rate

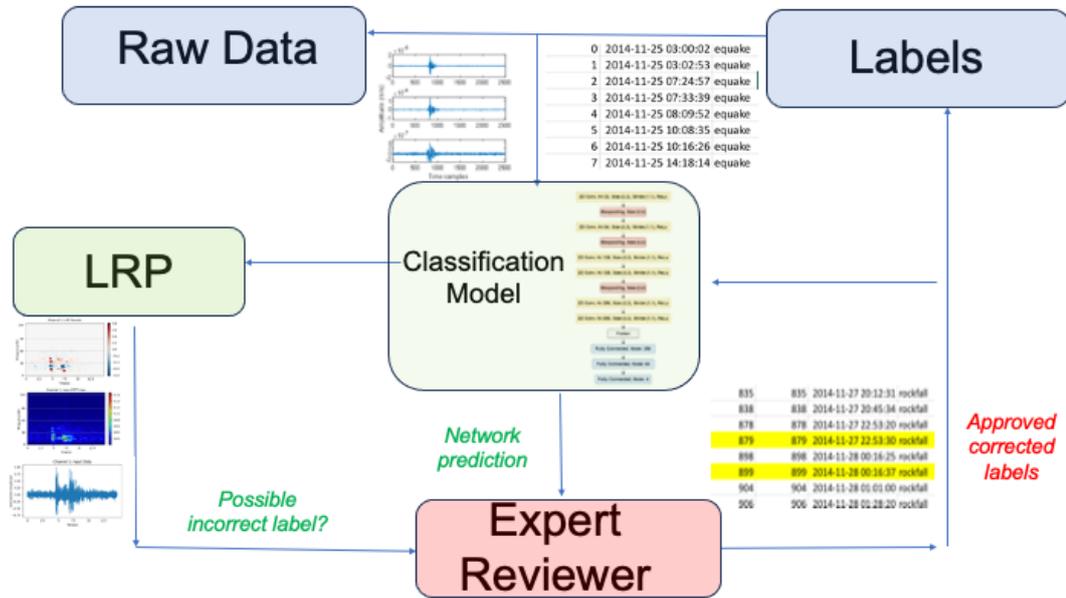


Figure 4.2: Flowchart of the proposed human-on-the-loop process.

adjustment is implemented, which reduces the learning rate by a factor of 0.9 when loss improvements plateau for 5 epochs. Training is performed over 100 epochs with a batch size of 128. For the second training session, utilizing the data from November 25, the model is trained over a total of 50 epochs. To prevent the risk of overfitting due to additional training, early stopping is implemented; that is, if the training accuracy did not exhibit significant improvement within 5 consecutive epochs, the training process is terminated early.

In the 3-day testing period (26th-28th Nov.), the expert labelled 46 quakes, 18 earthquakes, 74 rockfalls and 719 noise events. The confusion matrix in Table 4.2 compares the output of the proposed CNN-based network, with post-processing (Sec. 4.4.3), to the expert labels. As is common practice for seismic signal classification on continuous data [26], the confusion matrix also includes recall or sensitivity values in brackets. Recall is the ratio of true positives to the sum of true positives and false negatives. In Section 4.3.1, it is demonstrated that during the 4-day period from November 25th to 28th, there are 6 additional earthquakes not labelled in the original catalogue [26]. The model discussed in Chapter 3 detected a much larger number, specifically 174 additional, earthquakes. This comparison shows the significant improvement in the

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

precision of earthquake classification achieved by our model. Additionally, our model achieved high recall (sensitivity) for rockfall events. As expected, quake and noise events can be confused with the other 3 classes, due to heterogeneity of the noise signal and very low signal amplitude of quake signals.

Table 4.2: Confusion Matrix - Proposed CNN-based network with post-processing against expert labels (the numbers in brackets indicate recall rates).

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

### 4.5.2 Explainability

Figure 4.3(a) shows an example of a correctly classified earthquake event. Positive and negative values of the LRP relevance represent positive and negative contributions to the classification results, of the corresponding STFT, respectively. The distribution of relevance is focused on the high frequencies (about 40 to 50Hz) when the P-wave is picked as well as the low frequencies (around 15 to 20Hz) of the P-wave and, after roughly 5sec, the low frequencies of the S-wave with intermediate noise shown in light blue correctly identified as not contributing (negative contribution). This example shows that the model learnt, and uses as basis for its predictions, that the P-waves of earthquake events tend to have both high and low frequencies (around 50Hz and 20Hz, respectively) and that high energy content of S-Waves follows in time.

Figure 4.3(b) shows an example of a correctly classified quake event. Quake events are of shorter duration than earthquakes, have lower amplitudes, and energy focused in low frequencies. The relevance is concentrated in the single peak (positive and negative) of the event waveform, suggesting that the normalised maximum amplitude is the key distinguishing feature. In the frequency domain, the LRP map clearly shows the importance of the peak that has energy mainly focused below 30Hz while there is also a small positive contribution between 30 to 40Hz.

Figure 4.3(c) shows an example of a correctly classified rockfall event. While the

relevance of quake events is concentrated on a single peak, the relevance of rockfall events is concentrated on multiple peaks, which also shows an important property of rockfall events – multiple significant peaks. Looking at the LRP map, the relevance has multiple focused points corresponding to multiple short waves – a characteristic of rockfalls. In addition, although both, the rockfall and the quake events have a frequency band between 10 to 30Hz, the relevance are mostly concentrated at frequencies greater than 20Hz for rockfalls and below 20Hz for quakes.

Similar visualisation maps are produced for other correctly classified events. In summary, the model searches: (a) for P-wave and S-wave peaks and their corresponding frequency contributions to predict an earthquake; (b) a short wave with a single peak below 20Hz to decide quake; (c) multiple significant frequency components around 25Hz to decide that the target signal is rockfall. This is in accordance to the characteristics of the three signal classes [26], [19], [94]. Next, we will analyse misclassified events to explain why they occur and how they can be avoided.

### 4.5.3 Explaining origin of misclassification

In this section, the way LRP can be used for model diagnosis is shown. The confusion matrix presented in Table 4.2 shows that the quake signals are sometimes misclassified as rockfalls. Interestingly, however, rockfall signals are rarely misclassified as quakes (only 2 misclassified events). To investigate this further, Figure 4.4(a) shows an example of a quake event misclassified as rockfall. In the LRP map, the relevance distribution is very scattered. That is, the LRP relevance is not focused on the quake event’s peak, but instead picked up several consecutive peaks, where the positive relevance is correctly concentrated at 5 seconds. This indicates that the model correctly recognised a quake event’s peak appearing around 5 seconds, but there was a high energy signal in nearby frequency bands, influencing the final prediction. On the other hand, there are many positive relevancies at different times that correspond to frequencies between 20Hz to 30Hz, which is akin to the learnt rockfall ‘behaviour’. Thus, the main reason of misclassification between quake and rockfall is that the signal-to-noise ratio of the quake event was very low, with a noise signal appearing immediately after, mimicking

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

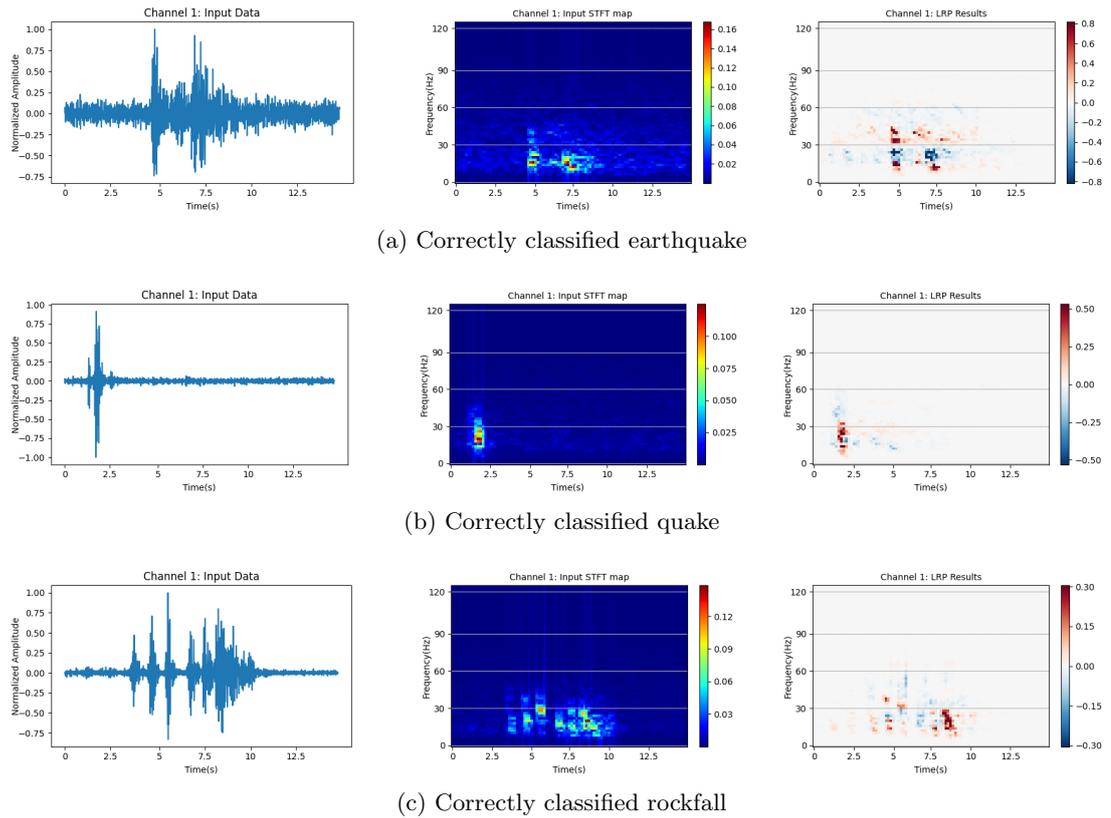


Figure 4.3: Correctly classified examples of earthquake, quake and rockfall: The first column shows the time-series signal, middle column the STFT, and the right column is the LRP relevance heatmap.

multiple peaks of rockfall events.

In Figure 4.4(b), we show an instance in which a rockfall event is misclassified as a quake. The rockfall event displays multiple peaks; however, these peaks, aside from the principal one, are of low magnitude and the event has a very short time span. Analysis of the LRP representation illustrates a concentration of positive effects (depicted in red) at the primary peak of the event. Conversely, numerous negative contributions (depicted in blue) are observed at the secondary peaks, suggesting that the presence of these multiple peaks is not taken into account due to their limited magnitudes; hence, the model finally classifies this event as a quake.

In Figure 4.4(c), we present an instance of a quake misclassified as an earthquake. This misclassification is evident in the LRP map, where both high-frequency and low-frequency components simultaneously exhibit positive contributions around the 3-second period. Thus, the model interprets this segment as a P-wave. Furthermore, at approximately 5 seconds into the waveform, a positive contribution appears in the low-frequency range. Although the primary peak of this event occurs around 3 seconds, the spectrogram reveals that the low-frequency component persists for an extended duration. Moreover, the event is influenced by higher-frequency noise (exceeding 30Hz), and this high-frequency noise coincides with the primary waveform peak around the 3 seconds. Consequently, this led the model to mistakenly identify it as a P-wave, with the prolonged low-frequency component being mistakenly identify as a S-wave. These observations align with seismic features of earthquakes, thereby causing the model's misclassification as an earthquake event.

In Figure 4.4(d), we encounter an instance where an earthquake is mistakenly classified as a rockfall. The LRP map highlights multiple spectral peaks, which is a feature of rockfall events. However, this event may have resulted from an earthquake occurring amidst background noise, exhibiting a distinctive multi-peak pattern. Thus, despite the presence of a P-wave at approximately 1 second and an S-wave at roughly 4 seconds, complex background noise caused misclassification.

In Figure 4.4(e), the misclassification of noise as an earthquake is shown. The noise signal exhibits prominent peaks around 4 seconds and 5.5 seconds. Examination

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

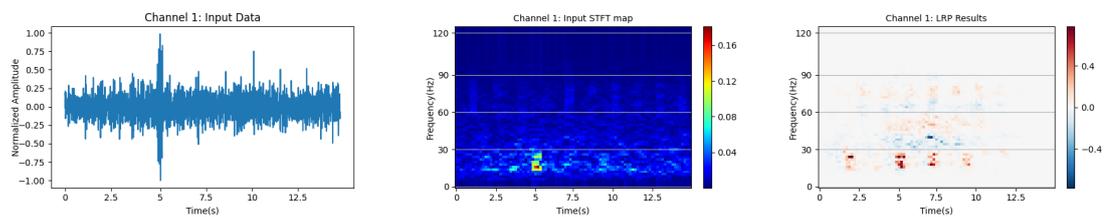
of the LRP map reveals the model’s recognition of low-frequency and high-frequency components (15-20Hz) around the 4-second mark, along with low-frequency signals at 5.5 seconds (15Hz). This aligns with the characteristic features of P-waves and S-waves in earthquake signals, resulting in the model’s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

It can be seen from these examples that most misclassifications are due to high level of background noise. The next example highlights another origin of error related to the filtering process. Figure 4.5 displays an unfiltered earthquake waveform with a frequency below 3 Hz, characteristic of low-frequency earthquakes that are rarely associated with active landslides [116]. Since the focus is on detecting local seismic events related to landslides, we apply a BPF in the 5-60 Hz range (see Sec. 3.3.1), which excludes these low-frequency earthquakes. Consequently, this filter removed the low-frequency event’s waveform, leaving only background noise as input to the CNN. As illustrated in Figure 4.6, the LRP map indicates that the model failed to extract meaningful features from the filtered input, resulting in the earthquake being misclassified as noise. This misclassification can be attributed to the rarity and uniqueness of low-frequency earthquakes on landslides, as the filter inadvertently eliminated their distinctive waveforms, confounding the CNN’s classification process.

