UNIVERSITY OF PALERMO
PHD JOINT PROGRAM:
UNIVERSITY OF CATANIA - UNIVERSITY OF MESSINA
XXXVI CYCLE

DOCTORAL THESIS

# Computational Models for the Analysis of seismo-volcanic traces

*Author:*
ANDREA DI BENEDETTO

*Supervisors:*
Prof. Giosué LO BOSCO
Dr. Antonino D'Alessandro

*Co-Supervisor:*
Dr. Alessandro Amato

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in*

*Mathematics and Computational Sciences*

June 24, 2025

Signed: 

Date: 24/06/2025

*"The Sun sets Forever over Blackwater Park."*

Mikael Lars Akerfeldt, Opeth

UNIVERSITY OF PALERMO

# *Abstract*

Department of Mathematics and Computer Sciences

Doctor of Philosophy

**Computational Models for the Analysis of seismo-volcanic traces**

by ANDREA DI BENEDETTO

This PhD thesis investigates computational models for the detection and picking of seismo-volcanic events from continuous waveforms. The research explores three complementary approaches: (1) a Grid-search method for optimizing Short Term Average over Long Term Average (STA/LTA) parameters, (2) an Active Learning (AL) strategy combining machine learning and human intervention to enhance classification accuracy, and (3) the application of a deep learning model, Earthquake Transformer (EQTransformer), based on transformer architectures, for local and regional seismic event detection. In the volcanic domain, the Grid-search method was applied to optimize the STA/LTA parameters for the detection of Explosion Quakes (EQs) at Stromboli volcano by using the overall measure proposed: Quality Numerosity index (QNi). The results demonstrated a substantial improvement in detection precision and consistency over standard parameter settings, with the optimized configuration increasing the QNi by more 0.24 on test data. The Active Learning methodology, using a Random Forest classifier, showed that human-guided relabeling and signal editing significantly improved the model's classification performance, particularly in low-confidence events and hybrid signal scenarios. These enhancements enabled better discrimination between EQs and non-EQs and reduced the need for operator intervention over time. For tectonic seismicity, the EQTransformer model was tested both on local seismicity in the Groningen region (Netherlands) and on a regional dataset from the 2023 seismic sequence in Turkey. The model achieved high accuracy (up to 98%) in event detection and phase picking, even on previously unseen data. Evaluation using Quality and Numerosity indices (Qi, Ni, QNi) confirmed that the deep learning model performed robustly across diverse seismic contexts, supporting its generalizability. Overall, this thesis demonstrates how the integration of traditional signal processing, machine learning, and deep learning can enhance the reliability of seismic and volcanic monitoring systems. The proposed methodologies are adaptable and scalable, offering valuable tools for real-time monitoring and early warning applications in both volcanic and seismic hazard assessment.

# *Acknowledgements*

The first acknowledgment goes to my supervisor, Prof. Giosue' Lo Bosco, who gave me this great opportunity to demonstrate that I can give some contribution to the scientific world and try to undertake this career in the future. To my co-supervisors, Dr. Antonino D'Alessandro and Dr. Alessandro Amato, who gave me a great support during the path, both in scientific and human terms.
To the PhD coordinator, Prof. Mariacarmela Lombardo, through whom she was always available for any of my requests during my path.
To Dr. Luca Trani, my supervisor during the research period in Utrecht, who taught me a lot about research, how the scientific method works and also for being a friend during my first time abroad experience alone.
To Dr. Anna Figlioli, friend and colleague, with whom we have faced most of our PhD paths together.
I also want to thank the ONT team in Palermo: Dr. Giovanni Vitale, Dr. Salvatore Scudero and Mr. Stefano Speciale, who supported me during the path and introduced me to the world of INGV.
Finally, I want to thank the CAT Team in Rome for supporting me during this journey, with particular reference to Fabrizio Bernardi for all the suggestions.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **AI** | **A**rtificial **I**ntelligence |
| **DL** | **D**eep **L**earning |
| **EEW** | **E**arthquake **E**arly **W**arning |
| **INGV** | **I**stituto **N**azionale di **G**eofisica e **V**ulcanologia |
| **STA** | **S**hort **T**erm **A**verage |
| **LTA** | **L**ong **T**erm **A**verage |
| **CF** | **C**haracteristic **F**unction |
| **Qi** | **Q**uality **i**ndex |
| **Ni** | **N**umerosity **i**ndex |
| **QNi** | **Q**uality **N**umerosity **i**ndex |
| **EQs** | **E**xplosion **Q**uakes |
| **NEQs** | **N**on **E**xplosion **Q**uakes |
| **LP** | **L**ong **P**eriod |
| **VLP** | **V**ery **L**ong **P**eriod |
| **AL** | **A**ctive **L**earning |
| **RF** | **R**andom **F**orest |
| **EQTransformer** | **E**arthquake Transformer |
| **ML** | **M**achine **L**earning |

*To myself and to music, which has been a guiding light throughout this Journey…*

# Chapter 1

# Introduction

The research project has the general objective of deepening and providing contributions to the use of Machine Learning systems for the creation of automatic recognition methods of seismo-volcanic and seismo-tectonic events. In this PhD thesis, it will be described the machine learning and deep learning methodologies and the case studies where the specific aim is the recognition of seismic and volcanic events on seismograms, starting from raw data. The methodological approach starts from the assumption of having data coming from seismic stations installed on site, which once extracted are used to train appropriate models with the aim to recognize seismic events and precursors of an effusive (or explosive) volcanic eruption, such as Explosion Quakes (EQs). To perform an analysis of this type of data, a tool has been developed in Python programming language for the analysis and manipulation of seismo-volcanic waveforms. This tool (described in detail in 3.3) was then used for the applications of the grid-search approach (Chapter 2) and the Active Learning application (Chapter 3).

The grid-search system has been developed to systematically identify the optimal parameter combination for the STA/LTA method (Allen, 1978), used to picking seismic phases onset. This approach is used to extract volcanic events — such as EQs — with the goal of creating a large, consistent dataset for machine learning applications. The grid-search approach and the evaluation measures used within it, were then reused to evaluate the seismic waveform pickings of the Deep Learning model described in chapter 4; demonstrating that this application (Di Benedetto et al., 2024b) can also be used in different contexts (seismo-volcanic and seismo-tectonic).
The other contribution, namely the application related to the topic of Active Learning, gave an improvement in terms of detection of seismo-volcanic events with the help of a human operator on a subset of the input dataset (D'Alessandro et al., 2022). Human operator review of a subset of volcanic events (or other types of events) that are detected with an accuracy below a set threshold (for instance, 90%) can improve the picker's performance since these events are re-proposed to the expert operator for reanalysis; allowing the operator to accurately crop the data and relabel it, in order to enrich the initial dataset.

For the earthquake analysis component, a deep learning model based on the transformer architecture, namely, EQTransformer (Mousavi et al., 2020), was employed in Chapter 4. This model was applied in two complementary approaches (local and regional), both aimed at detecting and picking seismic events. The local approach regards the training and testing EQTransformer for detecting and picking seismic events in the northern netherlands region (Groningen region), where the seismicity is rich in low-magnitudes events (Di Benedetto et al., 2024a). The regional

approach, on the other hand, deals with seismic events covering a wider magnitude range and focuses on the geographical area of Turkey. This area was chosen because in February 2023 a seismic event with magnitude 7.9 started a seismic sequence rich in events with magnitudes that exceed those that were considered in the paper of EQTransformer.

As a resume of the topics and the applications here proposed, following is the list of the thesis chapters:

- Chapter 1: Introduction, this chapter.
- Chapter 2: Grid-search method and application
- Chapter 3: Active Learning
- Chapter 4: Deep Learning: Transformer
- Chapter 5: Discussion
- Chapter 6: Conclusions

## 1.1 Picking on seismic and volcano signals

Seismic and volcanic signals provide crucial insights into Earth's internal dynamics, helping scientists monitor and predict natural hazards such as earthquakes and volcanic eruptions. One of the fundamental tasks in seismology and volcanology is signal picking—the process of detecting and identifying key phases within seismic waveforms. Accurate picking of seismic signals, including P- and S-wave arrivals, volcanic tremors, and other geophysical anomalies, is essential for determining event locations, analyzing subsurface structures, and assessing volcanic activity.

Traditional methods for seismic picking rely on manual interpretation or classical algorithms, such as short-term average to long-term average (STA/LTA) ratio detection (Maggi et al., 2009; Withers et al., 1998; Bai and Kennett, 2001). However, these approaches can be time-consuming and sometimes lack precision, especially in noisy environments or complex volcanic settings. Recent advances in machine learning and automated signal processing have revolutionized the field, enabling more efficient and accurate automatic detection of seismic and volcanic events. This study explores different approaches to seismic and volcanic signal picking, comparing traditional and modern techniques to improve the accuracy and efficiency of event detection. By enhancing our ability to analyze seismic signals, we contribute to better hazard assessment and early warning systems, ultimately improving disaster preparedness and response strategies. The picking of seismic and volcanic signals leads to significant differences, described in the next paragraph.

### 1.1.1 Differences Between Picking Seismic and Volcanic Signals

The picking of seismic and volcanic signals involves detecting and identifying key phases within waveforms, but the characteristics of these signals and the challenges associated with their analysis differ significantly due to the nature of the sources and the environments in which they occur. Below are some key differences:

1. **Signal Characteristics**

- *Seismic Signals*
  - Typically consist of distinct, impulsive arrivals corresponding to different seismic phases, such as P-waves (primary waves) and S-waves (secondary waves).
  - Generated by sudden ruptures in the Earth's crust, such as earthquakes, explosions, or tectonic movements.
  - More predictable in terms of wave propagation and velocity models starting from the seismic source.

- *Volcanic Signals*
  - Often characterized by emergent, continuous, or harmonic waveforms rather than sharp, impulsive arrivals.
  - Include volcanic tremors, long-period (LP) events, very long period (VLP) signals, and explosion signals.
  - More complex and variable due to the influence of magmatic movement, gas emissions, and changes in the volcanic conduit.

2. **Picking Challenges**

- *Seismic Picking*
  - Well-established algorithms such as STA/LTA, Akaike Information Criterion (AIC), and machine learning approaches are commonly used.
  - P- and S-wave arrivals can often be detected with quite high precision, given their sharp onsets.
  - Noise can interfere, especially in urban areas or near industrial activity, but signal processing techniques help mitigate this.

- *Volcanic Picking*
  - More challenging due to the continuous and emergent nature of volcanic tremors.
  - Traditional picking methods are less effective, requiring alternative approaches such as spectrogram analysis, wavelet transforms, or machine learning-based classifications.
  - Often needs a multidisciplinary approach, combining different sensors (e.g., seismic, infrasound, gas, and thermal data) for a comprehensive interpretation.

3. **Frequency and Duration**

- *Seismic Signals*
  - Typically short-duration (seconds to minutes) with a broad frequency range, depending on the earthquake magnitude and depth.
  - High-frequency components are often dominant, especially for local and regional events.

- *Volcanic Signals*
  - Can last from seconds to days (e.g., volcanic tremors), making it difficult to define clear onset and termination points.
  - Often dominated by low-frequency components, especially for LP and VLP events.

4. **Interpretation and Applications**

   - *Seismic Signal Picking*
     - Used for earthquake early warning, hypocenter determination, seismic hazard assessment, and tectonic studies.
     - Helps in identifying faults, aftershock sequences, and stress distribution in the Earth's crust.

   - *Volcanic Signal Picking*
     - Primarily used for monitoring volcanic activity, forecasting eruptions, and understanding magma dynamics.
     - Can indicate changes in volcanic systems, such as pressurization, gas movement, or conduit blockages.

Advances in signal processing and artificial intelligence are helping to improve the accuracy and efficiency of both fields, leading to better hazard monitoring and early warning systems. There are many scientific works that have addressed this picking challenge, summarized below.

### 1.1.2 State of the Art in AI techniques for picking seismo-volcanic events

While both seismic and volcanic signal picking aim to identify key waveform features, the differences in signal characteristics, picking methods, and interpretation make volcanic signal analysis more complex. Seismic and volcanic signal picking is a critical component in monitoring Earth's internal processes, enabling the detection and analysis of earthquakes and volcanic activities. The nature of these signals varies significantly, necessitating distinct approaches for effective picking. Recent studies have explored both traditional and modern methodologies, highlighting their respective advantages and limitations. In the field of seismic signal picking, traditional methods have long been employed. As early as the 1970s, seismologists developed automatic phase-picking algorithms based on single-trace waveform data to cope with the increasing volume of seismic data, reducing the reliance on manual picking.

In recent years, deep learning has emerged as a powerful tool in this domain. For instance, in 2018, a deep-neural-network-based method called PhaseNet (Zhu and Beroza, 2018) was introduced with the purpose of seismic arrival time picking. This approach utilizes three-component seismic waveforms to generate probability distributions of P and S wave arrivals, achieving higher picking accuracy and recall rates compared to existing methods. Similarly, Wang et al., 2024 proposed the Multi-Stage Segmentation Picking Network (MSSPN), an automatic first arrival picking method that addresses challenges in low signal-to-noise ratio scenarios. Their approach simulates the manual picking process through a multi-stage network, demonstrating improved generalization across different work sites. Zhu et al., 2022 introduced an end-to-end earthquake detection method that jointly optimizes phase picking and event detection using deep learning. This approach integrates a neural network architecture capable of processing seismic waveforms from multiple stations, enhancing the overall detection accuracy. The model includes sub-networks for feature extraction, phase picking, and event detection, utilizing a shift-and-stack module based on back-projection to generalize across different velocity models and station geometries. While effective, this method requires substantial computational

resources and may face challenges in regions with sparse seismic networks. Münchmeyer et al., 2022 conducted a comprehensive evaluation of six deep learning models across eight datasets, encompassing both local and teleseismic distances. Their findings indicate that models such as EQTransformer, Generalized Seismic Phase Detection (GPD, Ross et al., 2018), and PhaseNet outperform classical methods in terms of accuracy. However, the study also revealed that these models experience performance degradation when applied to data with characteristics differing from their training sets, particularly when transferring between regional and teleseismic data. In the context of volcanic signal picking, the continuous and emergent nature of volcanic tremors presents unique challenges. Fenner et al., 2021 developed AWESAM, a Python module for automated volcanic event detection applied to Stromboli volcano. This tool identifies potential volcanic events by analyzing ground-velocity amplitudes and employs an adaptive filtering approach, achieving a detection rate of approximately 95% for eruptions with a signal-to-noise ratio above three. Furthermore, Yoma et al., 2021 introduced an end-to-end Long Short-Term Memory (LSTM) based scheme for volcano event epicenter localization. This method addresses the inaccuracies associated with automatic phase picking in volcanic signals by modeling the temporal dynamics of the data, resulting in a success rate significantly superior to traditional approaches.