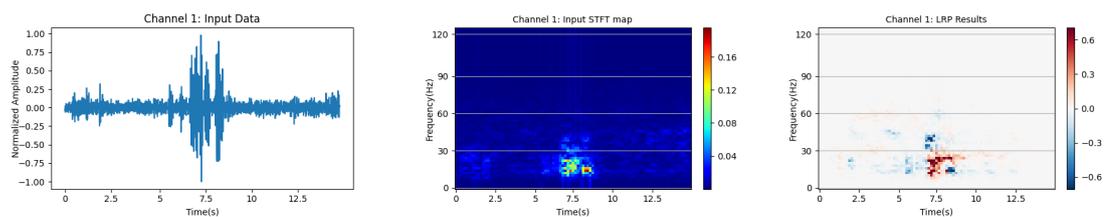
#### 4.5.4 Re-labelling results

Figure 4.7 shows three examples of misclassifications, which could be due to human error during expert labelling. The example shown in Figure 4.7(a), is an event classified by the model as noise, though the domain experts labelled it as a quake. In the STFT representation of the signal, no obvious peak corresponding to the event was discernible. Moreover, the LRP map exhibits a disordered distribution of relevance. Collectively, these findings lead to the argument that the event in question is more likely to be anthropogenic noise rather than a quake. Figure 4.7(b) illustrates a similar situation where the event is mistakenly labelled as an earthquake. There are no clear P-waves at

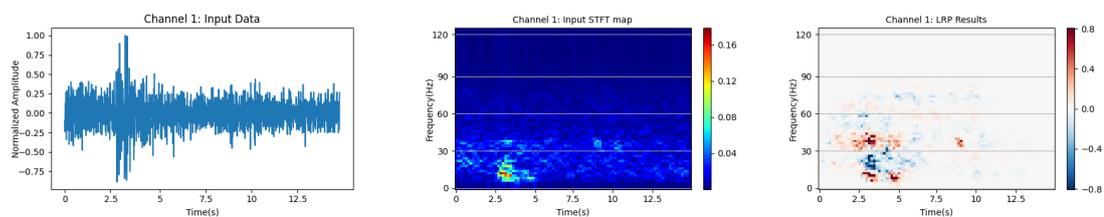
Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model



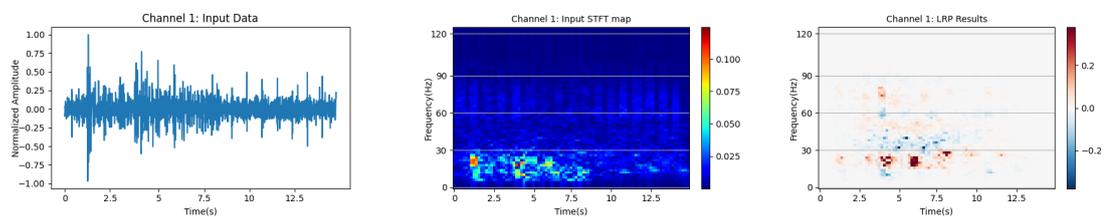
(a) Quake misclassified as rockfall



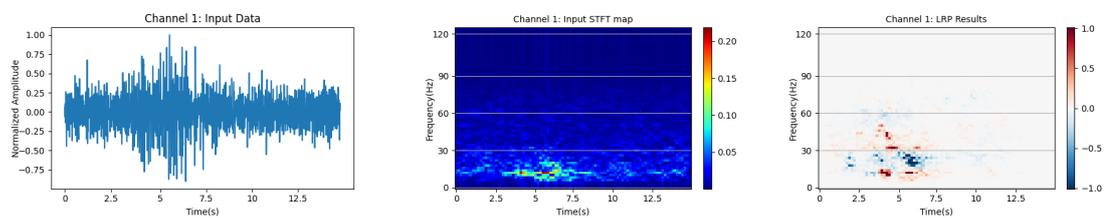
(b) Rockfall misclassified as quake



(c) Quake misclassified as earthquake



(d) Earthquake misclassified as rockfall



(e) Noise misclassified as earthquake

Figure 4.4: Misclassified examples.

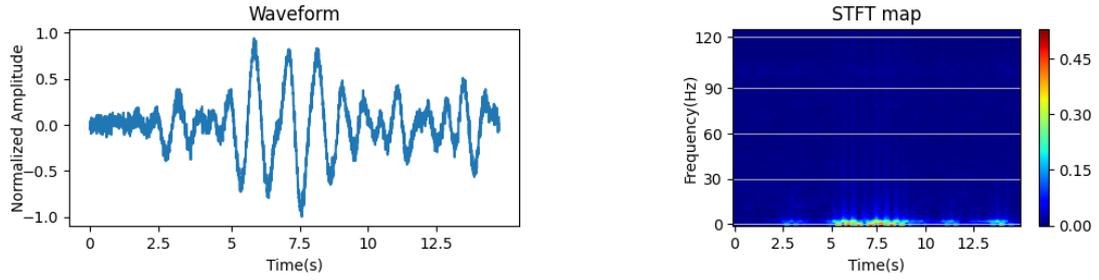


Figure 4.5: Waveform (left) and STFT map (right) of the unfiltered low-frequency earthquake.

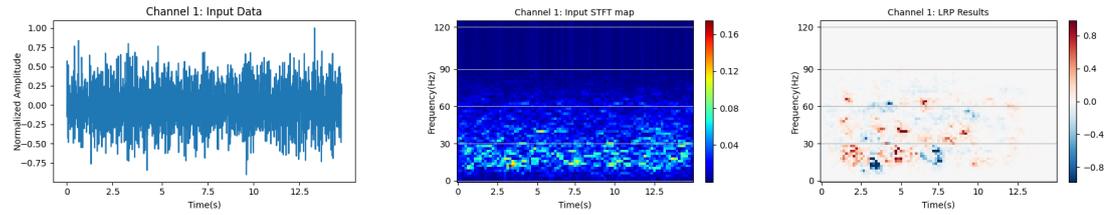
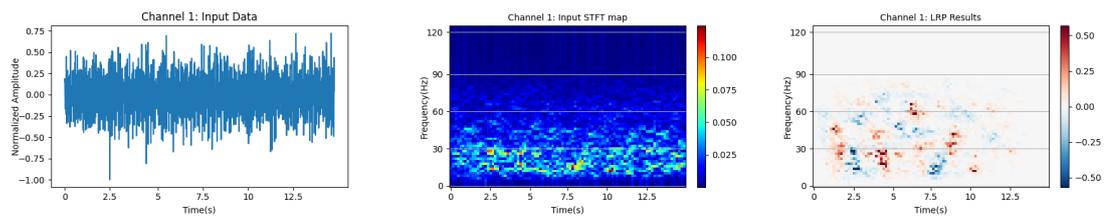


Figure 4.6: Waveform (left), STFT map (middle) and the LRP map (right) of the filtered low-frequency earthquake.

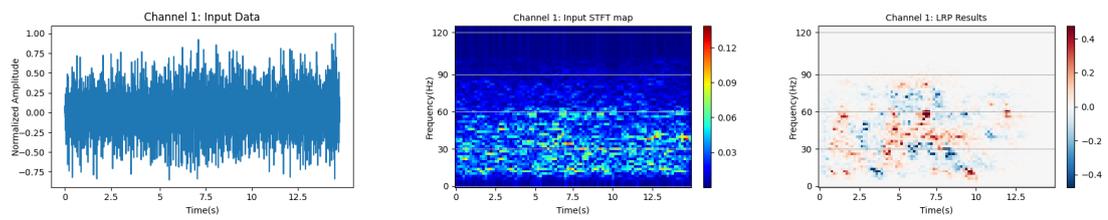
both low and high frequencies, and there are no S-waves with high energy content. For this earthquake event, we also examined the unfiltered raw signal, and it still did not exhibit any earthquake waveform characteristics. Figure 4.7(c) shows an example that was classified as a rockfall by the CNN model, while the expert labelled it as a seismic quake. It can be concluded from the LRP map that the model focused on multiple peaks in the event, with a frequency distribution centred around 30Hz, characteristics that align with typical rockfall patterns. In contrast, quakes tend to exhibit a single dominant peak, a feature that was notably absent in the input STFT map, where multiple peaks were discernible. Consequently, based on these distinctive patterns and spectral features, it becomes evident that the event in question is more accurately classified as a rockfall.

Here we list all corrections made to the expert catalogue, following above explainability and queries. Specifically, 7 quakes were relabelled as noise as per example Figure 4.7(a), 1 earthquake was relabelled as noise (shown in Figure 4.7(b)), and 1 quake as rockfall (Figure 4.7(c)). In addition, some noise events were labelled by the expert though these events occurred very close to earthquake, quake and rockfall events,

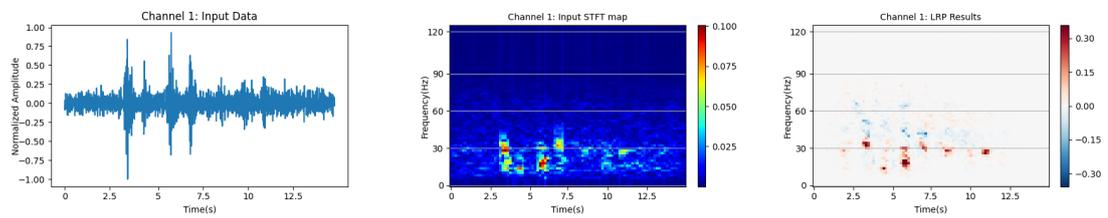
Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model



(a) Noise mistakenly labelled as quake



(b) Noise mistakenly labelled as earthquake



(c) Rockfall mistakenly labelled as quake

Figure 4.7: Three examples of events with labels corrected.

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model

which potentially caused confusion. Hence, we removed all noise events that occurred in close proximity (within 30s) to the earthquake, quake and rockfall events - this way 38 noise events were removed.

Thus, after this relabelling there are 38 quakes, 17 earthquakes, 75 rockfalls and 689 anthropogenic noise events in total. The verified catalogue of events is publicly available in [115], as a contribution to address the second and third principles of Trustworthy AI, related to reproducibility and data access. Specifically, the 260 verified events on the 25th Nov. 2015 are listed in the Training events identified by the date. The 819 verified events on 26th to 28th Nov. 2014 are listed in the Additional 3-day catalogue. In order for other researchers to enable benchmarking, Table 4.3 and Table 4.4 show the confusion matrix and classification performance after the re-labelling, respectively. Although the F1-score for quake events is low, we have a high Recall but precision is low because of 8 instances of false positives for rockfall. There are relatively few instances of quake and earthquake, which explains why the F1-score is not the best indicator of performance and the confusion matrix provides a more explainable and trustworthy measure of performance.

Table 4.3: The confusion matrix after label correction. The numbers in the brackets show the recall.

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

Table 4.4: The classification performance after label correction.

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

## 4.6 Summary

This chapter reflects on the relevance and application of the seven principles of Trustworthy AI: Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity and Fairness, Societal and Environmental Well-Being, and Accountability. These principles are particularly important in the context of seismic signal analysis, where AI systems are used for tasks ranging from understanding subsurface processes to managing geological hazards. Depending on the use case, AI systems in this domain may be considered minimal to high risk and therefore must adhere to stricter requirements to ensure safe and ethical deployment.

To align with these principles, we propose a human-on-the-loop seismic classification framework that incorporates explainable AI techniques, specifically Layer-wise Relevance Propagation (LRP). This approach enhances transparency, strengthens technical robustness, and supports expert oversight during the classification process. LRP is used to highlight which parts of the input signals contribute most to the model's decisions, helping to identify potential human labeling errors. For example, LRP maps have revealed cases where quake events are misclassified as rockfalls due to high-frequency noise that resembles rockfall signatures. These visualizations assist experts in diagnosing such misclassifications and refining event labels accordingly. The use of LRP supports informed and accurate decision-making by experts, thereby reinforcing the principle of Human Agency. Through this collaboration between human expertise and AI-assisted interpretation, the accuracy of microseismic event catalogues is improved, contributing to more reliable geological assessments. The system also addresses data governance by facilitating quality control during data labeling, which is essential for maintaining trust in the dataset and downstream AI applications.

However, training deep learning models typically requires a large amount of labeled data, but labeling seismic data is time-consuming and inefficient, making labeled seismic events costly. This challenge limits the advancement of deep learning-based seismic event classification. In the following chapter, SSL technology is introduced to reduce the model's dependency on extensive labeled data by assigning unsupervised

Chapter 4. A human-on-the-loop approach for labelling seismic recordings from landslide site via a multi-class deep-learning based classification model and semi-supervised downstream tasks to address varying levels of annotation needs. Additionally, a novel algorithm that significantly enhances the efficiency of manual annotation by experts is proposed.

## Chapter 5

# Generative Self-Supervised Learning for Seismic Event Classification

### 5.1 Introduction

Deep learning-based models achieve state-of-the-art performance in detecting and classifying seismic signals, surpassing traditional machine learning methods that require manual feature generation, selection, and extraction. Their key advantage is the ability to automatically learn the most relevant features from raw recordings. However, achieving high accuracy in seismic classification through deep learning often relies on supervised learning techniques, which necessitate large labeled datasets. The labeling process requires professional expertise and is inherently time-consuming, costly and prone to human error; thus, gathering sufficient data for supervised inference can be prohibitively difficult. Hence, there is a need for methods that can detect and process continuous seismic signals without or with minimum labelled data.