In summary, deep learning-based techniques have demonstrated significant potential in enhancing the accuracy of seismic signal picking, though they often necessitate extensive training data and may struggle with generalization across diverse datasets. For volcanic signal picking, automated methods like AWESAM offer promising results, yet their effectiveness can be influenced by data quality and the specific characteristics of the volcanic environment.

# Chapter 2

# Grid-search method and application

## 2.1 Grid-search: State of the art

Grid Search is a widely used technique in machine learning and algorithm parameter optimisation. It is used to search for the optimal combination of parameters for a model or algorithm while varying multiple parameters simultaneously. In Machine Learning, grid-search is used as an exhaustive exploration of the hyperparameter combinations in order to find the one who fits at the best with the model considered. A basic diagram that illustrates the core idea behind grid-search for hyperparameter tuning is proposed in Figure 2.1.

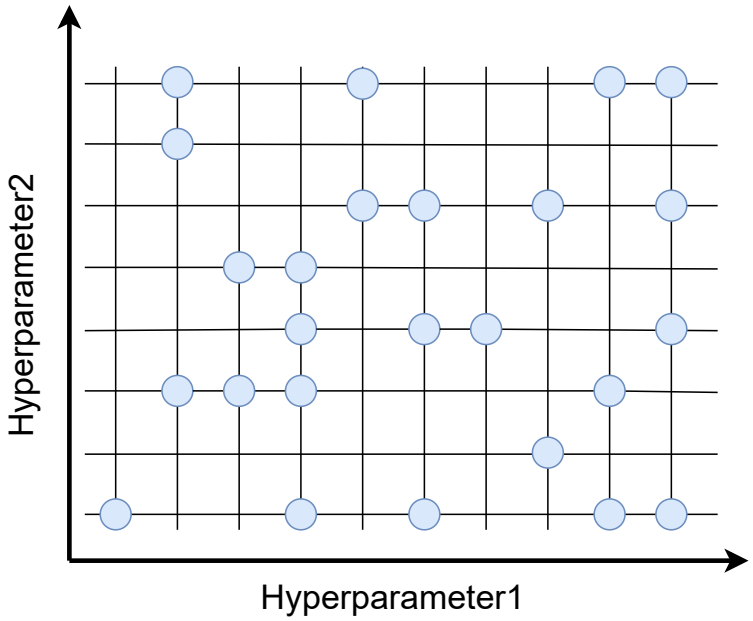



FIGURE 2.1: Illustration of a general grid search diagram: each point on the grid represents a unique combination of the two hyperparameters. The model is trained and evaluated at each point on the grid to identify the best configuration in terms of model performance

Each axis on the plot represents a different hyperparameter (e.g., learning rate, regularization strength), and the grid points correspond to specific combinations of these ones. At each point in the grid, a model is trained and evaluated. The goal is

to identify the combination that provides the best performance, usually based on a validation metric. It is a brute-force but effective method to systematically explore the hyperparameter space.
For instance, Belete and D H, 2021 use the grid-search among several ML models to identify the optimal hyperparameter values that produce the best-performing model, while accounting for uncertain training computing costs, and evaluate the predictive performance of the models; the aim of this work is to predict the outcomes of HIV tests from the Ethiopian Demographic and Health Survey (EDHS), HIV/AIDS dataset. Another work that resume some of the most efficient works that performs at best in the Hyperparameter Optimization (HPO) is the one from Bischl et al., 2023; for instance, in reinforcement learning and deep learning topic, the dynamic configuration approach adopted by Li et al., 2019 use Population-Based Training (PBT), enabling hyperparameter optimization during training. PBT maintains multiple training runs with varied settings and periodically updates them using evolutionary strategies such as mutation and selection. The disadvantage is that this approach demands significant computational power, as it involves training multiple models simultaneously. Furthermore, long HPO runs can produce biased performance estimates, potentially leading to weak hyperparameter configurations (HPC), especially in presence of small datasets or few resampling iterations. Using additional validation data or data augmentation might help, but it reduces data efficiency. A smarter, dynamic resampling strategy during HPO could be a better solution, though it's still an underexplored area.

## 2.2 Short-Term Average over Long-Term Average

The analysis of seismological data is crucial for understanding and monitoring seismo-volcanic activity, and the Short Term Average/Long Term Average (STA/LTA) method (Allen, 1978) is one of the most widely used techniques for detecting seismic events within continuous waveform data. STA/LTA is a ratio-based approach that compares the average signal amplitude over a short time window (STA) with the one over a longer time window (LTA) providing an efficient way to discriminate between seismic events and background noise in seismogram data. The short-term window typically spans a few seconds, capturing the immediate variations in the signal caused by seismic waves. The long-term window, on the other hand, is usually several times longer, capturing the overall background noise level. When a seismic event occurs, the short-term window will capture the strong amplitudes associated with the event, causing a significant increase in the STA/LTA ratio. This increase is identified using detections threshold defined by the user. This method is particularly effective in identifying the onset of seismic events, such as volcanic tremors, explosions, and microseismicity, by highlighting abrupt changes in amplitude that indicate the start of an event. The two windows STA and LTA are applied to the "character" of the signal by a function called Characteristic Function (CF) $E_k$. There are different CF that can be used; here, the classic one from Allen, 1978 is described as:

$$E_k = x_k^2 + (x_k')^2 + C_k \tag{2.1}$$

where, $x_k$ is the seismic trace, $x_k'$ its derivative and $C_k$ is an empirical weighting constant described as:

$$C_k = \frac{\sum_{j=1}^{k} |x_j|}{\sum_{j=1}^{k} |x_j - x_{j-1}|} \tag{2.2}$$

The use of the derivative allows highlighting abrupt frequency variations in the signal, making the CF useful for detecting abrupt changes in the signal, as it is shown in the Figure 2.2. In this figure, the most informative content occurs between the second ~460 and ~1000.

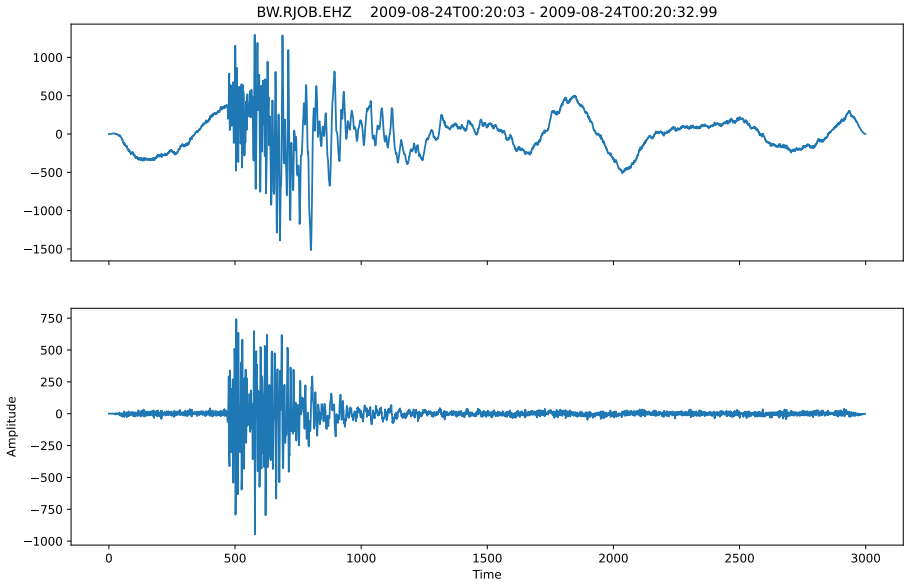


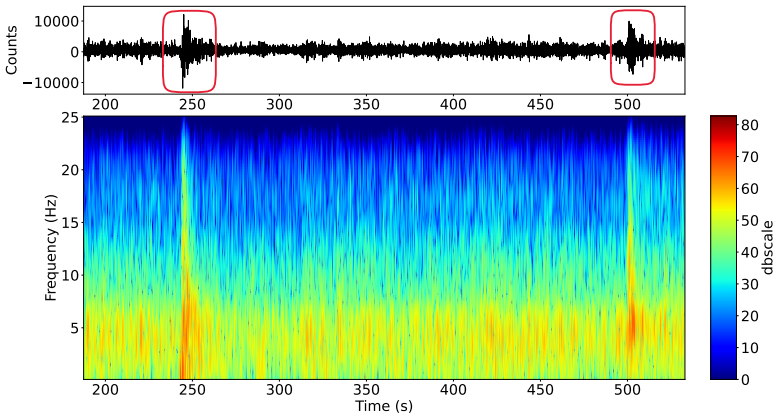

FIGURE 2.2: Example of a plot showing the seismic trace (top) and its corresponding derivative (bottom)

Instead, the role of the constant $C_k$ is to remove a specific offset that may cause amplitude differences between signals belonging to the same seismic event.

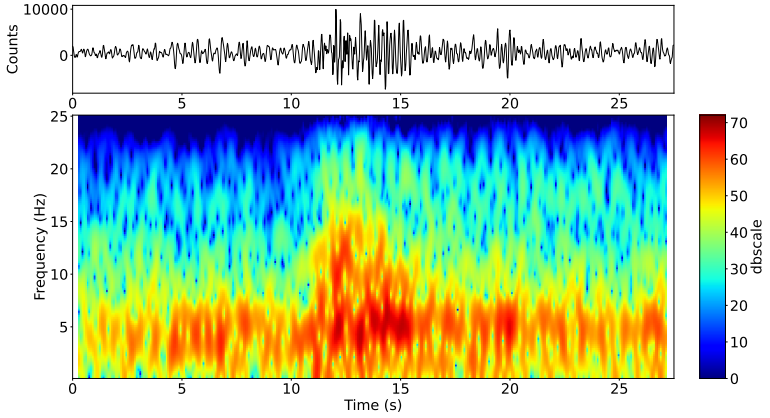
## 2.3 Grid-search method for STA/LTA parameters selection

STA/LTA is a parametric approach where two basic parameters are STA and LTA window lengths expressed in seconds. The ratio between the two windows defines a quantity that is controlled through two other parameters, i.e. Trigger on and Trigger off. These parameters determine the start and end points of a detected event, enabling detailed analysis of the detection's acacuracy. The selection of these parameters can be an iterative process based on real data analysis, where one must remember to continuously monitor the performance of this detection system and make the necessary updates to adapt it to the evolution of the observed phenomenon. The application of the grid-search here proposed is an exploratory research conducted to find the combination of parameters of STA/LTA method that automatically cut at best EQs, compared to the cuts made by the expert operator. As an example of events present in the dataset, Figure 2.3a shows the case of two EQs in raw signal, with a zoom on the last one (Figure 2.3b). A Spectrogram is computed from the raw signal to represent its frequency content, thereby facilitating the identification of significant events by the analyst. To extract it, a short-time Fourier transform was

calculated using 0.5 s sliding time windows with 90% time overlapping. Figure 2.4 shows the STA/LTA ratio (bottom) calculated on zoomed EQ of Figure 2.3 and the triggers on the signal (top): the red bar consists of Trigger on threshold and the Blue bar on Trigger off threshold. When the slope of the curve exceeds the value of the trigger threshold, both for trigger on and trigger off, the red and blue bars are applied on the plot, respectively. The presented tool to analyse continuous waveforms and manually cut EQs (D'Alessandro et al., 2022) is the same as the one used for the Active Learning application in section 3.3.




(a)




(b)

FIGURE 2.3: This figure presents a representative case of Explosion Quakes observed in raw seismo-volcanic signals. Each subfigure consists of two components: the raw signal displayed at the top and the corresponding spectrogram at the bottom. The raw signal is plotted with time on the x-axis, while the spectrogram illustrates the representation of this raw signal in the frequency-time domain. The side color bar indicates the energy of the signal measured in decibels; useful to precisely detect significant events in raw signal. In subfigure (a), multiple Explosion Quakes from the Stromboli volcano's signal are marked with red circles. In subfigure (b), a detailed view of the second explosion quake is provided.