Self-supervised learning (SSL) is a paradigm in machine learning where a model is trained on a task using the data itself to generate supervisory signals, rather than relying on external labels provided by humans. This makes SSL approaches time- and cost-effective, hence their popularity in fields such as computer vision and natural lan-

guage processing that require large amounts of labeled data to train models [117] [118].

This chapter proposes a microseismic classification method based on SSL. The proposed approach enables the model to achieve good classification performance for landslide seismic signals while using no or minimal labeled data. SSL is used for representation learning, to extract features from raw waveform recordings without relying on labels. The learnt features are then employed in three downstream signal classification tasks. Firstly, using K-means clustering, we classify the SSL-generated features into four classes: earthquake, micro-quake, rockfall, and noise signal, providing this way a fully unsupervised learning-based classification method. Secondly, we utilize a small portion of the labeled data (5%, 10%, 20% and 33.3%) and the corresponding extracted features to train an Artificial Neural Network (ANN), thereby proposing a semi-supervised learning approach. Our models are trained on the publicly accessible Résif dataset, which contains 1375 labeled (micro)seismic events. We use all event waveforms for SSL without labels and a small portion of the labels for the semi-supervised downstream tasks. Thirdly, leveraging the superior feature extraction capability of SSL, we employ SOM with various shapes ( $8 \times 8$ ,  $12 \times 12$ ,  $17 \times 17$ , and  $21 \times 21$ ) to cluster the SSL-based extracted features into multiple nodes. The key motivation for using SOM its ability to cluster high-dimensional SSL features into topologically ordered nodes. Each node in the SOM can be treated as a representative unit encompassing similar seismic events. Instead of labeling individual events, experts can label each SOM node once, and this label can then be propagated to all events mapped to that node. For example, using an  $8 \times 8$  SOM reduces the manual labeling task to just 64 nodes, which collectively represent the entire dataset. This significantly alleviates the labeling burden and can reduce the required human effort by 95.4%.

## 5.2 Related work

SSL can be generative-, contrastive-, or adversarial-based [119]. Generative-based SSL trains models to generate data, similar to the input, learning the underlying data distribution, with autoencoding being an example. Contrastive-based SSL trains models

to distinguish between similar and dissimilar data instances, bringing similar instances closer in the representation space (e.g., the SimCLR model [120]). Adversarial-based SSL involves a generator creating data indistinguishable from real data and a discriminator distinguishing between real and generated data, with Generative Adversarial Network (GAN) exemplifying this approach.

Murshed et al. [121] propose a semi-supervised Seismic Contrastive Graph Neural Network (SC-GNN) for an earthquake early warning system, that uses a GNN to propagate spatio-temporal information through a graph representing seismic station distribution and wave propagation. A contrastive-based SSL approach is used to train the network with larger time windows, enabling predictions using shorter initial waveforms, where similarity in terms of seismic intensity is defined, i.e., waveforms from the same seismic event are considered similar. The SC-GNN provides accurate seismic intensity predictions using only the initial seismic waveforms from a limited number of seismic stations. In [122], Li et al. introduce a generative-based Self-Supervised Convolutional Clustering Picking (SCCP) method for automatically picking the first break of microseismic recordings. The SCCP method decomposes and reconstructs the time-frequency features of the microseismic recordings using accurate convolutional encoding and decoding under self-supervision. The output is then clustered by Fuzzy C-means.

Using seismic geophone data from the unstable Åknes rock slope in Norway, Lee et al. [123] develop seismic event waveform classification to distinguish 8 different signal classes, namely noise, regional earthquake, rockfall, slopequake high-frequency, slopequake low-frequency, multi slopequake, tremor and spike. There are 1818 labelled events and 1611 unlabelled events. The authors use Neighboring and Random Crop for input data augmentation, and employ a contrastive-based SSL method, Variance Neighboring-Invariance better-Covariance Regularization (VNIbCReg) model [124], to extract features. After pre-training using unlabelled data, the authors evaluate performance using linear and fine-tuning methods as downstream tasks. In the linear evaluation protocol, only the encoder from the SSL-trained model was kept with frozen weights and an additional linear layer, trained using 80% of the labelled data. For the

fine-tuning evaluation, the authors again kept only the encoder from the SSL-trained model, but this time the encoder’s weights were left trainable, with fine-tuning using 5%, 10%, and 80% of the labelled data. The linear evaluation is to assess how well the learned features perform in classification tasks by SSL. The fine-tuning evaluation is particularly effective at revealing the generalizability of a SSL method. Note that [123] uses 80% of the labeled data for training in linear evaluation, without exploring the impact of using less training data, and does not explain the learning process, which could be done through the visualization of features extracted by the encoder. This limits the comprehensiveness of the analysis regarding the model’s feature extraction capabilities.

Song et al. [125] pre-train the contrastive-based Simple Framework for Contrastive Learning of Representations (SimCLR) model [120] using a large amount of unlabeled data. Then, a linear classifier is added to the encoder part (with weights frozen), leveraging small labeled datasets for linear classification. The extracted features are classified into microseismic and blasting events. A Data Balancing Algorithm is used to enlarge the labeled dataset from 3,966 to 7,932 samples. The augmented data is then split, with 60% used to train the downstream classifier and 20% used for performance evaluation and testing. The paper uses only a fixed proportion of labeled data for linear evaluation. In [126], Mousavi et al. propose a generative-based SSL method, using an auto-encoder structure model to extract features and then apply K-means clustering to classify the features into local and teleseismic earthquakes. No other downstream task is considered to demonstrate efficiency of the learnt features.

In summary, SSL has been applied for various seismic signal analysis tasks. However, there is relatively little research on SSL techniques for event classification of seismic waveforms. Unlike abstract inputs such as images or text in Computer Vision (CV) and Natural Language Processing (NLP), which often use data augmentation and contrastive learning methods, seismic signal recordings have precise numerical values. This makes generative self-supervised learning particularly effective for extracting meaningful features. The uniqueness of our proposed approach, compared to the above reviewed literature, lies in its ability to classify various types of landslide micro-seismicity, along

with rockfalls and anthropogenic noise, using SSL with only a small amount of labeled data. We tackle the challenge of using at most 1375 labeled events in total and evaluate the model’s performance with various proportions of training data for linear evaluation. Additionally, we explore the effects of different dimensionality reduction methods on clustering, which involves four categories (in contrast to two-category clustering of [126], [121], [122], and [125]). To further enhance our analysis, we employ feature visualization methods to identify the primary causes of misclassification and to demonstrate how the inclusion of a small number of labels improves model performance. Additionally, we propose a novel approach to improve the efficiency of manually labelling data, using SOM, to perform multi-node clustering on the features learned by SSL. Human experts can thus label clusters of nodes instead of directly labeling raw data, significantly enhancing the efficiency of the labeling process.

## 5.3 Methodology

### 5.3.1 Data Pre-processing

Firstly, a 15-second seismic recording window is extracted from the continuous time-series data. Given the sampling rate of 250Hz and measurements recorded from three channels (East, North and Vertical), each extracted 15-sec window is a  $3750 \times 3$  array (3750 time samples over 3 channels), as described in Chapter 4. To focus on signals of interest, a BPF is applied, effectively removing low- and high-frequency noise that might obscure relevant seismic features. Unlike the 5 to 60 Hz filter used in Chapter 4, the filter frequency range is adjusted to 2–60 Hz to capture low-frequency earthquakes [127], [128]. Subsequently, normalisation is carried out by dividing each channel’s data by the maximum absolute value within that channel, thereby standardising the data range to  $[-1, 1]$ . A window size of 128 samples and an overlap of 70% are chosen for the STFT as in Chapter 3, resulting in an output shape of  $65 \times 98 \times 3$ , representing 65 frequency bins, 98 time steps (within each 15-sec window), and 3 channels. To prepare the data for SSL feature learning via auto-encoder, each STFT map is resized to a uniform size of  $64 \times 64 \times 3$  using bilinear interpolation, which is suitable

for model training including  $2 \times 2$  maxpooling.

### 5.3.2 Generative-based SSL using auto-encoder

An automatic feature extraction approach using an auto-encoder for generative-based SSL is proposed. The structure of the auto-encoder consists of an input layer followed by an encoder and a decoder, shown in Figure 5.1. The input layer takes in the pre-processed STFT maps of seismic events. The encoder part consists of several convolutional and pooling layers, progressively reducing the spatial dimensions while increasing the depth of the features extracted.

To train the auto-encoder to extract features from the STFT maps, the autoencoder input and output correspond to the same event’s STFT map, ensuring that the model learns to faithfully reconstruct the original input. The training process involves optimizing the autoencoder to minimize the mean squared error between the encoder input and decoder output STFT maps using the Adam optimizer. Additionally, a callback is employed to dynamically adjust the learning rate during training based on the loss. Specifically, if the training loss does not show a significant decrease in 5 epochs, the learning rate will decrease by 10%. Once trained, the decoder part, responsible for reconstructing the input from the learned features, is discarded. Only the encoder part is retained to extract features, producing a feature vector of shape  $4 \times 4 \times 16$ . This vector represents a compressed and abstract representation of the input seismic event, capturing its salient features.

### 5.3.3 Downstream task

After employing generative-based SSL to automatically extract features from seismic data recordings, we proceed with three downstream tasks.

#### Unsupervised learning with k-means

The proposed fully unsupervised method employs k-means clustering on the extracted features, flattened into a 256 ( $= 4 \times 4 \times 16$ )-length vector. Considering that high-dimensional data can negatively impact clustering performance [126], we utilize vari-

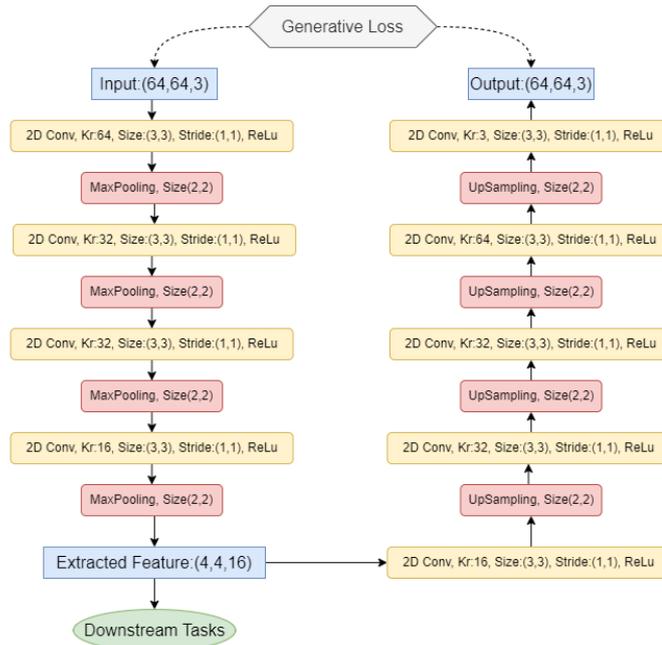


Figure 5.1: Generative-based SSL using auto-encoder. 'Kr' stands for kernel. The left part presents the encoder structure, while the decoder, removed after training, is shown on the right.

ous dimensionality reduction methods. Specifically, we apply t-distributed Stochastic Neighbour Embedding (t-SNE) [129] to reduce the 256-length vector to both 2 and 3 dimensions and PCA to reduce the dimensions to 2, 3 and 10 (with 10-D explaining 95.3% of the variance). Additionally, we use the original features with 256 dimensions for comparison. These features are clustered into  $k=4$  classes, which are labeled post-clustering as earthquake, rockfall, quake, and natural/anthropogenic noise.

### Semi-supervised ANN-based classification

The proposed semi-supervised ANN method constructs a classifier with a sequential architecture comprising several dense layers. The input shape of the ANN is 256 which is determined by the length of the flattened feature. The ANN consists of four dense layers: the first layer contains 128 neurons with a Rectified Linear Unit (ReLU) activation function; the second dense layer has 64 neurons and ReLU activation; the third dense layer comprises 32 neurons with ReLU activation; the final dense layer contains four neurons, representing the number of classes, with a softmax activation function

to output class probabilities. The model is compiled with the Adam optimizer and categorical cross-entropy loss function. We train the model using 5%, 10%, 20%, and 33.3% of the data.

### **SOM-based Method for Enhancing Manual Labeling Efficiency**

Finally, to significantly improve the efficiency of expert manual labeling data, we propose a SOM-based clustering, where we initialise various SOMs with the shape of  $8 \times 8$  (64 nodes),  $12 \times 12$  (144 nodes),  $17 \times 17$  (289 nodes), and  $21 \times 21$  (441 nodes) and use all unlabelled events for training. The SOM is trained with the following parameters that are heuristically set: the input length is determined by the shape of the flattened feature vectors (i.e., 256); the sigma parameter controls the spread of the neighborhood function (set to 1.5 for  $8 \times 8$ , 2 for  $12 \times 12$ , 2.5 for  $17 \times 17$  and 3.2 for  $21 \times 21$ ); the learning rate, initialized at 0.5, determines the rate at which the weights of the SOMs are adjusted. The neighborhood function was set to ‘Gaussian’. During training, the SOMs undergoes 2000 iterations to adapt their weights and organize the data based on similarities. Following training, the SOMs organize the large dataset into a reduced number of nodes without requiring labels. This approach enables experts to label the nodes rather than the entire dataset. However, the SOM may sometimes organize events incorrectly, and theoretically, the more nodes the SOMs have, the higher the accuracy. To reduce SOM errors, we use a very small amount of labels to perform majority voting and post-label the grid map produced by the SOMs.

## **5.4 Results and discussion**

### **5.4.1 Simulation setup**

The dataset used in this chapter is also the Résif dataset. The waveform recordings from the 3-component sensor are used as in Chapter 4.

In the unsupervised learning task, we use features from all 1375 labeled events. For the semi-supervised tasks, we utilize a small portion of the labeled data (5%, 10%, 20%, and 33.3%) to train the ANN. To ensure a balanced class representation in both training

and testing sets, we employ a stratified split strategy. Given that these portions of the dataset are very small and could potentially affect the training process, we employ k-fold cross-validation to ensure reliable evaluation results. When using 5%, 10%, 20%, 33.3% of the data, we employ a 20-fold, 10-fold, 5-fold, and 5-fold cross-validation strategy, respectively, with one fold used for training and all remaining data for testing.

The ANN model is implemented using the Keras framework and the SOM model is created by MiniSom [130] in Python. The K-Means clustering, PCA and t-SNE are all implemented using scikit-learn [131]. We utilized StratifiedKFold from scikit-learn to create the folds, ensuring that each fold preserves the percentage of samples for each class label.

#### 5.4.2 Unsupervised learning results and feature visualization

The results of the SSL-based clustering using k-means and different dimensionality reduction methods are shown in Table 5.1, averaged over all 4 classes. It can be seen that due to their ability to preserve local similarities, capturing non-linear relationships between data points, and robustness to outliers, 2D and 3D t-SNE achieve the highest accuracy of 72%, which is close to a 5% improvement compared to the case without dimensionality reduction (i.e., of 256 dimension) and the 10D PCA case that explains over 95% of variance.

Detailed results using 2D t-SNE features are presented in Table 5.2, while the results of other dimensionality reduction methods are provided in Appendix B. K-means achieves its highest F1 score (0.88) for rockfall events with N/A also displaying high precision and recall. However, the performance is lower for the seismic signals, namely quake and earthquake, where precision and recall scores vary significantly. While earthquake achieves a higher precision (0.82), quake shows better recall (0.67). The confusion matrix 5.3 further illustrates this. Although k-means have a certain ability to distinguish between quakes and earthquakes, some earthquakes are mistakenly clustered as quakes, resulting in low recall and precision of quakes.

To further analyse the misclassification of k-means, we visualize the 2-D features extracted through SSL and t-SNE in Figure 5.2, where we show the ground truth clus-

tering (left) and those generated by the proposed unsupervised approach. It can be seen from the ground truth that rockfall events are well separated with a clear cluster center and few outliers, resulting in highly accurate clustering results as can be seen from the right figure. On the other hand, some features of earthquake events closely resemble those of quake events, as evidenced by their proximity in the feature visualization map, thereby resulting in misclassification. Indeed, due to factors such as distance, intensity, and seismic source diversity, waveform patterns among earthquake events vary, and some of them are very close to the characteristics of quakes due to their low intensity and short duration. Furthermore, quake events typically manifest themselves as short duration, low-amplitude signals, often with poor signal-to-noise ratios, particularly when originating from distant sources relative to the sensor. Consequently, they may resemble N/A noise signals, as corroborated by the feature visualization map where some quake event features align with the N/A noise features. If we consider seismic signals, i.e., quakes and earthquakes, as one class, then a recall (sensitivity) of up to 80% for seismic events is achieved (quake and earthquake), 71% for rockfall, and 78% for N/A noise. In summary, the 2D feature space does not fully separate earthquake, quake and N/A noise signals, leading to some k-means classification errors.

Figure 5.2 shows that there are many earthquake events that are mixed in the feature space with quake events are consequently mis-classified by the unsupervised method. Although some earthquake features are close to those of quakes in the feature visualization map, clear cluster heads still differentiate between them. For instance, some quake features are concentrated at coordinates (-20, -20) on the feature map, whereas some earthquake features are distributed in a line at coordinates (-30, -20) to (0, -20). Although the features of this subset of earthquakes bear strong similarities to quakes, there are still subtle differences that might not be caught by unsupervised clustering algorithms. Thus, our next step involves employing a limited number of labels to enable semi-supervised ANN.

Table 5.1: Comparison of K-means Classification Performance Across Different Dimensionality Reduction Methods.

Input Features	Original (256D)	PCA (2D)	PCA (3D)	PCA (10D)	t-SNE (2D)	t-SNE (3D)
Accuracy (%)	67.13%	56.07%	66.69%	67.13%	<b>72.00%</b>	<b>72.00%</b>

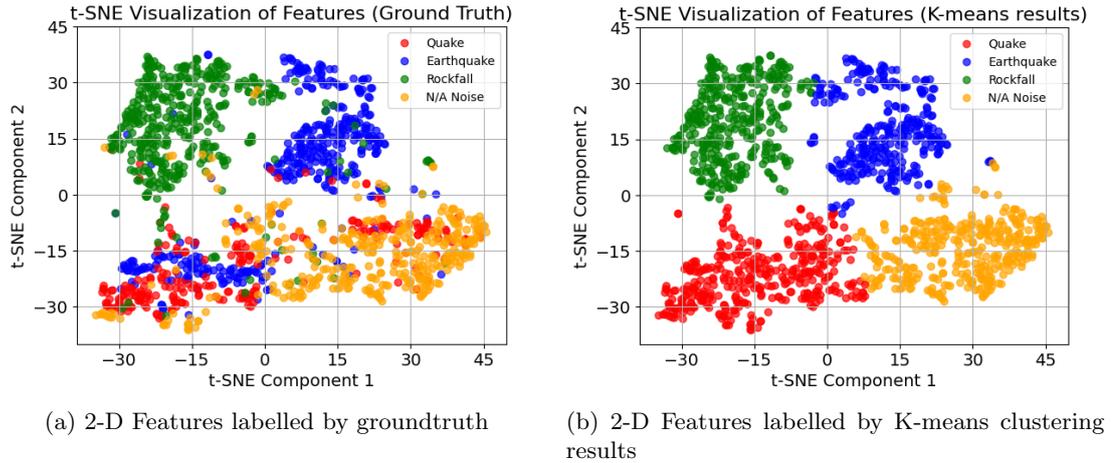


Figure 5.2: 2-D Feature visualization by t-SNE: (a) shows the features labelled by groundtruth, and (b) shows the features labelled by K-means clustering results.

Table 5.2: K-means Classification Performance Using 2D t-SNE Features

	precision	recall	F1-score
Quake	0.42	0.67	0.52
Earthquake	0.82	0.65	0.72
Rockfall	0.95	0.82	0.88
N/A	0.72	0.72	0.72

Table 5.3: K-means Confusion Matrix Using 2D t-SNE Features

	Quake	Earthquake	Rockfall	N/A
Quake	<b>157</b>	3	3	68
Earthquake	113	<b>252</b>	6	18
Rockfall	21	40	<b>331</b>	10
N/A	82	9	8	<b>251</b>

### 5.4.3 Semi-supervised ANN results

The classification performance results and confusion matrix for the ANN-based semi-supervised method, using 20% of the data for training, are presented in Tables 5.4 and 5.5, respectively. Each model was trained and tested using 5-fold cross-validation. The performance results are provided in the format “mean±standard deviation.”

As shown in Table 5.4, the ANN models provide better classification performance for all classes than K-means, as expected. Moreover, from the classification results

and the confusion matrix, using a small amount of labelled data for training effectively improves the results especially of earthquakes that are confused with quakes by k-means. Indeed, compared to K-means clustering, the recall of earthquakes has increased from 65% to 90%, while the precision of quakes has increased from 42% to 71%. Moreover, the obtained results are very close to those reported in Chapter 3, where a supervised CNN classifier was used with 70% of the data for training and validation, and 30% of the data for testing and validation. Specifically the F1-score for semi-supervised classification with SSL and ANN vs supervised CNN (in Chapter 3) on the same test STFT data are: earthquakes (0.88 vs. 0.96), quakes (0.68 vs. 0.86), rockfalls (0.91 vs 0.89) and N/A (0.83 vs 0.84). These clearly demonstrate that, compared to the requirement for large labelled datasets with supervised learning, performance is not negatively impacted through semi-supervised classification with robust SSL on smaller labelled datasets for training.

Table 5.4: The classification performance results for the proposed SSL-based ANN method.

	precision	recall	F1-score
Quake	0.715±0.029	0.646±0.040	0.677±0.017
Earthquake	0.864±0.020	0.900±0.033	0.881±0.015
Rockfall	0.916±0.027	0.909±0.021	0.912±0.014
N/A	0.822±0.033	0.837±0.037	0.828±0.016

Table 5.5: The confusion matrix for the proposed ANN-based method.

	Quake	Earthquake	Rockfall	N/A
Quake	<b>121.0±7.4</b>	22.8±2.9	6.0±2.0	37.4±8.7
Earthquake	17.2±5.5	<b>280.0±10.3</b>	8.6±4.1	5.4±1.85
Rockfall	6.6±3.6	14.0±4.3	<b>292.2±6.7</b>	8.8±4.7
N/A	25.6±14.2	7.6±4.2	12.4±5.8	<b>234.4±10.4</b>

Results for other training set sizes are included in the Appendix B. It can be seen from these results that, as expected, the overall accuracy steadily increases by the increase of the labelled training set size: for 5%, the accuracy is 75.8%; for 10%, it is 77.5%; for 20%, it is 84.3%; and for 33.3%, it is 87.5%.

#### 5.4.4 SOM-based method results

The previous two sub-sections demonstrated that by gently increasing the labelled data size, an increase in performance can be achieved from 72% without using any labels, to 87.5% when using one third of the dataset for training. In an attempt to strike better trade-off between labelling effort and performance, we leverage on SOMs. SOM is a powerful unsupervised tool for compressing large datasets into a manageable number of nodes, determined by the SOM model’s size. During training, when a new signal is introduced, the distance between the signal and each node in the map is calculated, and the node closest to the signal is designated as the Best Matching Unit (BMU). During the adaptation phase, the weights of the BMU are adjusted to more closely resemble the signal. This adjustment also influences the neighboring nodes, causing them to become more similar to the signal, thereby creating a neighborhood effect. This compression allows for significant reduction in workload and improved efficiency when manually labeling data, as annotators can focus on a smaller set of representative nodes instead of the entire dataset. However, despite the improved efficiency, SOM can sometimes incorrectly assign different types of events to the same node. Therefore, to enhance the efficiency of labeling using SOM, it is essential to specifically analyze the SOM clustering process by having experts label the SOM nodes.

Specifically, we assign each node to a class according to the majority of events it contains based on available labels. For example, if a SOM node contains 30 earthquakes, 5 rockfalls, and 1 N/A noise event, it will be labeled as an earthquake node. This approach simulates the process of human experts labeling nodes and helps in assessing the accuracy of SOM.

Table 5.6: The classification performance results for SOM.

	precision	recall	F1-score
Quake	0.74	0.67	0.70
Earthquake	0.83	0.91	0.87
Rockfall	0.93	0.91	0.92
N/A	0.83	0.81	0.82

The results and confusion matrix of SOM with  $17 \times 17$  size is shown in Table 5.6 and Table 5.7, respectively. The SOM results with other map sizes’ are shown in

Table 5.7: The confusion matrix for the SOM.

	Quake	Earthquake	Rockfall	N/A
Quake	<b>156</b>	35	5	38
Earthquake	12	<b>355</b>	11	11
Rockfall	9	17	<b>367</b>	9
N/A	34	20	13	<b>283</b>

Appendix B. The SOM-based clustering method achieves high precision, recall, and F1-scores across different classes, particularly for earthquake and rockfall events, very close to those achieved by the ANN-based method. Examining the confusion matrix, most quake events are correctly classified (156), but, similarly to ANN results, there are instances where quakes are confused with earthquake (35) and N/A noise (38). Figure 5.3a shows the  $17 \times 17$  SOM grid where each dot on the grid represents an event, and the grid structure visually clusters events based on their similarities. It can be seen that SOM separated well the N/A and rockfall events, occupying bottom left and top right corner of the map. Earthquake and quake events are in between, as expected, partly mixed, with quakes closer to the N/A noise signals due to their low signal-to-noise ratio, and earthquakes closer to rockfalls due to their longer duration and closer frequency representation.

Figure 5.3b shows the same SOM grid, but instead of plotting individual events, it displays pie charts at each node, based on the proportion of different seismic event classes assigned to the node. We can observe that single-coloured nodes are predominant which means that in most cases the events of the same type are assigned to the same node. Only few nodes have more than two events assigned to them, and these nodes are mainly positioned at the boundaries of the four class regions. For these nodes we apply majority voting to determine the output class. The more evenly a node is coloured, the lower our confidence in the classification output.