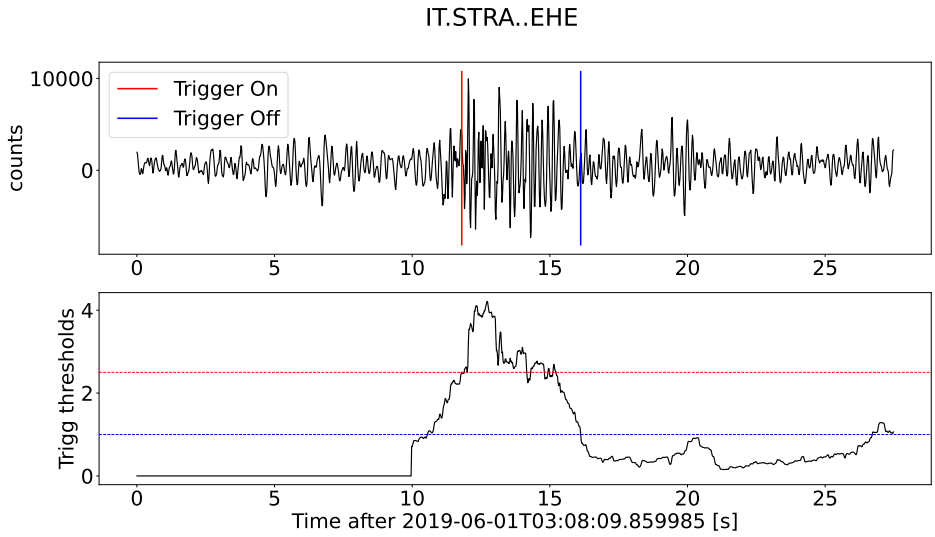



FIGURE 2.4: STA/LTA performed on the zoomed EQ of the Figure 2.3. Red Bar is Trigger On and Blue Bar is Trigger Off. The figure above illustrates the raw seismo-volcanic signal, while the figure below represents time along the x-axis and trigger threshold values along the y-axis. Precisely, on the y-axis, the amplitude of the STA/LTA ratio is shown, but, trigger threshold values are used to show when the slope of the curve exceeds them. In this example case, the chosen values of the parameters for this example were STA 1 s, LTA 10 s, Trigger On 2.5, Trigger Off 1.

In this application, as a first approach, the CF $E_k$ described in equation 2.1 is used for STA/LTA method. The underlying concept of this grid-search application is based on the start and end times outputs from the STA/LTA method and associated with a detected event. The quality of start and end times suggested by this application can be evaluated by comparing them to the selections performed by an expert operator and establishing a specific misfit measure. This measure is called Quality-Numerosity index (*qni*) and combines two preparatory measures: the Quality index (*qi*) and the Numerosity index (*ni*). *qni* is calculated as the product of these two measures, providing an overall measure (see Paragraph 2.3.1 for details on quality and numerosity indices). In this comparison, the absolute deviation in terms of temporal distance is computed. If this value does not exceed a certain residual value $k$, both for the start time and the end time of the event, then the cut is deemeed correct. This comparison enables the checking of the quality of the cut (*qi*) and the numerosity (*ni*) of the triggered events. A list of triggered events can be generated by varying the combination of the following four parameters: STA window size (in seconds), LTA window size (in seconds), Trigger on threshold, Trigger off threshold. From now on, the four parameters will be indicated with the term 'quadruple'. A grid search is performed on a quadruple set, by considering the overall index computed on the related triggered events.

### 2.3.1 Evaluation Measures

*Quality index* of a cut is designated as *qi* and *numerosity index* of a cut as *ni*. These measures were designed to analyse different phenomena represented by time series, in this case, they are used for EQs.
The qi measures the degree of greater temporally precise cut performed by the STA/LTA

compared to one performed by the human operator. Let $m$ be the *mean* of all the temporal deviation computed between the STA/LTA cuts and the operator cuts, and $k$ be the residual value as *threshold* in seconds. The qi is defined as:

$$qi = 1 - (m/k) \tag{2.3}$$

$k$ is an arbitrarily defined constant, dependent on the reference dataset. In this case, a constant has been empirically set with a value of 10 (pre-set value. Reported on the repository published on GitHub, see the repository on GitHub). This constant has been empirically set based on the average temporal uncertainty observed in local seismic datasets between automatic and manual picks. This value reflects a practical normalization window commonly encountered in STA/LTA applications for local events. While it may not generalize to other event types (e.g., teleseismic or anthropogenic sources), it provides a consistent reference scale for the scope of this work. Users applying this $qi$ to different datasets are encouraged to adapt the value of k accordingly.

Every time STA/LTA outputs a list of triggers, a check is performed to see whether the start time (also indicated as $t_{on}$, trigger on) and end time (also indicated as $t_{off}$, trigger off) of every trigger are temporally close to the start and end times of the EQs extracted by the expert. When a match is found (correct cut), the absolute value of the temporal distance is calculated either with the two start times and end times being within $k$, or that trigger is not considered, but it is simply taken into account when counting the triggers for the $ni$. If no match is found, the trigger is not considered, but is still accounted for when counting triggers for $ni$. Figure 2.5 shows a representation of this process.

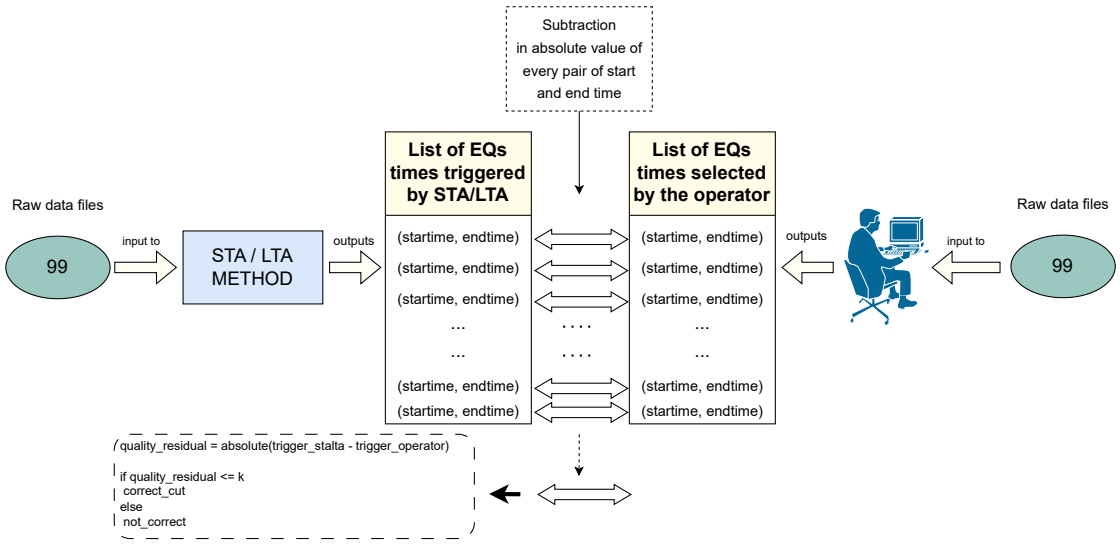



FIGURE 2.5: Selection of the trigger's start time and end time temporally close to the times of the EQs. The raw dataset indicated by the green oval, is the same dataset given as input for both the STA/LTA method and the human operator.

All the computed deviations are stored in a list, the deviations mean $m$ is computed and then normalised by $k$. The $qi$ is finally computed as the complement of the ratio $m/k$ so that it is defined in a range between 0 and 1 where 1 means perfect agreement among the automatic cuts and the expert cuts.

Then, the ni measures the agreement between the number of events triggered by the automatic approach and the number of cuts selected by the human operator.

Let *enq* be the *Experimental EQs*, namely the cardinality of the set of event trigger list produced by the STA/LTA method, and *tnq* be the *Theoretical EQs*, i.e. the number of the cuts performed by the human operator. The ni is thus defined:

$$ni = \begin{cases} enq/tnq, & \text{if } enq < tnq \\ \dfrac{tnq - mod(enq, tnq)}{tnq}, & \text{if } tnq \leq enq < 2 * tnq \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

The *ni* takes into account the discrepancy between *enq* and *tnq*. If *enq* is lower than *tnq*, or *enq* is between *tnq* and twice *tnq*, then the *ni* will result in a number in the range $[0, 1]$. Otherwise, it is set to 0. This latter case occurs when *enq* exceeds *tnq* by at least twice its value, resulting in an *enq*'s value being out of range. The choice of twice is justified by limiting the number of false events triggered by STA/LTA with the quadruple considered.

Finally, the combination of the number of events and their temporal precision selected by the automatic process, compared to human experts, make the *qni* the overall measure. This overall measure determines the effectiveness of the cut made by STA/LTA, and is defined as:

$$qni = qi * ni \tag{2.5}$$

The *qni* ranges between 0 and 1, and can be converted into percentages. These measures are mainly dependent on the results of the STA/LTA method (based on its few parameters) and the expertise of the operator because of the manual cut. The range of window values from STA and LTA and the threshold values for trigger on and trigger off can be determined based on the type of event one wants to detect and, therefore, cut. For instance, if one wants to detect teleseismic events compared to local ones, a wide range of STA and LTA window values must be set, to ensure the expansion of the grid and improve the search for the optimum. In the beginning of the paragraph 2.3.3 an in-depth analysis was conducted in this regard. On the other hand, the work by Jones and Baan, 2015 uses an STA/LTA adaptive method based on the Hidden Markov Model. This method is independent from data, meaning that it requires only minimal configuration by the user. The goal of this methodology is to determine the probability that a term y(t) is an outlier compared to the noise population. The term y(t) corresponds to the CF of a data point from the seismogram x. STA and LTA windows are composed with these probability levels. Using this model, the objective of this work is to detect and select a seismic event. Even though this method is adaptive, it is necessary to determine values for STA and LTA window length and threshold adjustment in the initial state.

### 2.3.2 Grid-search technique

In the present case, a grid-search enables an exhaustive search of the quadruples that correspond to the optimum *qni* values. The grid search is, by definition, an exhaustive method. In this case, the grid search is exhaustive in a local sense rather than a global one. The convergence of the search is based on specific thresholds defined for the evaluation metrics, as detailed later in this paragraph. In Figure 2.6, an example of a representative scheme of the search is proposed.

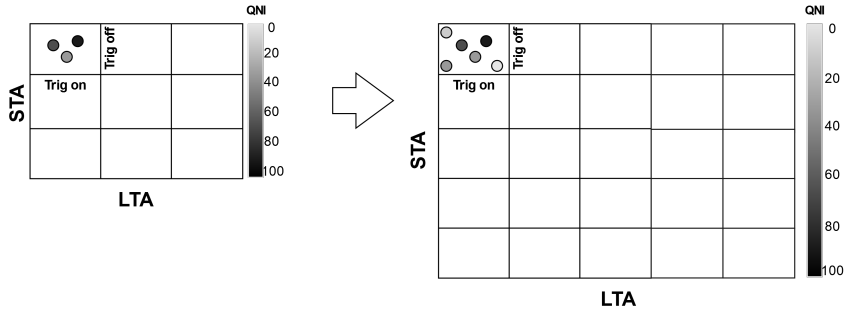



FIGURE 2.6: Grid-search technique for *qni* values

The grid is composed of STA window sizes represented in rows and LTA window sizes represented in columns. Every cell is also a grid, where in abscissa the $t_{on}$ and in ordinate the $t_{off}$ are shown. *qni* values are expressed as a percentage and represented as coloured circles: the darker the colour, the higher the value. We started from a basic grid (grid on the left) with a few quadruples (STA window size (in seconds), LTA window size (in seconds), $t_{on}$, $t_{off}$) and gradually expanded to a grid (grid on the right) still containing the previous grid. The cardinality of the grid is determined by the variation of the quadruples by one unit based on their order of magnitude. As an increase in the value of *qni* is measured, the grid is expanded until the optimum is found. The process is repeated until the *qni* values do not improve further and reach a maximum value. To formalise this concept, a recursive exploratory grid-search algorithm is proposed (Algorithm 1). Choosing to use a recursive algorithm is based on experimenting with all the possible combinations of the quadruples (STA, LTA, $t_{on}$, $t_{off}$) until the best result of *qni* is found. As input, the algorithm requires the range of values to construct all possible quadruples (*get_combinations* function), including a *step* value for each parameter (only one is needed for both trigger thresholds). The *step* determines the numerical distance between one value and its next value. After all possible combinations of quadruples have been calculated, the value of *qni* is determined for each of them through the *compute* function; if this value is greater than the previously calculated *qni*, then it is classified as the *best_qni* value (the quadruple associated with *best_qni* is also stored). The algorithm ends when the *best_qni* found in lines 5-16 exceeds the threshold value; in this case, the *best_quintuple* list is returned. Otherwise, the algorithm is called recursively by subtracting and adding the *step_parameter* (*step_sta*, for instance) associated with the individual parameter quantities. A series of checks are performed for each parameter, to verify that the lower bound of each one is respected. The lower bound is determined by the minimum size of the windows used for the detection, which are usually around 1 s for STA and 5 s for LTA. The upper bound for STA and LTA is determined for the search for local and regional events. This limit can however be varied based on one's needs (types of events sought) (Küperkoch et al., 2010, Gentili and Michelini, 2006, Earle and Shearer, 1994). If no *qni* value exceeds the threshold value, the programme returns the best value of *qni*. In this experiments, 20 iterations were set as thresholds by trial and error.

**Algorithm 1** Pseudocode for recursive exploratory Grid-search algorithm

**procedure** GRID_SEARCH($min\_sta$, $max\_sta$, $step\_sta$, $min\_lta$, $max\_lta$, $step\_lta$,$min\_trig\_on$, $max\_trig\_on$, $step\_trig\_on$, $min\_trig\_off$, $max\_trig\_off$, $step\_trig\_off$, $num\_iterations$)
  $quadruples \leftarrow$ ***get_combinations***($min\_sta$, $max\_sta$, $step\_sta$, $min\_lta$, $max\_lta$, $step\_lta$, $min\_trig\_on$, $max\_trig\_on$, $min\_trig\_off$,$max\_trig\_off$, $step\_trig$)
  $max \leftarrow 0$
  $best\_quintuple \leftarrow []$
  **for all** $quadruple \in quadruples$ **do**
    $qni \leftarrow$ ***compute***($quadruple$) ▷ $qni$ calculation value
    **if** $qni > max$ **then**
      $max \leftarrow qni$
      $best\_sta \leftarrow sta$
      $best\_lta \leftarrow lta$
      $best\_trig\_on \leftarrow trig\_on$
      $best\_trig\_off \leftarrow trig\_off$
      $best\_qni \leftarrow qni$
      $best\_quintuple \leftarrow [best\_sta, best\_lta, best\_trig\_on, best\_trig\_off, best\_qni]$
    **end if**
  **end for**
  **if** $best\_qni > threshold$ *or* $num\_iterations > 20$ **then** ▷ For instance: threshold=80
    **return** $best\_quintuple$
  **else**
    $min\_sta \leftarrow min\_sta - step\_sta$
    **if** $min\_sta < 1$ **then**
      $min\_sta \leftarrow 1$
    **end if**
    $max\_sta \leftarrow max\_sta + step\_sta$
    **if** $max\_sta > 16$ **then**
      $max\_sta \leftarrow 16$
    **end if**
    $min\_lta \leftarrow min\_lta - step\_lta$
    **if** $min\_lta < 5$ **then**
      $min\_lta \leftarrow 5$
    **end if**
    $max\_lta \leftarrow max\_lta + step\_lta$
    **if** $max\_lta > 220$ **then**
      $max\_lta \leftarrow 220$
    **end if**
    $min\_trig\_on \leftarrow min\_trig\_on - step\_trig$
    **if** $min\_trig\_on < 0.5$ **then**
      $min\_trig\_on \leftarrow 0.5$
    **end if**
    $min\_trig\_off \leftarrow min\_trig\_off - step\_trig$

41: **if** $min\_trig\_off < 1$ **then**
42: $min\_trig\_off \leftarrow 1$
43: **end if**
44: **return** GRID_SEARCH($min\_sta$, $max\_sta + step\_sta$, $step\_sta$, $min\_lta$, $max\_lta + step\_lta$, $step\_lta$, $min\_trig\_on$, $max\_trig\_on + step\_trig$, $min\_trig\_off$, $max\_trig\_off + step\_trig$, $step\_trig$, $num\_iterations + 1$)
45: **end if**
46: **end procedure**

### 2.3.3 Experiments and Results

We consider 14 days of raw data from June 1st to June 14th recorded from STRA seismic station located on Stromboli (Italy) (more details on the dataset are also described in section 3.3.1). From these data, an EQs dataset is extracted thanks to the expertise of a human operator, by using the developed tool for manual cut of events (section 2.3, D'Alessandro et al., 2022). The entire dataset extracted contains 1506 EQs. The most significant number of EQs are in the first 4 days of June, i.e. 743 EQs. This set was used as a training set. The subdivision of the dataset for Training and Test is described in Table 2.1.

TABLE 2.1: Dataset subdivision for Training and Test

| | Num. Eqs extracted | Phase |
|---|---|---|
| First 4 days | 743 | Training |
| Days 5-14 | 763 | Test |
| Total | 1506 | |

#### Training Phase

A large grid was explored to better view the distribution of the *qni* values. In Training phase, the grid with the highest density of large *qni* values is obtained by combining a range of window values of STA and LTA, respectively from 2 s to 16 s and 20 s to 220 s, in steps of 2 s for STA and 20 s for LTA. The same representation of the parameter values as in figure 2.6 was used. For each combination of windows, $t_{on}$ and $t_{off}$ threshold values were combined, respectively from 1 to 7, in steps of 0.5 for both. Every cell is a $12x12$ matrix. Figure 2.7 shows a screenshot of the terminal grid (8x11), or the result of this experiment. At the beginning the grid was 5x5 and step values for STA and LTA were 2 s and 20 s respectively. After 6 iterations, a 8x11 grid was obtained, according to the STA, LTA lower and upper bounds defined in the Algorithm 1.

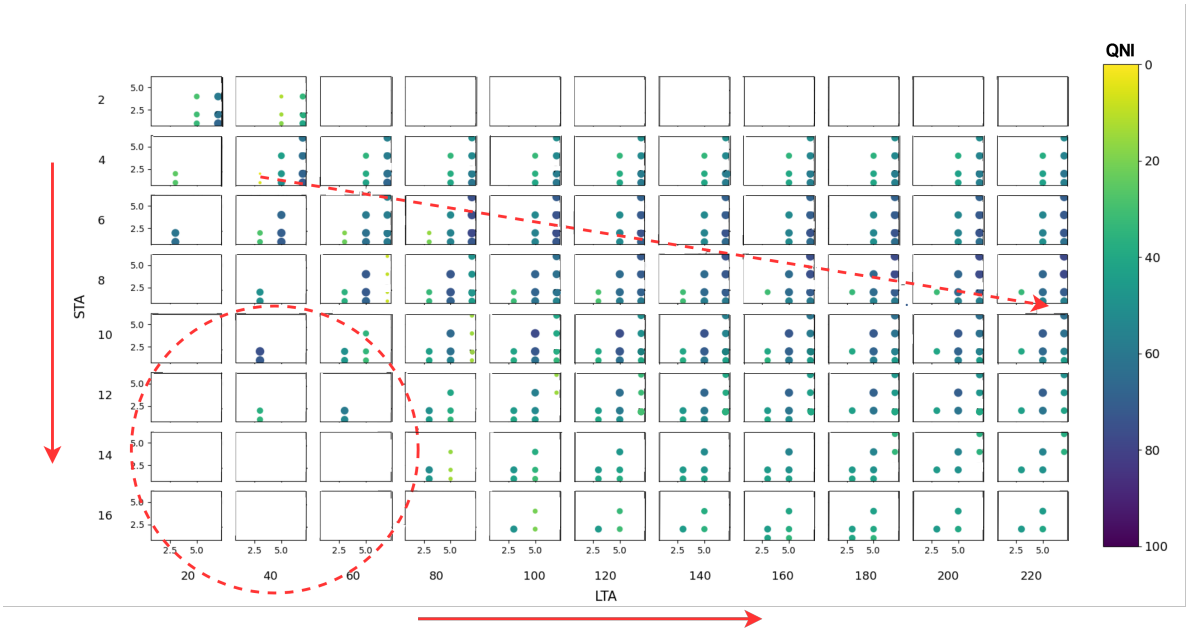



FIGURE 2.7: Grid-search exploration for STA windows (rows) and LTA windows (columns), represented by the red rows. Both STA and LTA windows are expressed in seconds. Trigger on / off thresholds, respectively in abscissa and ordinate, are shown in every cell of the matrix. The red circle shows a region where most of the *qni* values are 0. Red dashed line indicates a direction where the *qni* values are increasing.

In the lower-left region (highlighted by a red circle), where STA are between 12 s and 16 s and LTA are between 20 s and 60 s, most of the *qni* are 0 (no visible circles), only a few *qni* values around 20, which means that this parameter's combination is not suitable for detecting the EQs efficiently. A dashed red line was outlined to show a direction where the *qni* values are increasing. The highest values of *qni* are found in the central region from left to right, where STA is between 6 s and 10 s, LTA between 60 s and 220 s, $t_{on}$ between 5 and 7 and $t_{off}$ between 2 and 5. Neither *qni*, in the first and last rows of the grid, shows an improvement; while the continuous shift towards the right region shows a saturation of the values. This means that increasing the LTA values does not improve the search for local events.

From the training phase, a list of the *qni* values in descending order was extracted and generated a plot, shown in Figure 2.8. An index identifying the value of *qni* is indicated on abscissa, while the value of *qni* is indicated on the ordinate. The *qni* decreases with a moderate slope up to the value 30 and then rapidly drops to 0. As a consequence, it was decided to extract the quadruple associated with the highest *qni* value to carry out the test phase, i.e. the one with value 0.78 resulting from the quadruple: 6, 80, 7, 2. As can be seen, the quadruple is positioned exactly in the distribution indicated by the dashed red line in figure 2.7. The selected $t_{on}$ and $t_{off}$ define a good balance between the quality index and numerosity index (see eq. 2.3, 2.4) of the detected EQs. This is because a lower $t_{on}$ than 7 can increase the number of false positives, and higher $t_{off}$ can worsen the time's precision of the extraction.

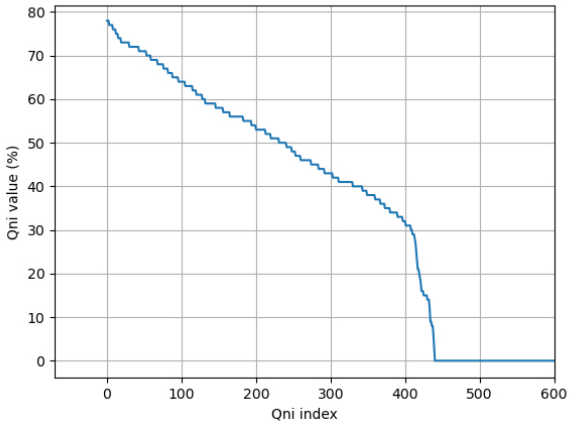



FIGURE 2.8: *qni* values distribution

To confirm the analysis of how *qni* varies with different STA and LTA window lengths, multiple parameter grids were examined to illustrate the evolution of this trend, as shown in Figure 2.9. In this case too, the *qni* values are defined for various window sizes as well as for different values of $t_{on}$ and $t_{off}$. Grid (a) explores a smaller range with fewer combinations, while grid (b) extends the exploration to include a larger set of STA and LTA pairs. In both cases, a diagonal trend in *qni* variation is visible, highlighted by the red arrow.

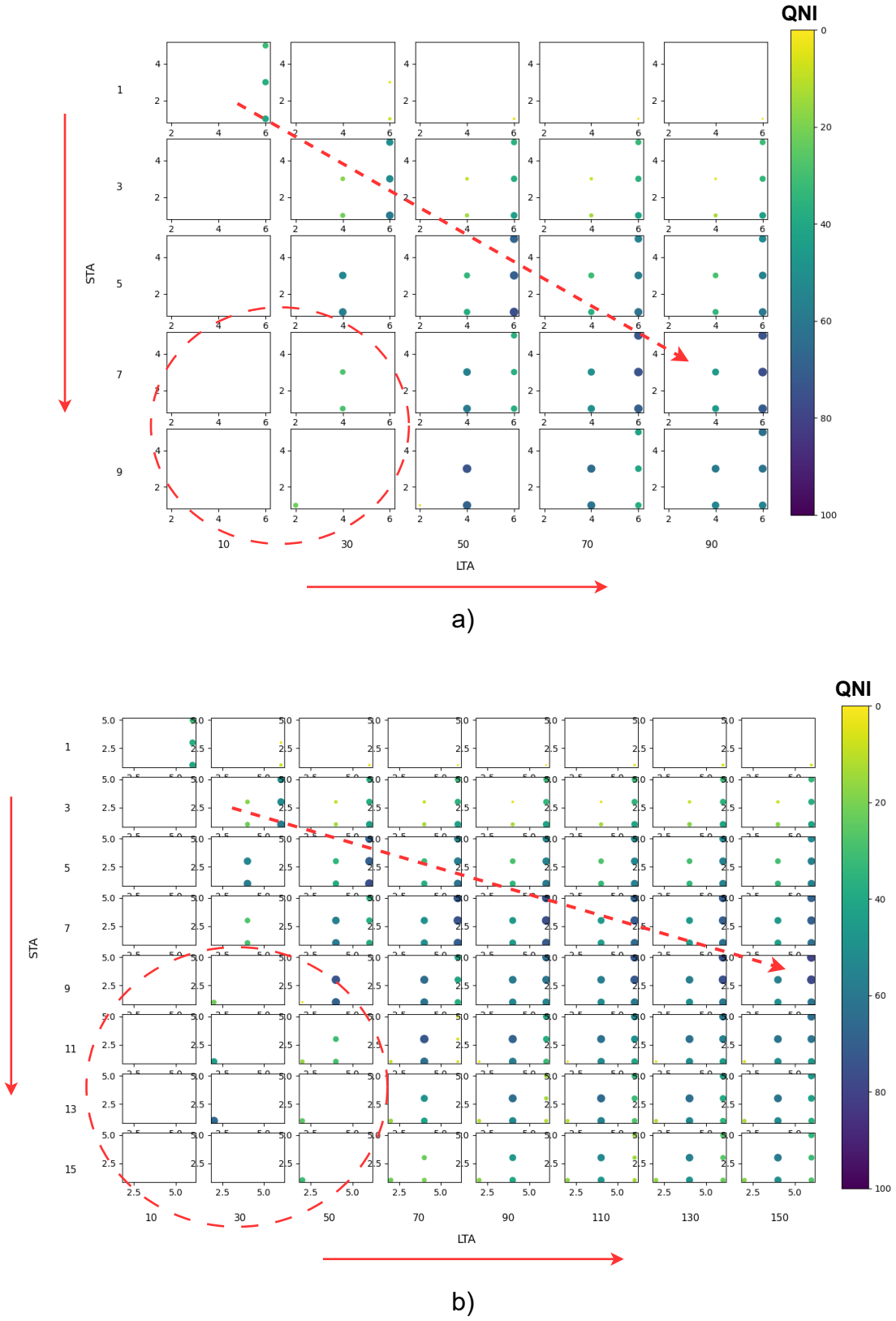



FIGURE 2.9: Here the same representation of the grid of the Figure 2.7 was used. *qni* values obtained by varying STA and LTA window lengths for two different configurations: (a) a denser grid with narrower STA and LTA intervals; (b) a broader grid exploring a wider range of window combinations. The color scale represents QNI values increasing

The 11×12 search grid shown in Figure 2.10 is the result of successive refinements based on the two previous grids (Figures 2.9a and 2.9b), where the initial experiments were conducted. The STA and LTA windows range from 1 s to 23 s and from 10 s to 210 s, respectively, with step sizes of 2 s for STA and 20 s for LTA—following the same representation used in Figure 2.7. For simplicity, the $t_{on}$ and $t_{off}$ values were left unchanged. A similar pattern to that observed in Figure 2.7 can be noted: in the bottom-left region (highlighted by the red circle), most *qni* values are zero, while the highest *qni* values are concentrated in the central region, extending from left to right. This trend thus confirms the choice of the parameter combination that yielded the best results in the grid-search optimization.