#### 5.4.5 Labelling Effort vs. Accuracy Trade-off

For an entire dataset consisting of 1375 seismic events, if an  $8 \times 8$  SOM is used, then experts need to annotate only 64 nodes. These 64 annotated nodes can then be used to label the entire dataset, reducing the experts' workload to just 4.6% compared to

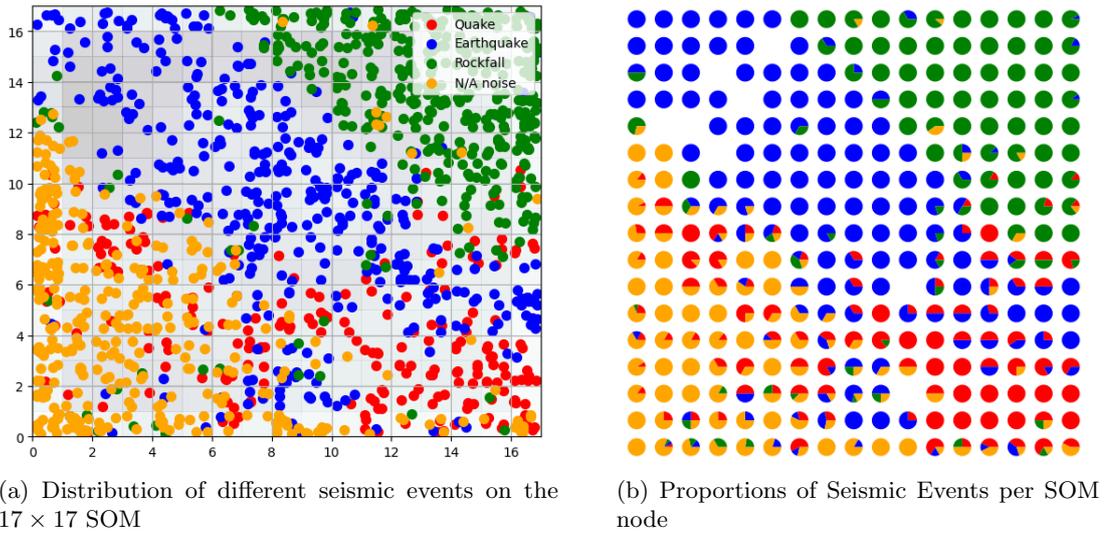


Figure 5.3: Visualization of the  $17 \times 17$  SOM Clustering Results. (a) Distribution of seismic events on the SOM, and (b) Proportions of seismic events per SOM node.

the case when each event is separately labelled. Note that, a 64-node SOM, requiring 4.6% of the data to be labeled, is comparable with an ANN trained using 5% of the data. Similarly, a 289(=  $17 \times 17$ )-node SOM, requiring 21% of the data to be labeled, is comparable with an ANN trained using 20% of the data. Hence, we use SOM configurations of  $8 \times 8$  (4.6%),  $12 \times 12$  (10.2%),  $17 \times 17$  (21%), and  $21 \times 21$  (32.1%) to make a fair comparison with ANNs trained on 5%, 10%, 20%, and 33.3% of the data.

Figure 5.4 shows the obtained accuracy performance of three different algorithms—ANN, SOM, and K-means across different levels of expert workload effort expressed as the percentage of data that needs to be labelled. The accuracy results are averaged over all four classes. ANN’s accuracy increases from approximately 75.8% to 87.5% as the expert workload percentage rises from 5% to 33.3%. SOM’s accuracy initially improves from around 79.8% to 84.4% but then plateaus, reaching 85.3% at 33.3% workload. It can be seen that 20% is a tipping point when the ANN method becomes superior.

This behaviour is expected since the optimization process of SOM is purely unsupervised. The labeled data is used for post-labeling, which limits SOM’s ability to improve accuracy significantly as more labeled data is provided, unlike ANN, which benefits more directly from the increased amount of labeled data during supervised

training. Additionally, as each iteration of a SOM feeds the same data, the blending of similar events (e.g., small earthquakes and quakes) may pull the SOM nodes closer together, which also contributes to the plateauing accuracy.

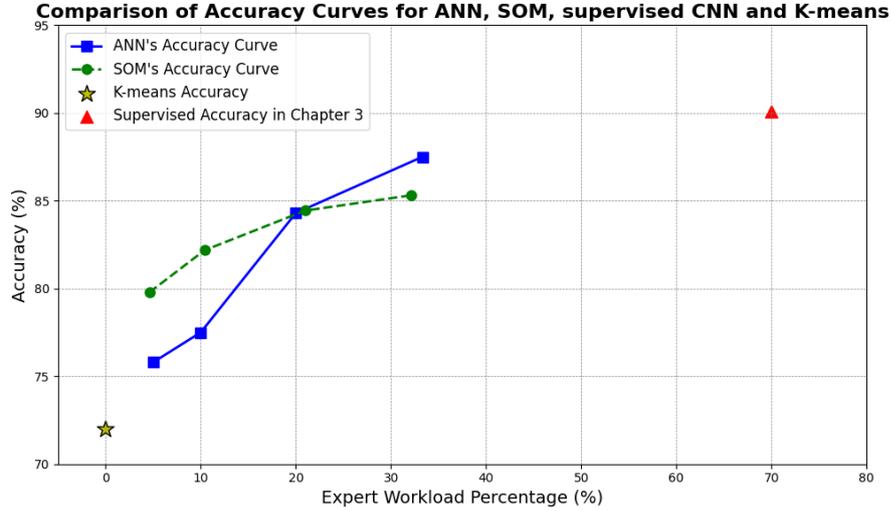


Figure 5.4: Accuracy Comparison of k-means, ANN, and SOM Methods with Varying Expert Labelling Workload Percentages

## 5.5 Summary

This chapter proposes a comprehensive approach to efficiently and accurately classify seismic events, regardless of the amount of labeled data, minimizing the resource-intensive manual labeling process. The proposed SSL framework extracts features from STFT maps of seismic events, producing compressed and informative feature vectors, which are used for different classification task, namely, unsupervised learning, semi-supervised learning, and enhancing manual labeling efficiency via SOMs.

The experiments demonstrate that the accuracy of unsupervised, k-means clustering benefits from dimensionality reduction techniques, with t-SNE achieving the highest accuracy due to its ability to preserve local similarities and non-linear relationships between data points. The semi-supervised ANN-based approach, trained with varying amounts of labeled data, better learns salient features of each class and outperforms k-means, particularly for earthquakes, which are often confused with micro-seismic sig-

nals, to due their similar frequency characteristics. As expected, ANNs rapidly improve performance as more labelled data are used for training, reaching 87.5% accuracy when one third of available data is used to train the model. Furthermore, SOM provides a valuable tool for reducing expert workload by organizing large datasets into a small number of map nodes, significantly enhancing labeling efficiency, by allowing experts to focus on representative nodes rather than the entire dataset. Our analysis shows that SOM outperforms ANN when the amount of labelled data is below 20% of the total dataset.

## Chapter 6

# Conclusion and Future Work

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

### 6.1 Summary

Recent advancements in passive seismic monitoring have accelerated the use of automated microseismic event detection and classification, especially through deep learning techniques such as CNNs. However, achieving accuracy, interpretability, and scalability in seismic monitoring using deep learning remains challenging, particularly due to the limitations of black-box models and the high demand for labeled data. This research introduces a comprehensive system addressing these challenges by leveraging CNNs for microseismic classification, explainable AI (XAI) for interpretability, and self-supervised learning (SSL) for reducing the need for extensive labeled datasets.

To address the key challenging in seismic event analysis, the following methodologies are proposed. First, Chapter 3 presents an end-to-end CNN-based classification model, designed to identify seismic events directly from continuous recordings without requiring multi-steps. Three CNN models were developed using different types of inputs: temporal waveform, STFT, and CWT maps. These models were trained on the Résif dataset to classify seismic events into earthquakes, quakes, and rockfalls. The models demonstrated strong performance, achieving an overall classification accu-

racy of approximately 90% on the Résif dataset. Testing on an independent dataset from Larissa, Greece, further confirmed the model’s robustness and adaptability across diverse geological settings and monitoring configurations.

Chapter 4 addresses the interpretability of CNNs in seismic monitoring by incorporating LRP, a technique from XAI that maps feature importance within the model’s decision-making process. This chapter emphasizes how LRP helps human experts by visually identifying features leading to specific classifications, such as noise signals contributing to misclassifications between quake and rockfall events. By combining human oversight with automated classification, the approach improves the accuracy and trustworthiness of seismic event catalogues.

To minimise the reliance on extensive labeled data, Chapter 5 introduces an SSL framework based on STFT maps of seismic signals, which reduces manual labeling efforts while maintaining classification accuracy. The framework employs t-SNE for dimensionality reduction within SSL features and self-organizing maps (SOMs) to enhance efficiency in expert labeling. Experimental results show that using the SSL-based approach, the proportion of manually labeled data required for training could be reduced by approximately 50% (from 70% in the fully supervised case to only 20%) while still achieving classification performance close to that of supervised learning models. Additionally, SOMs are particularly effective in allowing experts to focus on a small subset of representative map nodes, significantly optimizing the manual annotation process.

Importantly, the approaches developed in this thesis are highly transferable to other domains involving time-series or event-based data. Domains such as environmental monitoring, structural health assessment, and industrial sensor data analysis could benefit from the proposed CNN classification models, explainability techniques, and SSL-based label reduction frameworks. This cross-domain applicability highlights the broader potential impact of this research beyond seismology.

Together, the methodologies presented in this thesis offer a scalable, transferable, explainable, and efficient system for seismic event detection and classification. This work advances the field of passive seismic monitoring by providing a robust frame-

work that meets the demands for high accuracy, transparency, and reduced reliance on labeled data, facilitating proactive management of seismic risks and furthering our understanding of subsurface processes.

## 6.2 Future Work

Recent advancements in deep learning techniques enhance the accuracy and effectiveness of seismic monitoring. While this research addresses many challenges associated with the analysis of (micro)seismic events in practical applications, several limitations remain. This section outlines a number of recommendations for future research.

### 1. **Enhancing Generalization, Transferability, and Robustness Through Diverse and Noisy Data**

Although the models developed in this thesis showed strong performance across varied geological settings, further improvements in generalization, robustness, and cross-domain applicability are essential. Future work should focus on expanding the diversity of seismic datasets, including data from geologically distinct regions and environments with high ambient noise. Training models on such data can improve resilience to noise and enable more accurate classification in real-world scenarios where signal quality may be low. In addition, the transferability of the proposed methods to other domains—such as environmental monitoring, structural health analysis, and industrial sensor data interpretation—presents a valuable avenue for exploration. Investigating the adaptability of the frameworks developed here to different types of time-series or event-based data could extend their impact and uncover new insights into model limitations and scalability. Collectively, these efforts would foster the creation of more robust and widely applicable deep learning systems for complex signal analysis.

### 2. **Quantitative Analysis of Interpretability in Seismic Event Classification**

Although the human-in-the-loop framework has successfully provided interpretability for deep learning-based seismic classification and assisted experts in labeling,

future work could focus on developing quantitative metrics to evaluate the quality of the generated LRP maps. Metrics commonly used in generative text or image models to assess generated outputs could inspire further advancements in interpretability evaluation [132]. For example, coherence, which evaluates the logical consistency of generated content, could be adapted to ensure that relevance maps consistently highlight meaningful and relevant features. Similarly, coverage, a metric used to measure whether all critical information is included, could be applied to ensure that the relevance maps capture all essential signal features without overlooking key details. By integrating such metrics, researchers can achieve a more robust evaluation of the reliability and quality of interpretability methods, such as LRP, further refining these approaches to enhance trust, accuracy, and usability in seismic event detection systems.

### 3. **Enhancing Seismic Event Annotation with Expert Confidence Modeling**

One limitation of our approach for enhancing manual labeling efficiency is the assumption that labels provided by the expert are always considered to be correct. Extending the method to allow labelling errors and possibly multiple labels per event provided by different experts that include expert confidence levels, is an interesting direction of future research. Another possible approach to enhance labelling is the use of explainable AI to assist experts in labelling. Using a human-in-the-loop labelling framework based on XAI will enable experts to have clear, explainable insights into the AI's decision-making process, thereby supporting more accurate annotations. This integrated approach is expected to enhance both the efficiency and accuracy of the annotation process, reducing the risk of human error.

### 4. **Analysing Feature Learning Dynamics via LRP Under Varying Labelling Regimes**

A promising direction for future research is to systematically analyze how the amount of labeled training data influences the internal feature representations

learned by the model. By applying LRP to models trained with different proportions of labeled data (e.g., 5%, 10%, 20%, and 33%), researchers can visualize and compare the relevance maps to track how the model’s attention to different signal components evolves as more annotated examples are provided. This analysis may reveal whether the model increasingly focuses on geophysically meaningful features, suppresses irrelevant noise, or learns more abstract, high-level patterns with greater data availability. Such interpretability-driven diagnostics can enhance our understanding of self-supervised and semi-supervised learning dynamics, guide the design of more label-efficient training strategies, and improve model reliability in data-scarce scenarios.

# Bibliography

- [1] L. Trani, G. A. Pagani, J. P. P. Zanetti, C. Chapeland, and L. Evers, “Deep-quake—an application of cnn for seismo-acoustic event classification in the netherlands,” *Computers & Geosciences*, vol. 159, p. 104980, 2022.
- [2] J. Gomberg, P. Bodin, W. Savage, and M. E. Jackson, “Landslide faults and tectonic faults, analogs?: The slumgullion earthflow, colorado,” *Geology*, vol. 23, no. 1, pp. 41–44, 1995.
- [3] A. Tonnellier, A. Helmstetter, J.-P. Malet, J. Schmittbuhl, A. Corsini, and M. Joswig, “Seismic monitoring of soft-rock landslides: the super-sauze and valoria case studies,” *Geophysical J. Int.*, vol. 193, no. 3, pp. 1515–1536, 2013.
- [4] M. Walter, J. Gomberg, W. Schulz, P. Bodin, and M. Joswig, “Slidequake generation versus viscous creep at softrock-landslides: synopsis of three different scenarios at slumgullion landslide, heumoes slope, and super-sauze mudslide,” *J. Env. & Eng. Geophy.*, vol. 18, pp. 269–280, 2013.
- [5] S. L. Gariano and F. Guzzetti, “Landslides in a changing climate,” *Earth-Science Reviews*, vol. 162, pp. 227–252, 2016.
- [6] S. Fuchs, “Vulnerability to natural hazards—the challenge of integration,” *Natural Hazards*, vol. 52, no. 1, pp. 3–20, 2009.
- [7] S. M. Mousavi and G. C. Beroza, “Deep-learning seismology,” *Science*, vol. 377, no. 6607, p. eabm4470, 2022.