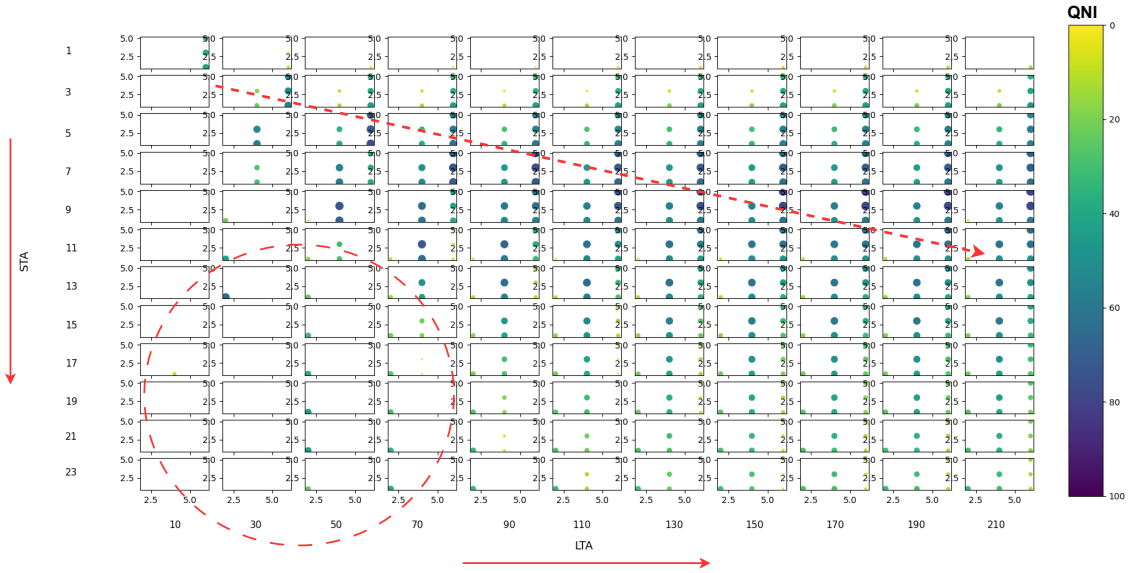



FIGURE 2.10: The 11×12 grid as the outcome of refinements developed from the two previous grids (Figures 2.9a and 2.9b), where the initial experiments were performed. This grid-search exploration confirms the optimal parameter selection determined through the grid-search procedure.

### Testing Phase

The common approach to the selection of the STA/LTA values is the adoption of literature-suggested ones. Specifically, these values are STA 1 s and LTA 10 s. Regarding the $t_{on}$ and $t_{off}$ values, they were set at 7 and 2, such as the ones found through the training phase. The *qni* computed is indicated using these values as literature quadruple. The test results are shown in table 2.2. In particular, every row corresponds to a range of days in which a certain number of EQs have occurred. This number is indicated in the "Num. EQs extracted" column; in the other columns, the *qni* values are reported after an experiment with the associated days was performed with the corresponding quadruples: training quadruple (third column, quadruple from training phase) and literature quadruple (fourth column, quadruple extracted from the literature). An overlap of the days for testing purposes was carried out. In the first row, the testing result is shown using the same dataset used in the training phase. This first comparison was made to show the result compared to the use of the literature quadruple.

TABLE 2.2: Testing results

| | Num. EQs extracted | (*qni*) Training quadruple 6 s 80 s 7 2 | (*qni*) Literature quadruple 1 s 10 s 7 2 |
|---|---|---|---|
| First 4 days (train data) | 743 | 0.78 | 0.59 |
| Days 5-8 | 395 | 0.65 | 0.36 |
| Days 6-9 | 425 | 0.64 | 0.38 |
| Days 7-10 | 373 | 0.56 | 0.33 |
| Days 8-11 | 397 | 0.54 | 0.26 |
| Days 9-12 | 291 | 0.48 | 0.21 |
| Days 10-13 | 271 | 0.51 | 0.3 |
| Last 4 days (11-14) | 170 | 0.5 | 0.3 |
| | | $\mu$: 0.55 $\sigma^2$: 0.004 (Test) | $\mu$: 0.3 $\sigma^2$: 0.003 (Test) |

Table 2.2 also shows that *qni* changes over time as the data become more different from the training set. This is due to the rapid evolution of the volcanological phenomenon, and therefore the EQs generated by it. As is well known (Andronico et al., 2021), the duration, amplitude and frequency content of the EQs can vary rapidly as the volcanic process evolves. The results show that on average, the training quadruple produced a *qni* 0.24 higher than the literature quadruple. Furthermore, the mean value for testing phase is 0.55 for the Training Quadruple and 0.3 for Literature Quadruple. Another strength of this approach is the tuning of the evaluation measures: for instance, qi is one of the measures that can be adapted based on the data and objectives one wants to obtain; in this case (looking for EQs in the raw data), the mean of the deviations was used to evaluate the quality of the cut, but one can consider using different evaluations such as median, mode or kurtosis. This possibility can lead to extending the research to the general volcano domain and also in seismic domain in tectonic areas.
The main limitation of this study is the low-variability of the dataset caused by the presence of just one seismic station (STRA). Based on the time-consuming process of extracting the seismo-volcanic events from real raw data, future improvements will mainly concern the variability of the dataset; extending to other types of seismo-volcanic events, such as: VLP, landslides, volcanic tremors and others (Wassermann, 2012), but also to local, regional and teleseismic events. To achieve this, the human operator can be replaced by automating the validation process through the use of multiple seismic stations recording the same signal within the same area. This approach replicates the current method used by operators, but will be enhanced by deploying seismic stations positioned at approximately equal distances around the crater (Fenner et al., 2022), ensuring more uniform coverage and reducing potential sources of error. Another limitation is specific to the Grid Search technique. In general, Grid Search is a powerful technique for optimising algorithm parameters, but it ought to be used judiciously as it may be time-consuming when there are many parameter combinations to evaluate. Other techniques such as Random Search or Bayesian optimisation may be more efficient alternatives in some cases. The choice depends on the specific problem domain and available resources. Stromboli remains a case study, but the method is applicable to any type of seismo-volcanic signal and can therefore be used on other volcanoes as well. For instance, it is also suitable for earthquakes in non-volcanic areas.

# Chapter 3

# Active Learning

## 3.1 Random Forest

Random Forest (RF) is a classic method used in Machine Learning for classification and regression problems that works by creating a multitude of decision trees during training (Breiman, 2001), improving classification performance. That is, RF returns interpretable results through decision trees, where the output is the class selected by most trees, in case of a classification problem. Instead, for a regression task, the output is the average of the predictions of the trees.
Each tree is trained on a different subset of the data, typically using a technique called Bagging (Bootstrap Aggregating, Figure 3.1), where subsets of dataset are generated by random sampling. This approach helps reduce variance and prevents overfitting by averaging the predictions of multiple diverse trees.

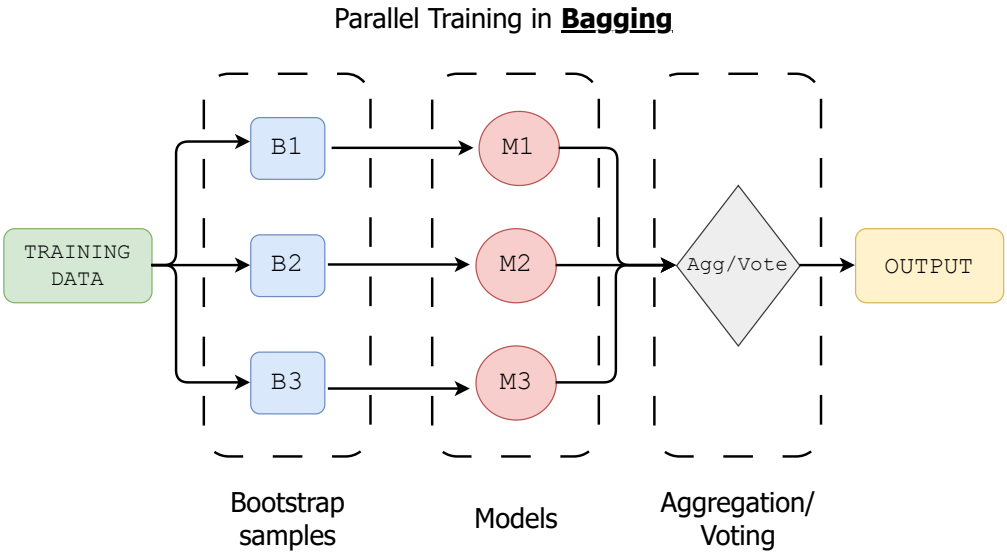



FIGURE 3.1: Bagging ensemble method. It involves training multiple models in parallel computing.

In contrast to boosting (Figure 3.2), which sequentially trains classifiers with a focus on correcting previous errors, bagging trains all classifiers independently in parallel. As reported in Bishop, 2006, bagging is especially effective when the base classifiers are unstable, as in the case of decision trees, since aggregating their outputs significantly improves overall stability and performance. The sequential training performed on decision trees in boosting, imposes each tree attempts to correct

the errors made by its predecessor, by assigning weights based on accuracy. Instead of using random subsets like in bagging, boosting reweights the training instances—placing more emphasis on those that were previously misclassified. This iterative process, focus on difficult-to-predict samples, reducing bias and building a strong classifier from a series of weak learners. Common implementations like AdaBoost (Schapire, 2013) or Gradient Boosting (Friedman, 2001) fine-tune this process by adjusting the way errors are propagated and models are combined.

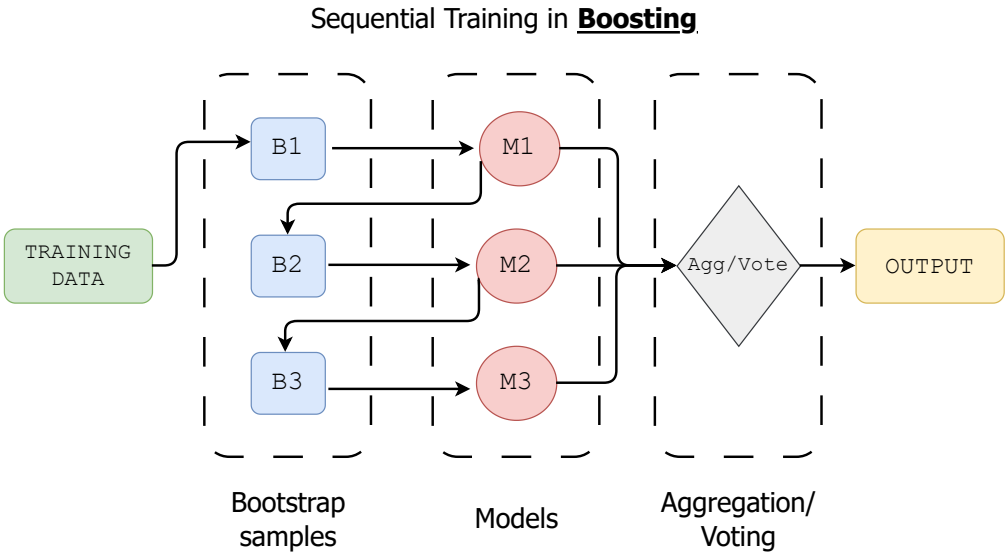



FIGURE 3.2: Boosting ensemble method. It involves training multiple models in sequential computing.

Bagging has also been used in various domains such as bioinformatics (Diaz-Uriarte and Andrés, 2006) for gene selection, and remote sensing (Pal, 2005), where it enhanced classification accuracy in land cover mapping. These works highlight the versatility of bagging across tasks with high-dimensional data and noisy labels, confirming its role as a fundamental technique in modern machine learning ensembles.
Bagging can be used with unstable models like decision trees while boosting works better for stable models like linear regression.

An example of the use of RF in the geoscience field is given by the study of Wenner et al., 2021. They used RF for a near-real-time events classification on a windowed seismic data. The signal classification involves noise, slope failure and earthquake classes, exploring the RF performance in case of non-ideal encountered conditions; like, poor network coverage, imbalanced data sets and low signal-to-noise ratios.

## 3.2 Active Learning: State of the art

Generally, an Active Learning model selects a subset of data in which there is high uncertainty, subsequently asking a human operator an intervention of manual labelling; later this updated data is added to the training set for a new training tour. An overview of this approach is represented in Figure 3.3

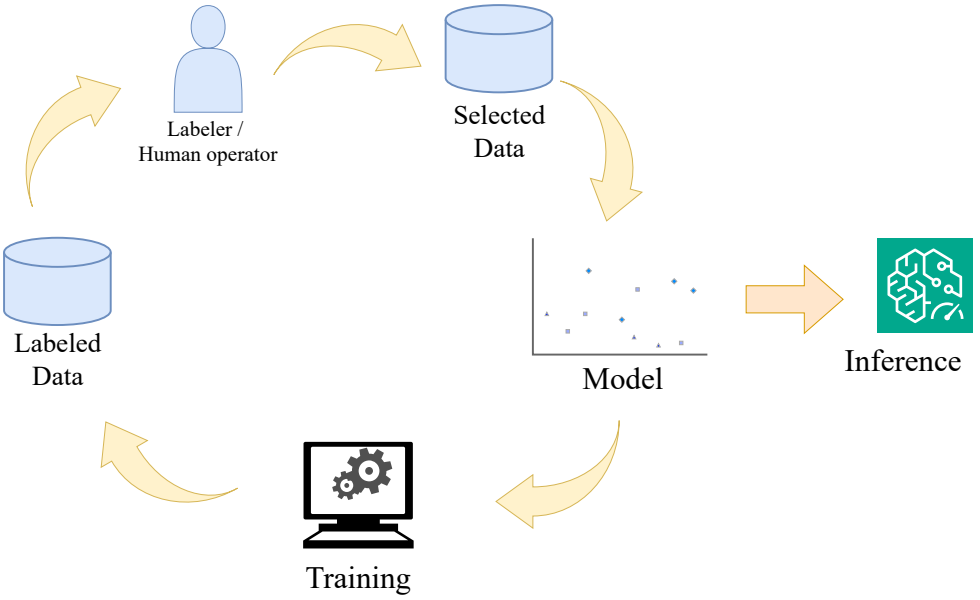



FIGURE 3.3: The Active Learning workflow.

This operation allows improving the performance of the classifier, making it more and more precise in the prediction. A general introduction to active learning together with its theoretical foundations can be better investigated by the book of Settles (Settles, 2012).
In the specific field of geoscience, there are several works that use active learning techniques. For instance, a new active learning technique to perform noise detection in seismic traces is proposed in the work from Chambefort et al., 2022. The detection of the noise here is important to detect information on the composition of the subsurface for oil industry. Each denoise step is then ideally followed by a Quality Control (QC) stage performed by means of human expertise. The novel active learning methodology helps to sequentially select the most relevant data, which are then given back to a human expert for labeling.
It was also implemented a strategy for the first break picking task, where the parameters to be chosen are selected with an active learning approach (Richardson, 2021). It uses the samples that have already been labeled and those that have yet to be labeled, to choose the samples that may be most useful for the human operator to be labeled in order to obtain a correctly labeled dataset. Using the active learning approach, it has been shown that the error in break picking decreases rapidly.
A study who introduces an active learning framework to enhance the classification of volcano-seismic events using deep learning models is the one from Manley et al., 2022. This research is similar to the application carried out for the classification of Explosion Quakes (EQs) presented in this chapter. The researchers applied a diversity-based active learning strategy, that is, diversifying the dataset from existing labeled data to maximize the performance. They integrated this approach with a convolutional neural network (CNN) classifier and tested it on datasets from two different volcanoes: Nevado del Ruiz in Colombia and Llaima in Chile. The results revealed that, for the Nevado del Ruiz dataset, active learning significantly improved model performance and reduced the amount of labeled data required. However, for the Llaima dataset, offers no increase in performance because the benefits were less pronounced, likely due to differences in data separability.

## 3.3 Active Learning approach for Classifying Explosion Quakes

In this thesis, an application of the Active Learning approach is proposed. This application improves the performance of the RF classifier for the classification of EQs vs Non Explosion Quakes (NEQs) thanks to data filtered by human intervention. This approach allows for enriching the training set to specifically recognize the EQs. The tool has been developed with Python programming language and uses the RF classifier as a classification method. It also allows for an editing phase of the signal before the relabelling, consisting in cutting and reducing the time window of the original signal.

### 3.3.1 Data Collection and Feature Extraction

On Stromboli volcano, data are recorded by the seismic monitoring network of the Istituto Nazionale di Geofisica e Vulcanologia (INGV), which is composed by three-component, short-period and broadband seismometers (Figure 3.4). The analog signals are sampled at 50 Hz and transmitted to the Vesuvius Observatory acquisition centre, where they are processed by a distributed system for seismic data acquisition and analysis.

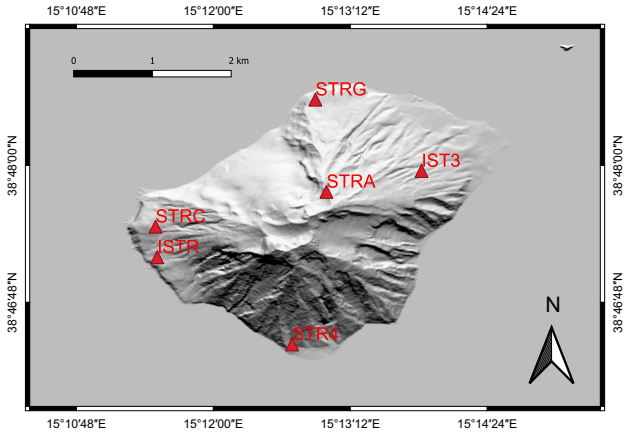
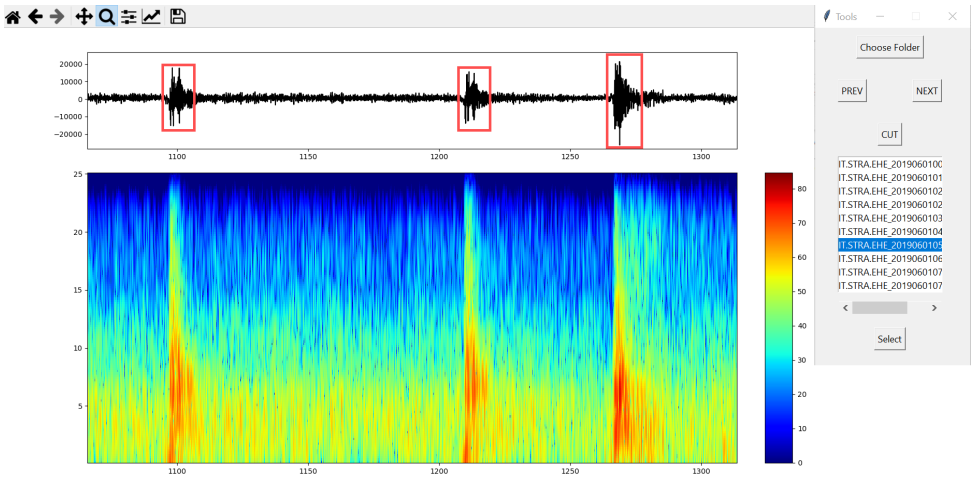


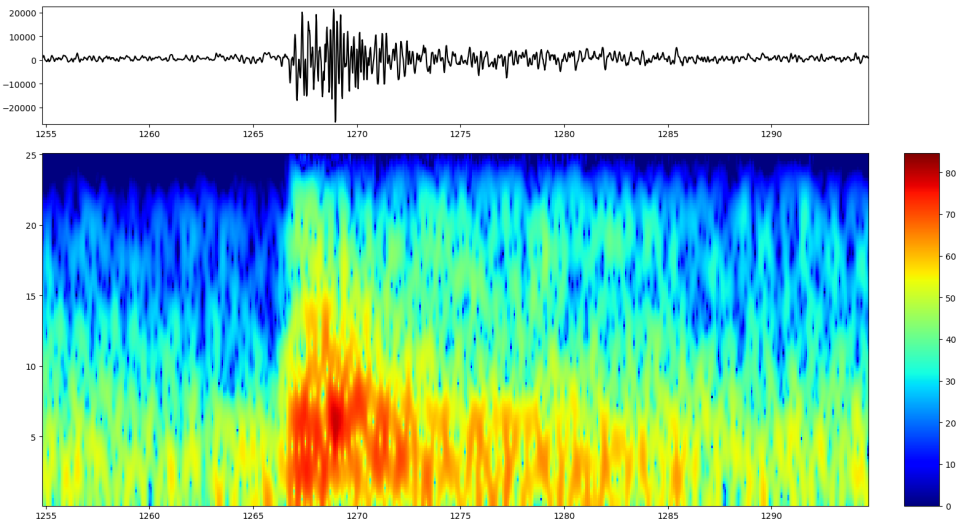

FIGURE 3.4: Permanent seismic network of Istituto Nazionale di Geofisica e Vulcanologia (INGV) on Stromboli. Map created using QGIS Development Team, 2024.

The dataset for the Active Learning application consists of the signals collected by STRA seismic station (Guralp 40T-60s), managed by INGV-OV (Osservatorio Vesuviano), which was concerned to 20 days, starting from 1 June 2019. The STRA seismic station, located on Pizzo sopra la Fossa, in the proximity of the Stromboli crater area, is geophysically the most suitable for the detection and analysis of EQs. It is at a minimum distance from the seismogenic source. This position allows signals with a high signal-to-noise ratio to be recorded, minimizing the effects of attenuation, scattering, and interference due to propagation through the volcano's heterogeneous structures. From a volcanological point of view, this allows a more accurate characterization of source parameters (such as amplitude, duration, and dominant frequency) and better discrimination between surface explosive events and other forms of volcanic seismicity. This is compared to more distal stations located at the base of the volcanic edifice. For this application, the East-West component of the

seismic signal was considered as it is the one in which the EQs have the best signal-to-noise ratio. The continuous seismic signal was analysed by an expert operator who was in charge of the identification of EQs waveforms. The analysis of the EQs events was done both in time and spectral domains by means of time-spectral analysis. To make this task easier, a specific software tool was implemented. The software tool allows for viewing at the same time the seismic waveform and its spectrogram in logarithmic colour scale. The waveforms are temporally synchronized with the spectrogram for easier comparison. Moreover, the software tool implements the possibility of zooming in on the waveform and its spectrogram and saving the portion of the highlighted signal. For this specific application, short-time Fourier transform was calculated using 0.5 s sliding time windows with 90% time overlapping. The graphical interface of the tool is shown in Figure 3.5.




(a)



(b)

FIGURE 3.5: (a) The collection of the data. Seismogram on top, spectrogram at the bottom. A sidebar at the right allows navigation between the varieties of file. Explosion Quakes are visible as red rectangles on raw signal. (b) Zoom on the third EQ

In order to perform automatic detection of EQs events, it is necessary to have an initial labelled database of seismic waveforms that will be used for training, validation and testing of the procedure. Two classes of events are thus generated, EQs and NEQs, in order to check if the classifier is able to distinguish between the two different types of waveforms. Then the NEQ dataset was generated from raw signals with Random Shift. Random Shift refers to the random position on the time window of the raw signal which defines the start time of the NEQs plus the duration of the NEQs (the end time). To derive the NEQs durations, a statistical study was carried out on the EQs time windows; 1 second wide bins were chosen and the most numerous bins were extracted. As a result, the most numerous time windows were between 4 and 18 seconds. The first dataset extracted from the raw data consists of 1506 EQs waveforms were extracted. As is well known, the application of a classification procedure requires the identification of an individual measurable property or characteristic of a phenomenon, commonly called features. Similarly to Titos et al., 2018, it was decided to extract a total number of 20 features based on well known statistical measures that are Standard Deviation, Median, Skewness and Kurtosis. These measures were determined by employing the raw signal and its part extracted through frequency filtering. The characteristics were therefore determined by filtering the signal in four disjoint frequency bands by means of Butterworth bandpass filters. Butterworth's filters have been designed to have the following passbands: 0.025-0.5, 0.51-5.0, 5.1-10.0, 10.1-20.0 Hertz. So, the final set of features: median, std_dev, skew, kurt, median0.025-0.5, std_dev0.025-0.5, skew0.025-0.5, kurt0.025-0.5, median0.51-5, std_dev0.51-5, skew0.51-5, kurt0.51-5, median5.1-10, std_dev5.1-10, skew5.1-10, kurt5.1-10, median10.1-20, std_dev10.1-20, skew10.1-20, kurt10.1-20 (Figure 3.6). The second dataset contains 1506 NEQ events. All these features were determined for both of the generated dataset. Finally, a mix between EQs and NEQs was made, obtaining one unique dataset of 3012 events from which the labels were also extracted (Figure 3.6). Choosing informative, discriminating and independent features is a crucial element of effective classification algorithms. In this regard, an analysis was carried out specifically for each characteristic described previously. To make this analysis, the RF classifier was used, in particular, a permutation feature importance selection paradigm, which has shown that half of the features have the most informative content for the classification of EQs. Among them, it emerged that the standard deviation was the most representative feature. Based on these results, the datasets used for the active learning approach are composed of the most important features.

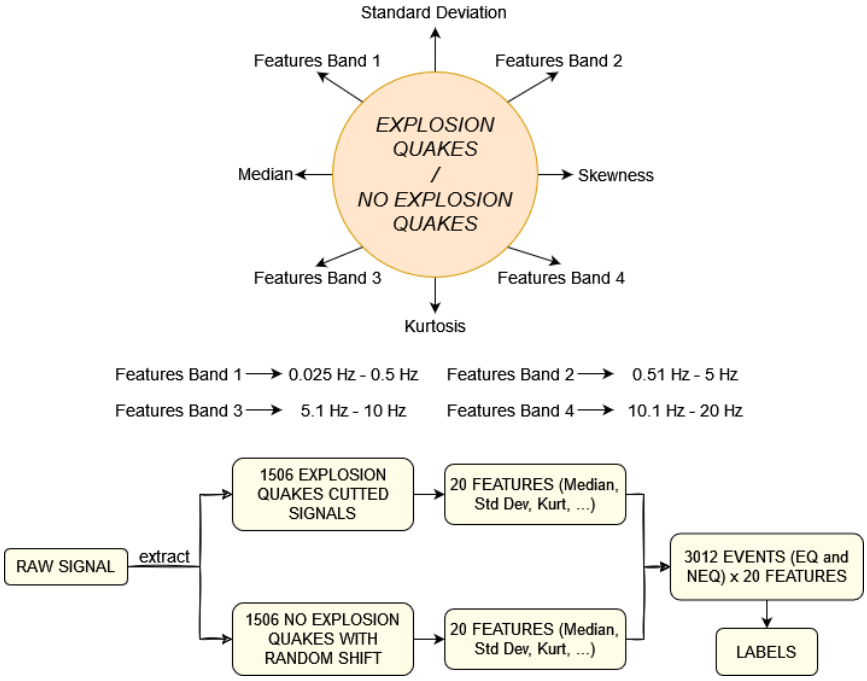



FIGURE 3.6: General diagram of the dataset structure and features

### 3.3.2 The proposed Active Learning tool

The RF algorithm outputs class probabilities for each input trace (specifically, the features extracted from the trace), representing the likelihood that the event belongs to a given class — that is, whether it is classified as an EQ or NEQ. These probability values are then used within the Active Learning approach to identify uncertain classifications, where human intervention can refine the labeling and improve the model's overall performance.
A small dataset of 400 events between EQs and NEQs was initially extracted to train the RF. For the test it was used another dataset of 400 events between EQs and NEQs: on the basis of the predicted class probabilities related to the single event belonging to the class (binary in this case), a subset of events with predicted class probabilities below a given threshold is chosen (in this experiment, 90% was used as threshold) and these events are re-proposed to the expert operator to re-examine them. The predicted class probabilities of an input sample are computed as the mean predicted class probabilities of the trees in the forest. The class probability of a single tree is the fraction of samples of the same class in a leaf. The operator can appropriately cut out the data and re-label it in order to enrich the starting dataset. Figure 3.7 shows a general scheme of the proposed Active Learning approach.