## Bibliography

- [8] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, “Stanford earthquake dataset (stead): A global data set of seismic signals for ai,” *IEEE Access*, vol. 7, pp. 179464–179476, 2019.
- [9] Y. Duan, Y. Shen, I. Canbulat, and G. Si, “Classification of clustered microseismic events in a coal mine using machine learning,” *J. Rock Mechanics and Geotech. Eng.*, 2021.
- [10] M. Shakeel, K. Itoyama, K. Nishida, and K. Nakadai, “Emc: Earthquake magnitudes classification on seismic signals via convolutional recurrent networks,” in *IEEE/SICE Int. Symp. Sys. Integrat.*, pp. 388–393, 2021.
- [11] T. Perol, M. Gharbi, and M. Denolle, “Convolutional neural network for earthquake detection and location,” *Science Advances*, vol. 4, no. 2, p. e1700578, 2018.
- [12] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature machine intelligence*, vol. 1, no. 5, pp. 206–215, 2019.
- [13] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, “Concrete problems in ai safety,” *arXiv preprint arXiv:1606.06565*, 2016.
- [14] V. I. Grechka and W. M. Heigl, *Microseismic monitoring*. Society of Exploration Geophysicists Tulsa, Oklahoma, USA, 2017.
- [15] 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.
- [16] 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.
- [17] N. Vouillamoz, S. Rothmund, and M. Joswig, “Characterizing the complexity of microseismic signals at slow-moving clay-rich debris slides: the super-sauze

## Bibliography

- (southeastern france) and pechgrabe (upper austria) case studies.,” *Earth Surface Dynamics*, vol. 6, no. 2, 2018.
- [18] Z. Ma and G. Mei, “Deep learning for geological hazards analysis: Data, models, applications, and opportunities,” *Earth-Science Reviews*, vol. 223, p. 103858, 2021.
- [19] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, “Automated platform for microseismic signal analysis: Denoising, detection, and classification in slope stability studies,” *IEEE Trans. Geoscience & Remote Sensing*, 2020.
- [20] 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.
- [21] C. Trabant, A. R. Hutko, M. Bahavar, R. Karstens, T. Ahern, and R. L. Newman, “Data products at the iris dmc: Stepping stones for research and other applications,” *Seismological Research Letters*, vol. 83, no. 5, pp. 846–854, 2012.
- [22] 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.
- [23] F. Li, J. Rich, K. J. Marfurt, and H. Zhou, “Automatic event detection on noisy microseismograms,” in *2014 SEG Annual Meeting*, OnePetro, 2014.
- [24] J. Akram, “Automatic p-wave arrival time picking method for seismic and microseismic data,” in *CSPG CSEG CWLS Convention*, CSEG Calgary, AB, Canada, 2011.
- [25] 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.

## Bibliography

- [26] F. Provost, C. Hibert, and J.-P. Malet, “Automatic classification of endogenous landslide seismicity using the random forest supervised classifier,” *Geophy. Research Let.*, vol. 44, no. 1, pp. 113–120, 2017.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] S. M. Mousavi and C. A. Langston, “Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding,” *Bulletin Seism. Society America*, vol. 106, no. 4, pp. 1380–1393, 2016.
- [33] J. Akram and D. Eaton, “Adaptive microseismic event detection and automatic time picking,” in *2012 CSEG Annual Convention*, vol. 15, 2012.
- [34] 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.

## Bibliography

- [35] 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.
- [36] J. I. Sabbione and D. Velis, “Automatic first-breaks picking: New strategies and algorithms,” *Geophysics*, vol. 75, no. 4, pp. V67–V76, 2010.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] J. Akram and D. W. Eaton, “A review and appraisal of arrival-time picking methods for downhole microseismic data arrival-time picking methods,” *Geophysics*, vol. 81, no. 2, pp. KS71–KS91, 2016.
- [41] 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.
- [42] 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.
- [43] R. Ghosh and H. Sardana, “Multi-feature optimization strategies for target classification using seismic and acoustic signatures (conference presentation),” in *Au-*

## Bibliography

- tomatic Target Recognition XXX*, vol. 11394, p. 113940I, Proc SPIE Int Soc Opt Eng, 2020.
- [44] A. Parasyris, L. Stankovic, and V. Stankovic, “Dimensionality reduction for visualization of hydrogeophysical and meteorological recordings on a landslide zone,” in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1864–1868, IEEE, 2024.
- [45] 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*, pp. 1398–1406, 2017.
- [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,” tech. rep., Technical Report, 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)*, pp. 2–6, IET, 2017.
- [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)*, pp. 1–6, IEEE, 2019.

## Bibliography

- [52] P. E. E. Lara, C. A. R. Fernandes, A. Inza, J. I. Mars, J.-P. Métaixian, 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] Y. LeCun, Y. Bengio, G. Hinton, *et al.*, “Deep learning. nature 521 (7553), 436-444,” *Google Scholar Google Scholar Cross Ref Cross Ref*, 2015.
- [54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [55] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [56] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [57] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [58] C. E. Yoon, O. O’Reilly, K. J. Bergen, and G. C. Beroza, “Earthquake detection through computationally efficient similarity search,” *Science Advances*, vol. 1, no. 11, p. e1501057, 2015.
- [59] FrenchLandslideObservatorySeismologicalDatacenter/RESIF, “Observatoire multi-disciplinaire des instabilites de versants (omiv).” Accessed: 2021. [online]. Available: <https://seismology.resif.fr/>. DOI: 10.15778/RESIF.MT.
- [60] J.-P. Malet, D. Laigle, A. Remaître, and O. Maquaire, “Triggering conditions and mobility of debris flows associated to complex earthflows,” *Geomorphology*, vol. 66, no. 1-4, pp. 215–235, 2005.

## Bibliography

- [61] A. Helmstetter and S. Garambois, “Seismic monitoring of séchilienne rockslide (french alps): Analysis of seismic signals and their correlation with rainfalls,” *J. Geoph. Res.: Earth Surface*, vol. 115, no. F3, 2010.
- [62] M. Walter, C. Arnhardt, and M. Joswig, “Seismic monitoring of rockfalls, sub-surface fracture processes, and superficial fissure development at the super-sauze, french alps, mudslide,” *Eng. Geology*, 2011.
- [63] C. Hibert, A. Mangeney, G. Grandjean, and N. M. Shapiro, “Slope instabilities in dolomieu crater, réunion island: from seismic signals to rockfall characteristics,” *Journal of Geophysical Research: Earth Surface*, vol. 116, no. F4, 2011.
- [64] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, “Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms,” *J. Volcanology & Geoth. research*, vol. 180, no. 1, pp. 1–8, 2009.
- [65] C. Hibert, A. Mangeney, G. Grandjean, C. Baillard, D. Rivet, N. M. Shapiro, C. Satriano, A. Maggi, P. Boissier, V. Ferrazzini, *et al.*, “Automated identification, location, and volume estimation of rockfalls at piton de la fournaise volcano,” *J. Geophy. Research: Earth Surface*, vol. 119, no. 5, pp. 1082–1105, 2014.
- [66] A. E. Ruano, G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, and P. M. Ferreira, “Seismic detection using support vector machines,” *Neurocomputing*, vol. 135, pp. 273–283, 2014.
- [67] M. Wenner, C. Hibert, A. van Herwijnen, L. Meier, and F. Walter, “Near-real-time automated classification of seismic signals of slope failures with continuous random forests,” *Natural Hazards and Earth System Sciences*, vol. 21, pp. 339–361, 2021.
- [68] R. Tibi, L. Linville, C. Young, and R. Brogan, “Classification of local seismic events in the utah region: A comparison of amplitude ratio methods with a spectrogram-based machine learning approach,” *Bulletin Seism. Society America*, vol. 109, no. 6, pp. 2532–2544, 2019.

## Bibliography

- [69] Y. Chen, G. Zhang, M. Bai, S. Zu, Z. Guan, and M. Zhang, “Automatic waveform classification and arrival picking based on convolutional neural network,” *Earth & Space Science*, vol. 6, pp. 1244–1261, 2019.
- [70] 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.
- [71] W. Zhu and G. C. Beroza, “Phasenet: a deep-neural-network-based seismic arrival-time picking method,” *Geophysical J. Int.*, vol. 216, no. 1, pp. 261–273, 2019.
- [72] B. Ku, J. Min, J.-K. Ahn, J. Lee, and H. Ko, “Earthquake event classification using multitasking deep learning,” *IEEE Geoscience & Remote Sensing Let.*, 2020.
- [73] X. Liao, J. Cao, J. Hu, J. You, X. Jiang, and Z. Liu, “First arrival time identification using transfer learning with continuous wavelet transform feature images,” *IEEE Geoscience & Remote Sensing Let.*, vol. 17, pp. 2002–2006, 2019.
- [74] G. Zhang, C. Lin, and Y. Chen, “Convolutional neural networks for micro-seismic waveform classification and arrival picking,” *Geophysics*, vol. 85, no. 4, pp. WA227–WA240, 2020.
- [75] R. M. Dokht, H. Kao, R. Visser, and B. Smith, “Seismic event and phase detection using time–frequency representation and convolutional neural networks,” *Seismological Research Let.*, vol. 90, pp. 481–490, 2019.
- [76] L. Linville, K. Pankow, and T. Draelos, “Deep learning models augment analyst decisions for event discrimination,” *Geophy. Research Let.*, vol. 46, no. 7, pp. 3643–3651, 2019.
- [77] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, “Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking,” *Nature Commun.*, vol. 11, pp. 1–12, 2020.

## Bibliography

- [78] O. M. Saad, Y. Chen, A. Savvaidis, S. Fomel, and Y. Chen, “Real-time earthquake detection and magnitude estimation using vision transformer,” *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 5, p. e2021JB023657, 2022.
- [79] O. M. Saad, G. Huang, Y. Chen, A. Savvaidis, S. Fomel, N. Pham, and Y. Chen, “Scalodeep: A highly generalized deep learning framework for real-time earthquake detection,” *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 4, p. e2020JB021473, 2021.
- [80] L. Liu, W. Song, C. Zeng, and X. Yang, “Microseismic event detection and classification based on convolutional neural network,” *J. Applied Geoph.*, p. 104380, 2021.
- [81] L. Linville, R. C. Brogan, C. Young, and K. A. Aur, “Global-and local-scale high-resolution event catalogs for algorithm testing,” *Seism. Research Let.*, vol. 90, no. 5, pp. 1987–1993, 2019.
- [82] U. M. Park, “Usgs northern california network,” *International Federation of Digital Seismograph Networks, Dataset/Seismic Network, doi*, vol. 10, 1967.
- [83] S. Horton, “Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central arkansas with potential for damaging earthquake,” *Seismological Research Letters*, vol. 83, no. 2, pp. 250–260, 2012.
- [84] NECIS. Accessed: 2021. [online]. Available: <http://necis.kma.go.kr/>.
- [85] IRIS. Accessed: 2021. [online]. Available: <https://www.iris.edu/>.
- [86] A. Savvaidis, B. Young, G.-c. D. Huang, and A. Lomax, “Texnet: A statewide seismological network in texas,” *Seismological Research Letters*, vol. 90, no. 4, pp. 1702–1715, 2019.
- [87] KNMI, “Netherlands seismic and acoustic network,” *Royal Netherlands Meteorological Institute (KNMI), Other/Seismic Network.(10.21944/E970fd34-23b9-3411-B366-E4f72877d2c5)*, 1993.

## Bibliography

- [88] EDIA. Accessed: 2021. [online]. Available: <http://www.orfeus-eu.org/data/eida/>.
- [89] Z. E. Ross, M.-A. Meier, E. Hauksson, and T. H. Heaton, “Generalized seismic phase detection with deep learning,” *Bulletin of the Seismological Society of America*, vol. 108, no. 5A, pp. 2894–2901, 2018.
- [90] O. M. Saad and Y. Chen, “Capsphase: Capsule neural network for seismic phase classification and picking,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.
- [91] G.I. Accessed: 2021. [online]. Available: <http://www.gein.noa.gr/en/>.
- [92] V. Karakostas, C. Papazachos, E. Papadimitriou, M. Foumelis, A. Kiratzi, C. Pikridas, A. Kostoglou, C. Kkallas, N. Chatzis, S. Bitharis, *et al.*, “The march 2021 tyrnavos, central greece, doublet (mw6. 3 and mw6. 0): Aftershock relocation, faulting details, coseismic slip and deformation,” *Bulletin of the Geological Society of Greece*, vol. 58, pp. 131–178, 2021.
- [93] I. Itasca Consulting Group, “Insite seismic processor, version 3.15,” 2019. Accessed: 2023. [online]. Available: [https://itasca-downloads.s3.amazonaws.com/documents/offices/UK/InSiteHelp\\_39.pdf](https://itasca-downloads.s3.amazonaws.com/documents/offices/UK/InSiteHelp_39.pdf).
- [94] J. Jiang, V. Stankovic, L. Stankovic, E. Parastatidis, and S. Pytharouli, “Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.
- [95] C. Sonesson, S. Gerster, and M. Delorenzi, “Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation,” *PloS one*, vol. 9, no. 6, p. e100335, 2014.
- [96] M. Hägele, P. Seegerer, S. Lapuschkin, M. Bockmayr, W. Samek, F. Klauschen, K.-R. Müller, and A. Binder, “Resolving challenges in deep learning-based anal-

## Bibliography

- yses of histopathological images using explanation methods,” *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [97] J. Li, L. Stankovic, V. Stankovic, S. Pytharouli, C. Yang, and Q. Shi, “Graph-based feature weight optimisation and classification of continuous seismic sensor array recordings,” *Sensors*, vol. 23, no. 1, 2023.
- [98] D. Bau, J.-Y. Zhu, H. Strobelt, A. Lapedriza, B. Zhou, and A. Torralba, “Understanding the role of individual units in a deep neural network,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30071–30078, 2020.
- [99] A. Holzinger, “From machine learning to explainable ai,” in *2018 world symposium on digital intelligence for systems and machines (DISA)*, pp. 55–66, IEEE, 2018.
- [100] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, “Evaluating the visualization of what a deep neural network has learned,” *IEEE transactions on neural networks and learning systems*, vol. 28, no. 11, pp. 2660–2673, 2016.
- [101] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, “Layer-wise relevance propagation: an overview,” *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193–209, 2019.
- [102] S. Lapuschkin, A. Binder, G. Montavon, K.-R. Müller, and W. Samek, “Analyzing classifiers: Fisher vectors and deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2912–2920, 2016.
- [103] D. Murray, L. Stankovic, and V. Stankovic, “Transparent ai: explainability of deep learning based load disaggregation,” in *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 268–271, 2021.
- [104] E. Commission, C. Directorate-General for Communications Networks, and Technology, *Ethics guidelines for trustworthy AI*. Publications Office, 2019.