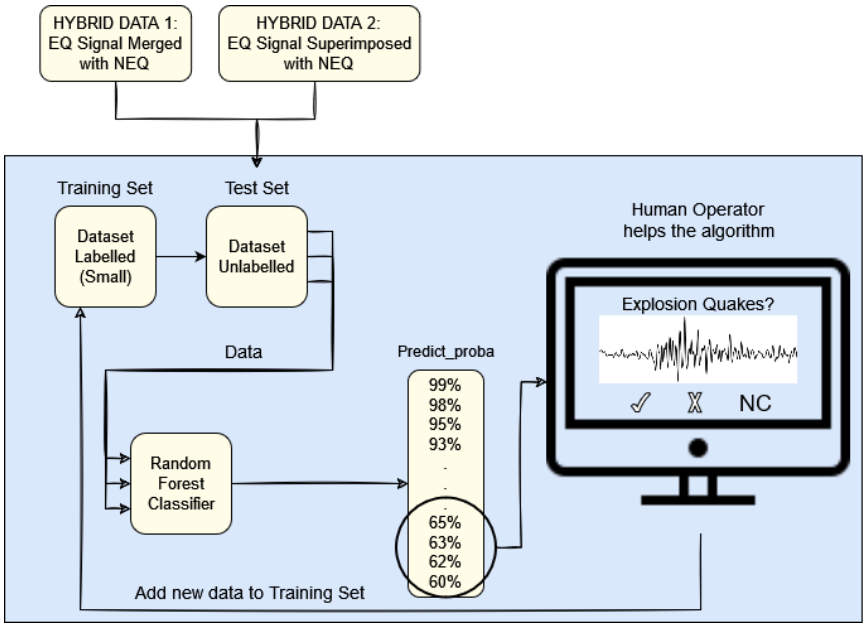



FIGURE 3.7: Active Learning Scheme

The tool uses the RF to suggest a prediction of a test set. The used parameters of the RF classifier are 20 decision trees with maximum depth 2, and each single decision tree uses the Gini impurity criterion (Breiman et al., 1984). In this work, the RF implementation provided by the Scikit-learn python library (Pedregosa et al., 2011) was used. After the prediction phase, the tool selects all the events that have a probability below a given threshold and show them in a window one by one. In this phase, the operator can optionally carry out the new cut of the signal and the relabeling (Figure 3.7). While the operator re-labels the signals (of the test set), they are added at the end of the previous training dataset for retraining using the added data. In the relabeling phase, the operator can also choose the "NC" option, which stands for "Not Classifiable". In this case, the events are not added to the Training Set as they cannot be labelled as EQs or NEQs. This option is particularly useful in the cases where the signal may represent instrumental anomalies, unexpected outliers or atypical events. In addition to the signals relabeled by the operator, those that have a probability above the threshold (which do not need to be relabeled) will also be added.

These added data increase the size of the starting training, hoping for an increase in classification accuracy. After a series of initial training and test steps with the intervention of the operator, the RF will become more and more precise in the classification of the EQs and consequently will have less and less need for operator intervention. This means that the algorithm will be able to label automatically as there will be fewer events that have a probability below 90%.
To further diversify the dataset, it was decided to create hybrid events between EQs and NEQs using the method reported in the Scheme as "Hybrid Data 1: EQs Signal Merged with NEQs". This method will be described in 3.3.3. The method reported as "Hybrid Data 2: EQs Signal Superimposed with NEQs" refers to the generation of hybrid events using the sum of EQs and NEQs. This last method will be used in future developments. In the case of "hybrid data 1" a wrong cut was simulated by an automatic method and this is corrected by a human operator. "Hybrid data 2" will be a simulation of an EQ with a low signal/noise ratio because there could be the possibility of having this type of events in the real data.

### 3.3.3 Experiments and Results

The reliability of classification techniques is critical for applications devoted to volcano monitoring. Initially, a balanced training set is used consisting of 400 events between EQs and NEQs and a balanced test set also formed by other 400 events between EQs and NEQs. This dataset is randomly extracted from the dataset represented in Figure 3.6. In this case, there is no active learning approach applied, but this experiment helped to demonstrate the effectiveness of the RF in recognizing EQs. In fact, the accuracy obtained for this first experiment was 97%.
After demonstrating that the RF classification works as expected, the subsequent experiments were carried out starting from the random extraction of the dataset shown in Figure 3.6. In these experiments, the active learning approach is used. As mentioned in Section 3.3.2, to further diversify the dataset and make it closer to real data, it was being generated hybrid events that contain both an EQs and a NEQs by dividing them into classes as follows:

- class_0.8 20% EQs - 80% NEQs
- class_0.6 40% EQs - 60% NEQs
- class_0.4 60% EQs - 40% NEQs
- class_0.2 80% EQs - 20% NEQs

The percentages correspond to the time windows of the EQs and NEQs signals. The time window that identifies the NEQs is the continuation of the time window of the EQs. In other words, NEQs is the noise following EQs in the raw signal. The resulting signal is a merge of the temporal windows into a single signal. It was generated 50 signals per class, for a total of 200 hybrid signals. These signals were added to the Test Set in this order: 200 Mix EQs and NEQs (No Hybrid Events) and 200 Hybrid Events. So, the Test Set is 400 events and it is represented as the `A2` green circle in Figure 3.8.

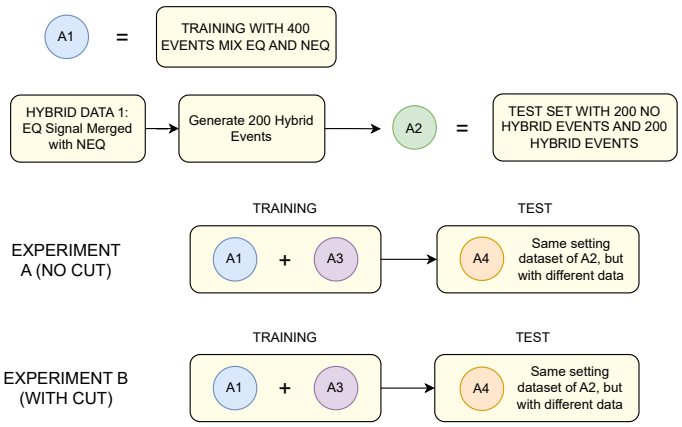



FIGURE 3.8: Active Learning Experiments and Results

The label `A1` indicate the Training Set with 400 events Mix EQs and NEQs randomly extracted from the dataset represented in Figure 3.6. These two sets have been used for a first initial experiment where 116 events are returned to be re-labelled and, opportunely, cutted out by the operator. Starting from the `A2` set, it was generated

an A3 set composed of both the elements labelled above 90% probability (the threshold used in the experiments) and the elements below opportunely re-labelled by the operator (the 116 events mentioned above). For both the experiments A and B, in the training phase, A1 set plus A3 set is used.
The difference between the training phase of the experiments A and B is that in A only relabelling (NO CUT) was made, while in experiment B both a relabelling and a new cut (WITH CUT) were made. In the test phase, the Test Set A4 is again composed by 200 non-hybrid events and 200 hybrid events, randomly extracted from the initial dataset (Figure 3.6), but, in this case, the data are different and disjoint from the A2 set.
Operator intervention through the new event editing proves the efficacy of the Active Learning tool. Four probability lists were extracted from the results of the Test Set A4 related to experiments A and B. In particular, the lists relating to EQs, NEQs, hybrid and non-hybrid events were extracted. Figures 3.9, 3.10, 3.11 and 3.12 show two curves: red curve refers to the Cumulative Distribution Function (CDF) of probability events of experiment A, while green curve refers to the CDF of the same data but of experiment B. In each plot, the abscissa reports the probability values referred to the events, and on the ordinate the percentage of the events that have that particular probability (i.e. the CDF). In these figures, it is possible to see an improvement in the CDF probability distribution after the cut-out has been made. The graphs show that the distribution improves both in classes and in hybrid and non-hybrid events. In fact, the CDF of experiment B, in which the active learning approach including waveform cut was applied, shows a trend of the curve that is below the red curve, in which only a relabeling operation was performed. This is a qualitative demonstration that human intervention by active learning could improve the prediction of the data since the uncertainty (here measured as the percentage of elements with low probability) is reduced.

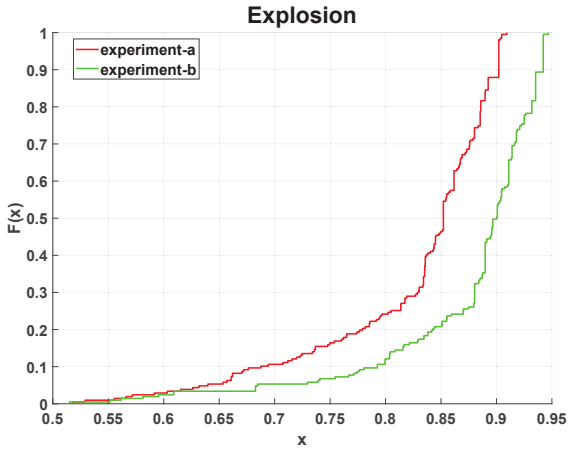



FIGURE 3.9: CDF of the EQs probabilities

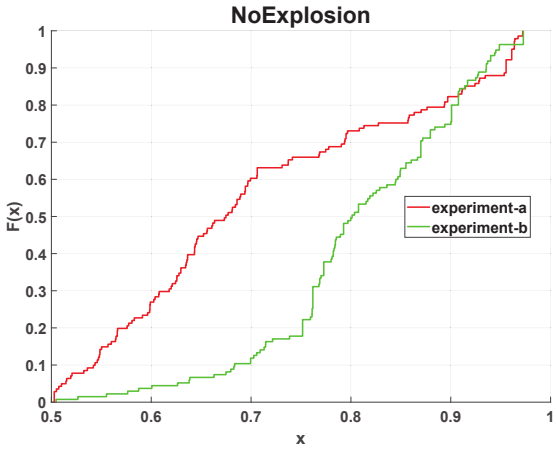



FIGURE 3.10: CDF of the NEQs probabilities

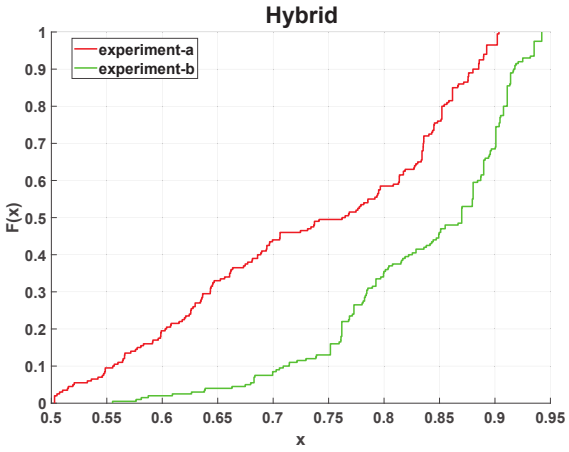



FIGURE 3.11: CDF of the hybrid events probabilities

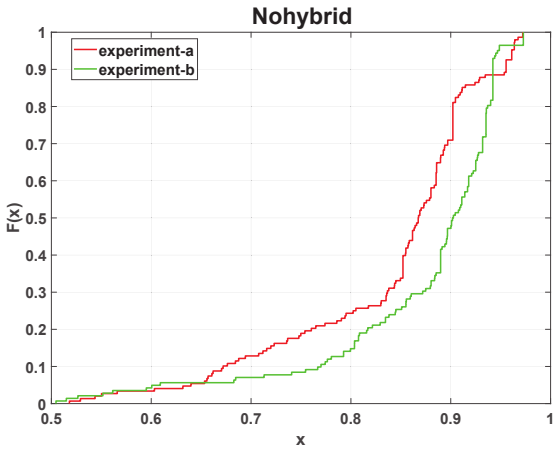



FIGURE 3.12: CDF of the non-hybrid events probabilities

# Chapter 4

# Deep Learning: Transformer

## 4.1 Transformer Architecture

Transformer Architecture represents nowadays the state of the art for natural language processing (Vaswani et al., 2023). Figure 4.1, shows an example of transformer application.

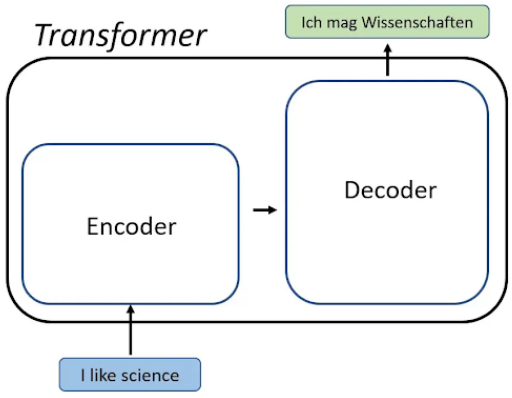



FIGURE 4.1: Transformer: Encoder and Decoder. Application on language translation (Unzueta, 2022)

Transformers consist of a simple architecture that uses attention mechanism intelligently. A transformer comprises two main components: an encoder and a decoder. The encoder processes the input, such as an English sentence, by vectorizing it and transforming it using attention to capture meaningful relationships between words. The decoder reverses this process, converting the transformed vectors into an output sequence, such as a sentence in a different language. This ability to encode and decode efficiently makes transformers effective for tasks like language translation.
In detail, the input, composed by tokens (e.g. words, numbers) is firstly converted into numerical vectors with fixed dimension, or, representative embeddings which are used as input for the deep learning model. This happens in the encoder, where the embeddings are converted into representative encodings using the attention mechanism; the decoder, instead, do the opposite, outputting the embeddings that are then converted into probabilities by the softmax function. The attention layer is composed by three neural networks where it takes the embeddings as input and, for each word embedded, produces three output vectors: Queries, Keys and Values (Figure 4.2).