## Bibliography

- [105] U. N. S. D. G. 13, 2023. Accessed: 2023. [online]. Available: <https://sdgs.un.org/goals>.
- [106] 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 pechgraben (upper austria) case studies,” *Earth Surface Dynamics*, vol. 6, no. 2, pp. 525–550, 2018.
- [107] A. Renouard, A. Maggi, M. Grunberg, C. Doubre, and C. Hibert, “Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning,” *Seismological Society of America*, vol. 92, no. 6, pp. 3725–3742, 2021.
- [108] X. Bi, C. Zhang, Y. He, X. Zhao, Y. Sun, and Y. Ma, “Explainable time–frequency convolutional neural network for microseismic waveform classification,” *Information Sciences*, vol. 546, pp. 883–896, 2021.
- [109] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” *International Journal of Computer Vision*, vol. 128, pp. 336–359, oct 2019.
- [110] F. Provost, J.-P. 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, pp. 1455–1473, 08 2018.
- [111] D. Murray, L. Stankovic, S. Pytharouli, and V. Stankovic, “Semi-supervised seismic event detection using siamese networks,” Apr. 2023. EGU General Assembly 2023, April 2023.
- [112] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, “On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation,” *PloS one*, vol. 10, no. 7, p. e0130140, 2015.

## Bibliography

- [113] A. Chan, M. Schneider, and M. Körner, “Xai for early crop classification,” in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2657–2660, IEEE, 2023.
- [114] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, “Explaining nonlinear classification decisions with deep taylor decomposition,” *Pattern recognition*, vol. 65, pp. 211–222, 2017.
- [115] J. Jiang, D. Murray, V. Stankovic, L. Stankovic, C. Hibert, S. Pytharouli, and J.-P. Malet, “A human-on-the-loop approach for labelling landslide-induced seismic recordings via a multi-class deep-learning based classification model,” 01 2024.
- [116] K. Masuda, S. Ide, K. Ohta, and T. Matsuzawa, “Bridging the gap between low-frequency and very-low-frequency earthquakes,” *Earth, Planets and Space*, vol. 72, no. 1, pp. 1–9, 2020.
- [117] P. Kumar, P. Rawat, and S. Chauhan, “Contrastive self-supervised learning: review, progress, challenges and future research directions,” *International Journal of Multimedia Information Retrieval*, vol. 11, no. 4, pp. 461–488, 2022.
- [118] X. Yang, Z. Song, I. King, and Z. Xu, “A survey on deep semi-supervised learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 9, pp. 8934–8954, 2022.
- [119] K. Zhang, Q. Wen, C. Zhang, R. Cai, M. Jin, Y. Liu, J. Y. Zhang, Y. Liang, G. Pang, D. Song, *et al.*, “Self-supervised learning for time series analysis: Taxonomy, progress, and prospects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [120] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.
- [121] R. U. Murshed, K. Noshin, M. A. Zakaria, M. F. Uddin, A. S. Amin, and M. E. Ali, “Real-time seismic intensity prediction using self-supervised contrastive gnn

## Bibliography

- for earthquake early warning,” *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
- [122] H. Li, J. He, X. Tuo, X. Wen, and Z. Yang, “Self-supervised convolutional clustering for picking the first break of microseismic recording,” *IEEE Geoscience and Remote Sensing Letters*, 2024.
- [123] D. Lee, E. Aune, N. Langet, and J. Eidsvik, “Ensemble and self-supervised learning for improved classification of seismic signals from the åknes rockslope,” *Mathematical Geosciences*, vol. 55, no. 3, pp. 377–400, 2023.
- [124] D. Lee and E. Aune, “Vibcreg: Variance-invariance-better-covariance regularization for self-supervised learning on time series,” *arXiv preprint arXiv:2109.00783*, vol. 2, no. 5, 2021.
- [125] Y. Song, E. Wang, C. Liu, Y. Li, H. Yang, B. Li, D. Chen, and Y. Di, “Intelligent recognition of valid microseismic events based on self-supervised learning,” *Measurement*, p. 114760, 2024.
- [126] S. M. Mousavi, W. Zhu, W. Ellsworth, and G. Beroza, “Unsupervised clustering of seismic signals using deep convolutional autoencoders,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1693–1697, 2019.
- [127] Q.-Y. Wang, W. B. Frank, R. E. Abercrombie, K. Obara, and A. Kato, “What makes low-frequency earthquakes low frequency,” *Science Advances*, vol. 9, no. 32, p. eadh3688, 2023.
- [128] X. Wei, J. Xu, Y. Liu, and X. Chen, “The slow self-arresting nature of low-frequency earthquakes,” *Nature communications*, vol. 12, no. 1, p. 5464, 2021.
- [129] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.,” *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [130] G. Vettigli, “Minisom: minimalistic and numpy-based implementation of the self organizing map,” 2018.

## Bibliography

- [131] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [132] D. Batic, V. Stankovic, and L. Stankovic, “Towards transparent load disaggregation—a framework for quantitative evaluation of explainability using explainable ai,” *IEEE Transactions on Consumer Electronics*, 2023.

# Appendix A

## Catalogue of Larisa data

The following tables present detection results comparing our CNN model's performance (Arrays A1/A2) against manual detection (M1/M2) and InSite software (I1/I2) for seismic events in Larissa, Greece. Tables show detailed detections during 18:53-19:53 on 17 March 2021, including earthquakes, quakes, and rockfalls.

Appendix A. Catalogue of Larisa data

Table A.1: Detection results of seismic events (Larisa Dataset, Part 1/3)

Class	Time	Detected						Duration (s)
		A1	M1	I1	A2	M2	I2	
earthquake	18:53:46	✓	✓		✓	✓		0.66
earthquake	18:54:20	✓	✓		✓	✓		1.14
earthquake	18:54:47	✓	✓	✓	✓	✓	✓	1.29
earthquake	18:54:57				✓	✓		0.99
earthquake	18:55:47	✓	✓		✓	✓	✓	0.20
earthquake	18:56:10					✓	✓	2.50
earthquake	18:56:25	✓	✓	✓	✓	✓	✓	1.40
earthquake	18:57:05	✓	✓		✓	✓	✓	1.10
earthquake	18:57:13	✓	✓		✓	✓		1.30
earthquake	18:57:32	✓	✓		✓	✓		2.40
earthquake	18:58:03	✓	✓	✓	✓	✓	✓	0.70
earthquake	19:00:09	✓	✓		✓	✓	✓	1.20
earthquake	19:01:02		✓			✓	✓	0.60
earthquake	19:02:55	✓	✓	✓	✓	✓	✓	2.10
earthquake	19:03:56	✓	✓			✓		0.90
earthquake	19:04:06	✓	✓			✓		2.50
earthquake	19:04:48	✓						10.00
earthquake	19:06:29	✓	✓	✓	✓	✓	✓	1.50
earthquake	19:06:51		✓			✓	✓	1.10
earthquake	19:07:04	✓	✓			✓	✓	4.10
earthquake	19:08:16	✓	✓	✓	✓	✓		2.30
earthquake	19:09:08	✓	✓	✓	✓	✓	✓	1.60
earthquake	19:11:02	✓	✓	✓	✓	✓	✓	2.00
rockfall	19:11:30	✓	✓	✓	✓	✓	✓	1.00
earthquake	19:12:47				✓			1.00
earthquake	19:13:53	✓			✓	✓		2.00
earthquake	19:14:09	✓		✓	✓			5.00
earthquake	19:14:57	✓	✓	✓	✓	✓	✓	3.00
earthquake	19:17:18	✓	✓	✓	✓	✓		1.00
earthquake	19:17:37	✓	✓		✓	✓	✓	1.50
earthquake	19:17:52		✓	✓		✓	✓	1.10
earthquake	19:18:38	✓	✓	✓	✓	✓	✓	1.00
earthquake	19:19:26	✓	✓	✓	✓	✓		0.60
rockfall	19:19:58			✓	✓	✓		0.70
quake	19:21:11	✓	✓	✓	✓			4.00
rockfall	19:21:23				✓			4.00
rockfall	19:21:37			✓	✓			3.00
earthquake	19:21:54				✓			1.00
quake	19:22:15	✓	✓		✓			3.00

Appendix A. Catalogue of Larisa data

Table A.2: Detection results of seismic events (Larisa Dataset, Part 2/3)

Class	Time	Detected						Duration (s)
		A1	M1	I1	A2	M2	I2	
earthquake	19:22:38	✓	✓	✓	✓	✓	✓	2.00
earthquake	19:22:50					✓		1.20
earthquake	19:23:00					✓		0.20
earthquake	19:23:17	✓	✓	✓	✓	✓	✓	1.80
rockfall	19:24:14		✓		✓	✓		2.50
earthquake	19:24:25		✓			✓		2.50
earthquake	19:24:35		✓	✓	✓	✓	✓	2.00
rockfall	19:25:00				✓			2.00
rockfall	19:25:12	✓	✓	✓	✓		✓	8.00
earthquake	19:25:40				✓			2.00
rockfall	19:25:00				✓			4.00
earthquake	19:26:27	✓	✓		✓	✓	✓	1.00
rockfall	19:26:48				✓			1.00
quake	19:26:57	✓	✓		✓		✓	4.00
rockfall	19:27:46				✓			5.00
earthquake	19:28:00	✓	✓	✓	✓	✓	✓	0.90
quake	19:28:08				✓			2.00
earthquake	19:28:21				✓			3.00
earthquake	19:28:40				✓			14.00
earthquake	19:29:23	✓	✓		✓			4.00
rockfall	19:29:39			✓	✓			4.00
earthquake	19:29:55	✓	✓	✓	✓	✓	✓	2.70
earthquake	19:30:42	✓	✓	✓	✓			5.00
earthquake	19:31:24	✓	✓					8.00
earthquake	19:31:09			✓	✓			2.00
earthquake	19:31:47				✓			1.00
rockfall	19:32:07	✓	✓	✓	✓			8.00
earthquake	19:33:06	✓						6.00
earthquake	19:34:00		✓	✓	✓	✓		0.80
quake	19:34:31				✓			2.00
earthquake	19:34:41				✓			2.00
quake	19:34:49	✓						1.00
quake	19:35:17	✓	✓		✓	✓		1.00
earthquake	19:35:40		✓		✓	✓		1.40
earthquake	19:36:08	✓	✓	✓	✓	✓	✓	2.10
quake	19:36:32	✓	✓					5.00
earthquake	19:37:08	✓	✓					10.00
earthquake	19:37:47	✓						17.00
rockfall	19:38:12				✓			3.00
rockfall	19:38:36				✓			4.00
earthquake	19:39:11	✓	✓	✓	✓	✓	✓	2.10

– May 30, 2025 –

Appendix A. Catalogue of Larisa data

Table A.3: Detection results of seismic events (Larisa Dataset, Part 3/3)

Class	Time	Detected						Duration (s)
		A1	M1	I1	A2	M2	I2	
quake	19:39:19	✓						3.00
rockfall	19:39:27	✓			✓			3.00
earthquake	19:39:41		✓		✓	✓	✓	1.20
earthquake	19:40:00	✓	✓					7.00
earthquake	19:40:17	✓	✓	✓	✓	✓	✓	0.40
earthquake	19:40:33				✓			1.00
earthquake	19:40:41	✓	✓	✓	✓	✓	✓	1.70
earthquake	19:41:29	✓			✓	✓		0.50
rockfall	19:42:07	✓						3.00
earthquake	19:42:20	✓						8.00
earthquake	19:42:44	✓	✓		✓	✓	✓	1.10
earthquake	19:43:10	✓	✓		✓			1.00
rockfall	19:43:41				✓			1.00
earthquake	19:43:51	✓	✓		✓	✓		1.00
rockfall	19:44:00				✓			2.00
quake	19:45:24	✓	✓		✓	✓	✓	0.40
earthquake	19:46:12	✓	✓	✓	✓	✓	✓	1.30
earthquake	19:46:40	✓	✓	✓	✓	✓	✓	1.80
earthquake	19:47:12	✓	✓		✓	✓		0.70
earthquake	19:47:22	✓	✓			✓		2.30
earthquake	19:47:33	✓	✓	✓	✓	✓	✓	1.30
quake	19:47:55	✓	✓		✓	✓	✓	0.50
earthquake	19:48:04	✓	✓		✓	✓		1.80
rockfall	19:48:25	✓	✓		✓			14.00
earthquake	19:48:45	✓	✓	✓		✓	✓	0.60
quake	19:49:03	✓	✓					4.00
rockfall	19:49:17	✓	✓					22.00
earthquake	19:49:41	✓			✓			3.00
rockfall	19:49:52	✓	✓	✓	✓	✓	✓	2.20
rockfall	19:50:14	✓	✓		✓	✓	✓	0.40
earthquake	19:50:32	✓	✓		✓	✓	✓	0.50
rockfall	19:50:42				✓			1.00
earthquake	19:51:12	✓	✓	✓	✓	✓	✓	1.80
earthquake	19:51:39				✓		✓	0.50
earthquake	19:51:57	✓	✓					9.00
quake	19:52:11	✓	✓			✓		0.60
earthquake	19:52:23	✓	✓	✓	✓	✓	✓	0.40
earthquake	19:52:36	✓	✓		✓	✓		0.40
earthquake	19:52:51	✓	✓			✓		0.60
earthquake	19:53:11	✓	✓					6.00

– May 30, 2025 –

## Appendix B

# Detailed Results and Confusion Matrices for Self-Supervised Learning Downstream Tasks

### B.1 K-means clustering

Detailed results and confusion matrices for K-means clustering using the original 256-D feature vector are presented. Additionally, results for t-Distributed Stochastic Neighbor Embedding (t-SNE) in both 2-D and 3-D, as well as Principal Component Analysis (PCA) with dimensions reduced to 2-D, 3-D, and 10-D (explaining 95.3% of the variance), are included.

Table B.1: t-SNE 2-D Classification Performance

	Precision	Recall	F1-score
Quake	0.4198	0.6709	0.5164
Earthquake	0.8289	0.6478	0.7273
Rockfall	0.9511	0.8234	0.8827
N/A Noise	0.7163	0.7143	0.7153

Appendix B. Detailed Results and Confusion Matrices for Self-Supervised Learning Downstream Tasks

Table B.2: t-SNE 2-D Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>157</b>	6	3	68
Earthquake	113	<b>252</b>	6	18
Rockfall	21	37	<b>331</b>	13
N/A Noise	83	9	8	<b>250</b>

Table B.3: t-SNE 3-D Classification Performance

	Precision	Recall	F1-score
Quake	0.4286	0.6667	0.5217
Earthquake	0.8546	0.6195	0.7160
Rockfall	0.8772	0.8532	0.8651
N/A Noise	0.7396	0.7143	0.7267

Table B.4: t-SNE 3-D Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>156</b>	8	4	66
Earthquake	110	<b>241</b>	27	11
Rockfall	24	24	<b>343</b>	11
N/A Noise	74	9	17	<b>250</b>

Table B.5: Original Dimension (256) Classification Performance

	Precision	Recall	F1-score
Quake	0.3790	0.6624	0.4821
Earthquake	0.8415	0.5321	0.6520
Rockfall	0.8746	0.7463	0.8054
N/A Noise	0.6923	0.7457	0.7180

Table B.6: Original Dimension (256) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>155</b>	4	2	73
Earthquake	133	<b>207</b>	21	28
Rockfall	55	32	<b>300</b>	15
N/A Noise	66	3	20	<b>261</b>

Table B.7: PCA 2-D Classification Performance

	Precision	Recall	F1-score
Quake	0.3902	0.5085	0.4461
Earthquake	0.8043	0.6350	0.7099
Rockfall	0.7974	0.7537	0.7750
N/A Noise	0.6918	0.7143	0.7029

Appendix B. Detailed Results and Confusion Matrices for Self-Supervised Learning Downstream Tasks

Table B.8: PCA 2-D Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>119</b>	60	6	49
Earthquake	46	<b>247</b>	14	82
Rockfall	49	31	<b>302</b>	20
N/A Noise	36	17	10	<b>287</b>

Table B.9: PCA 3-D (Explained variance: 79.59%) Classification Performance

	Precision	Recall	F1-score
Quake	0.3717	0.6624	0.4786
Earthquake	0.8410	0.5167	0.6404
Rockfall	0.8746	0.7512	0.8084
N/A Noise	0.6907	0.7400	0.7149

Table B.10: PCA 3-D (Explained variance: 79.59%) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>155</b>	4	2	73
Earthquake	133	<b>207</b>	21	28
Rockfall	56	32	<b>302</b>	12
N/A Noise	69	3	14	<b>264</b>

Table B.11: PCA 10-D (Explained variance: 95.3%) Classification Performance

	Precision	Recall	F1-score
Quake	0.3790	0.6624	0.4821
Earthquake	0.8415	0.5321	0.6520
Rockfall	0.8746	0.7463	0.8054
N/A Noise	0.6923	0.7457	0.7180

Table B.12: PCA 10-D (Explained variance: 95.3%) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>155</b>	4	2	73
Earthquake	133	<b>207</b>	21	28
Rockfall	55	32	<b>300</b>	15
N/A Noise	66	3	20	<b>261</b>

## B.2 ANN-based semi-supervised method

Detailed results and confusion matrices for Artificial Neural Network (ANN)-based semi-supervised method using different training set proportions of the entire dataset (5%, 10%, 20%, and 33.3%).

Table B.13: 33.3% (3-fold) Classification Performance

	Precision	Recall	F1-score
Quake	$0.777 \pm 0.024$	$0.774 \pm 0.048$	$0.774 \pm 0.015$
Earthquake	$0.917 \pm 0.025$	$0.920 \pm 0.007$	$0.918 \pm 0.007$
Rockfall	$0.916 \pm 0.032$	$0.906 \pm 0.024$	$0.913 \pm 0.017$
N/A Noise	$0.861 \pm 0.046$	$0.836 \pm 0.051$	$0.845 \pm 0.005$

Table B.14: 33.3% (3-fold) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b><math>120.6 \pm 7.5</math></b>	$6.7 \pm 2.4$	$6.3 \pm 0.5$	$22.3 \pm 9.1$
Earthquake	$8.3 \pm 1.9$	<b><math>238.7 \pm 8.9</math></b>	$9.3 \pm 1.5$	$3.3 \pm 2.1$
Rockfall	$5.6 \pm 1.6$	$6.7 \pm 2.4$	<b><math>249.3 \pm 3.7</math></b>	$3.3 \pm 0.6$
N/A Noise	$20.0 \pm 7.1$	$4.7 \pm 2.1$	$12.7 \pm 4.8$	<b><math>195.0 \pm 8.3</math></b>

Table B.15: 20% (5-fold) Classification Performance

	Precision	Recall	F1-score
Quake	$0.715 \pm 0.047$	$0.646 \pm 0.040$	$0.677 \pm 0.017$
Earthquake	$0.864 \pm 0.020$	$0.900 \pm 0.033$	$0.881 \pm 0.015$
Rockfall	$0.916 \pm 0.027$	$0.909 \pm 0.021$	$0.912 \pm 0.014$
N/A Noise	$0.822 \pm 0.033$	$0.837 \pm 0.037$	$0.828 \pm 0.016$

Table B.16: 20% (5-fold) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b><math>121.0 \pm 7.4</math></b>	$22.8 \pm 2.9$	$6.0 \pm 2.0$	$37.4 \pm 8.7$
Earthquake	$17.2 \pm 5.5$	<b><math>280.0 \pm 10.3</math></b>	$8.6 \pm 4.1$	$5.4 \pm 1.85$
Rockfall	$6.6 \pm 3.6$	$14.0 \pm 4.3$	<b><math>292.2 \pm 6.7</math></b>	$8.8 \pm 4.7$
N/A Noise	$25.6 \pm 14.2$	$7.6 \pm 4.2$	$12.4 \pm 5.8$	<b><math>234.4 \pm 10.4</math></b>

## B.3 SOM models

Detailed results and confusion matrices for using different Self-Organizing Map (SOM) models each requiring different amount of labelled data (shown in the brackets).

Appendix B. Detailed Results and Confusion Matrices for Self-Supervised Learning Downstream Tasks

Table B.17: 10% (10-fold) Classification Performance

	Precision	Recall	F1-score
Quake	$0.615 \pm 0.077$	$0.432 \pm 0.093$	$0.500 \pm 0.073$
Earthquake	$0.765 \pm 0.067$	$0.828 \pm 0.046$	$0.792 \pm 0.036$
Rockfall	$0.881 \pm 0.037$	$0.882 \pm 0.015$	$0.882 \pm 0.010$
N/A Noise	$0.763 \pm 0.042$	$0.821 \pm 0.049$	$0.788 \pm 0.018$

Table B.18: 10% (10-fold) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>90.9</b> $\pm$ 19.6	52.4 $\pm$ 14.2	9.4 $\pm$ 5.2	57.9 $\pm$ 16.3
Earthquake	27.2 $\pm$ 9.5	<b>289.0</b> $\pm$ 15.9	18.9 $\pm$ 8.2	18.1 $\pm$ 3.8
Rockfall	5.6 $\pm$ 2.9	26.6 $\pm$ 8.4	<b>319.9</b> $\pm$ 8.7	10.3 $\pm$ 5.3
N/A Noise	26.3 $\pm$ 12.4	14.2 $\pm$ 8.2	16.6 $\pm$ 7.9	<b>258.5</b> $\pm$ 15.5

Table B.19: 5% (20-fold) Classification Performance

	Precision	Recall	F1-score
Quake	$0.536 \pm 0.074$	$0.472 \pm 0.106$	$0.496 \pm 0.075$
Earthquake	$0.783 \pm 0.064$	$0.768 \pm 0.064$	$0.771 \pm 0.060$
Rockfall	$0.864 \pm 0.042$	$0.882 \pm 0.029$	$0.873 \pm 0.025$
N/A Noise	$0.774 \pm 0.021$	$0.796 \pm 0.050$	$0.787 \pm 0.018$

Table B.20: 5% (20-fold) Confusion Matrix

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>105.0</b> $\pm$ 25.4	46.0 $\pm$ 23.5	10.9 $\pm$ 5.2	60.4 $\pm$ 30.4
Earthquake	49.0 $\pm$ 25.1	<b>283.0</b> $\pm$ 23.5	23.1 $\pm$ 14.9	13.6 $\pm$ 9.2
Rockfall	6.7 $\pm$ 5.4	25.5 $\pm$ 18.4	<b>337.1</b> $\pm$ 16.6	12.5 $\pm$ 5.5
N/A Noise	15.3 $\pm$ 7.5	11.6 $\pm$ 5.2	15.9 $\pm$ 7.7	<b>254.6</b> $\pm$ 16.6

Table B.21: Classification performance results for SOM with 8x8 nodes (4.6%).

	Precision	Recall	F1-score
Quake	0.6589	0.6026	0.6295
Earthquake	0.8532	0.7918	0.8213
Rockfall	0.8878	0.9055	0.8966
N/A Noise	0.7282	0.8114	0.7676

Table B.22: Confusion matrix for SOM with 8x8 nodes (4.6%).

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>141</b>	24	7	62
Earthquake	29	<b>308</b>	24	28
Rockfall	6	16	<b>364</b>	16
N/A Noise	38	13	15	<b>284</b>

Appendix B. Detailed Results and Confusion Matrices for Self-Supervised Learning Downstream Tasks

Table B.23: Classification performance results for SOM with 12x12 nodes (10.4%).

	Precision	Recall	F1-score
Quake	0.6567	0.6538	0.6552
Earthquake	0.8278	0.8406	0.8342
Rockfall	0.9406	0.9055	0.9227
N/A Noise	0.7944	0.8171	0.8056

Table B.24: Confusion matrix for SOM with 12x12 nodes (10.4%).

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>153</b>	36	3	42
Earthquake	32	<b>327</b>	11	19
Rockfall	4	21	<b>364</b>	13
N/A Noise	44	11	9	<b>286</b>

Table B.25: Classification performance results for SOM with 17x17 nodes (21.0%).

	Precision	Recall	F1-score
Quake	0.7393	0.6667	0.7011
Earthquake	0.8314	0.9126	0.8701
Rockfall	0.9268	0.9129	0.9198
N/A Noise	0.8299	0.8086	0.8191

Table B.26: Confusion matrix for SOM with 17x17 nodes (21.0%).

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>156</b>	35	5	38
Earthquake	12	<b>355</b>	11	11
Rockfall	9	17	<b>367</b>	9
N/A Noise	34	20	13	<b>283</b>

Table B.27: Classification performance results for SOM with 21x21 nodes (32.1%).

	Precision	Recall	F1-score
Quake	0.7650	0.6538	0.7051
Earthquake	0.8596	0.9126	0.8853
Rockfall	0.9279	0.9279	0.9279
N/A Noise	0.8111	0.8343	0.8226

Table B.28: Confusion matrix for SOM with 21x21 nodes (32.1%).

	Quake	Earthquake	Rockfall	N/A Noise
Quake	<b>153</b>	34	7	40
Earthquake	8	<b>355</b>	7	19
Rockfall	5	15	<b>373</b>	9
N/A Noise	34	9	15	<b>292</b>

Appendix B. Detailed Results and Confusion Matrices for Self-Supervised Learning  
Downstream Tasks

– May 30, 2025 –