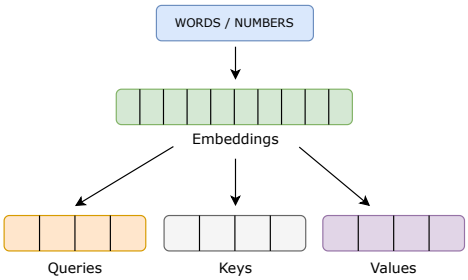



FIGURE 4.2: Content of attention layer: Queries, Keys and Values

The dot product between the query1, or the first word, and all keys in the sentence, produces the scores S (Figure 4.3).

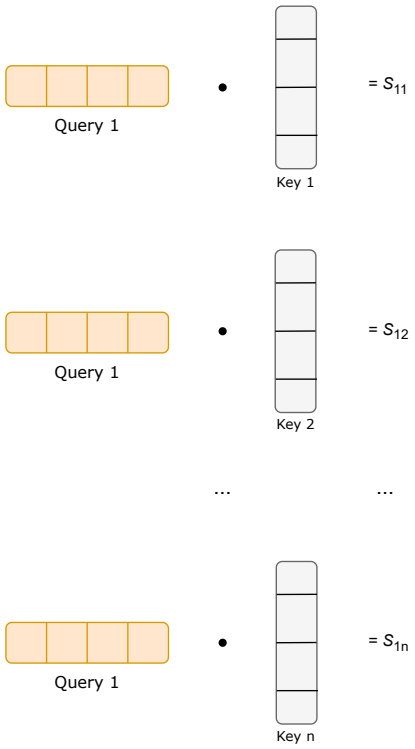



FIGURE 4.3: Product between Query of the work and all the keys

These scores represent the weights resulting from the product between Queries and Keys. They are then multiplied by the corresponding Values vectors, producing a set of scaled vectors that, when summed together, yield the attention vector for the word 1 (Figure 4.4).

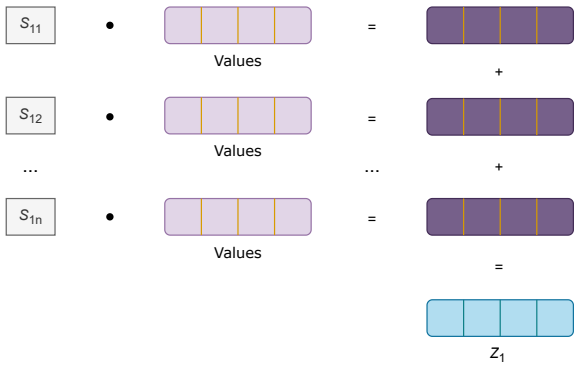



FIGURE 4.4: Resulting Attention Vector calculated for Word 1, depending on Query 1, Keys and Values vectors.

The same procedure is applied to all the other words of the input sentence to obtain the corresponding Attention vectors.

## 4.2 Transformer applications

We employed a transformer model that uses an attention mechanism called Earthquake Transformer (EQTransformer, Mousavi et al., 2020) for detecting earthquakes and the arrival times of seismic phases in two different approaches: local and regional earthquakes. The experiments are conducted following the held-out approach where the training is performed in one region for local approach and in other region for the global approach. In particular, the first approach focuses on local earthquakes in the northern Netherlands region, Groningen, where the magnitude of most events is below 3.0. Instead, the second approach concentrates on earthquakes with epicentral distance over 100Km occurred in the Turkey-Siria region. In the latter case, the earthquakes have magnitudes between 2.0 and 7.7, therefore with a much wider range than the local approach.
EQTransformer takes inspiration by the Transformer Architecture, which represent nowadays the state of the art for natural language processing (Vaswani et al., 2023), as previously described in 4.1. Transformer models, by their nature, are suitable for the analysis of time-series data; the seismic signal generated by earthquakes is an example of the latter. Starting from this principle, EQTransformer was developed. It also uses the concept of attention to understand the context of the input. An attention mechanism allows the neural network to process a subset of the input, focus on it, and then shift its attention to a different portion of the input. In this Deep Learning model, attention mechanism was applied on seismic waveforms, as illustrated in the Figure 4.5.

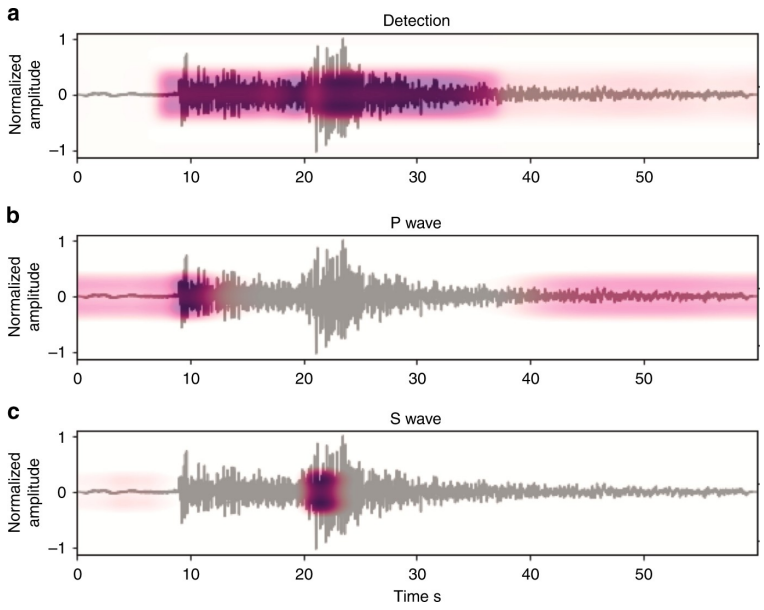



FIGURE 4.5: EQTransformer: Attention Mechanism

The highlighted part in every subplot consists of attention weights, the output vector calculated from the Queries, Keys and Values vectors. Figure 4.5a shows the detection of an earthquake by contextual information, emphasizing all the weights that represents it. Figure 4.5b focuses on local attention for P-phase, and Figure 4.5c the local attention for S-phase. The EQTransformer model is built around this concept. As illustrated in Figure 4.6, it incorporates Global Attention (highlighted in blue, in the end of Enconding phase) for earthquake detection and Local Attention (highlighted in green, in the beginning of Decoding phase) for phase picking.

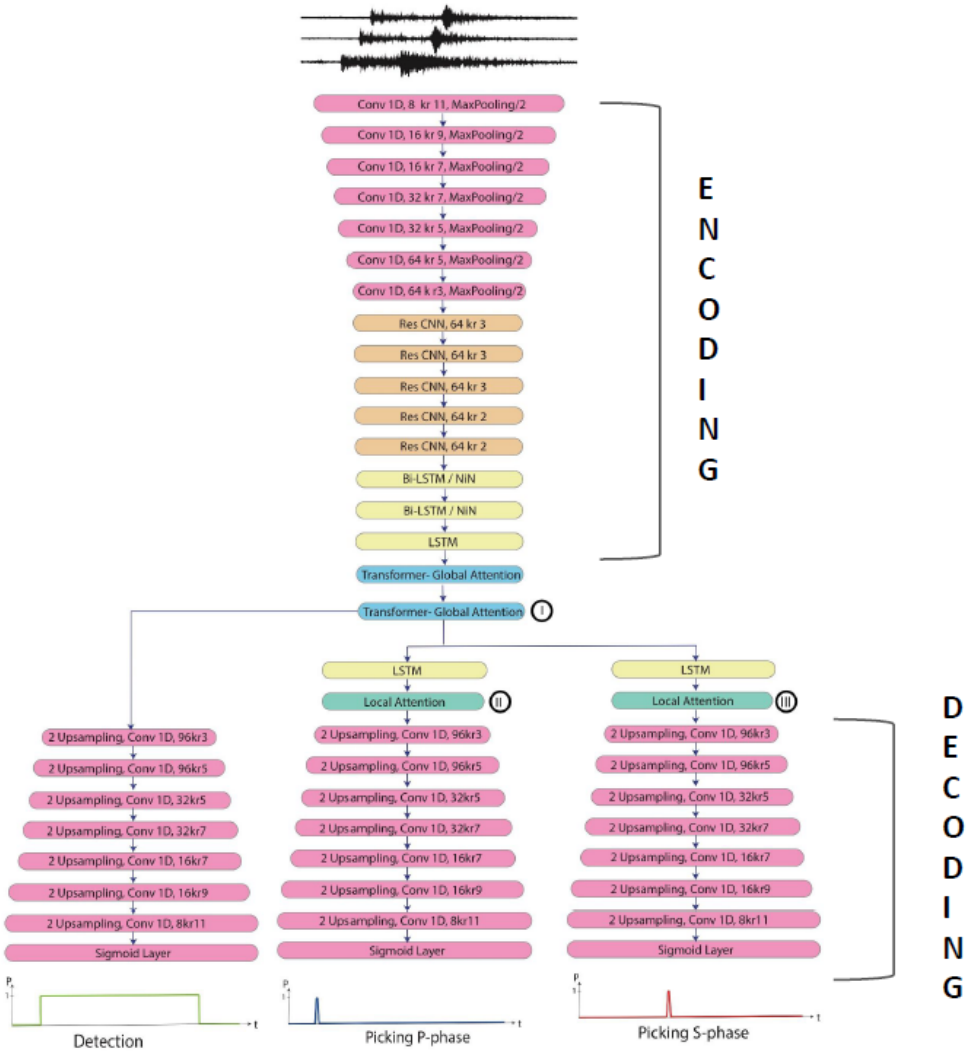



FIGURE 4.6: EQTransformer: Network Architecture. Global Attention (shown in blue) is responsible for earthquake detection, while Local Attention (shown in green) focuses on phase picking.

The model requires as input seismic traces with 6000 samples (1-minute long traces) x 3 channels as dimension, grouped into streams. Each stream is structured as follows:

- Trace 1 - (6000 samples, E Channel)
- Trace 2 - (6000 samples, N Channel)
- Trace 3 - (6000 samples, Z Channel)

This constraint has been considered for both local and regional applications.

The model combines several deep learning components to effectively process seismic waveform data. Initially, convolutional neural networks (CNNs) are used to extract local features from raw waveform inputs, capturing spatial patterns that are critical to identifying relevant signal features. These CNN layers are embedded

within residual blocks (ResCNN), which include skip connections to preserve information flow and improve training efficiency, especially in deeper networks. Following the feature extraction stage, the model employs bidirectional Long Short-Term Memory (BiLSTM) layers to capture temporal dependencies in the data. By processing both forward and backward input sequences, the BiLSTM layers enable the model to understand the context of past and future time steps, which is essential for accurate detection of seismic events and phase arrivals. After a representation of the input has been generated in the training phase, through the Attention mechanism, the signal is resampled by prioritizing regions with higher attributed weights, focusing on areas where more informative content is concentrated, enhancing the model's performance. In particular, Global attention is used to focus on relevant segments of the signal for overall earthquake detection, while local attention is specifically designed for precise phase picking of P and S waves. This multi-level attention structure allows the model to dynamically weight different parts of the signal, thus improving the detection accuracy and the timing precision. The output of EQTransformer is composed by three probabilities distributions, associated with Detection time, Picking P-Phase time and S-Phase time. Every time information is expressed in UTCDateTime format and shown with text and images. As an example case, Figure 4.13 shows the output of each plot associated with the same seismic event. Each plot shows the traces for each component (E, N and Z), meanwhile the last plot shows the probability of detection (Green curve), picking P Phase (blue curve) and S Phase (orange curve) in the time domain. The light blue and purple bars present for each component indicate, respectively, the P phase and the S phase that have been detected by the model. Sometimes, if the model wasn't able to make the prediction of the phase picking, the software search if there is an attribute called *p_status* setted to *manual*, prioritizing the time label indicated in the labelled dataset; in this case, the bars are coloured, respectevely, in blue and red.

### 4.2.1 Local Approach with EQTransformer: Groningen Region (The Netherlands)

For the local approach, EQTransformer was applied to 467 manually reviewed low-magnitude (0.1 to 3.4) local events cataloged in Groningen Region (The Netherlands) from 2016 to 2020. The seismic activity in this region is well-known in the Netherlands for its shallow events with low to moderate magnitudes (Dost, Ruigrok, and Spetzler, 2017). Focus on low-magnitude earthquakes (typically below M 3.0) pose challenges for both seismic monitoring and hazard assessment. These low-magnitude events, while not damaging individually, are crucial for understanding the evolving stress regime and for improving probabilistic seismic hazard models in the region. Furthermore, this region is also rich in non-earthquakes events, like Sonic Boom, Mine collapse, Induced or Triggered events; thus making it necessary to correctly distinguish between seismic and anthropogenic events.
The main objective of this work is to resolve a classification problem, distinguishing between seismic events and noise and, in case of a seismic event is detected, to perform the picking of P-Phase and S-Phase.
Starting from the seismic catalog containing metadata and continous waveforms, to create the dataset, a custom tool (Figure 4.7) was developed, enabling the download of both noise waveforms and seismic events in HDF5 format using the ObsPy library (Beyreuther et al., 2010).

```
9 Event(s) in Catalog:
2016-01-26T22:22:33.399999Z | +53.203,   +6.720 | 1.50 MLn | manual
2016-01-20T03:57:08.199999Z | +53.242,   +6.714 | 0.61 MLn | manual
2016-01-19T13:19:07.400000Z | +53.362,   +6.770 | 0.61 MLn | manual
2016-01-17T11:57:33.600000Z | +53.258,   +6.840 | 1.52 MLn | manual
2016-01-14T14:03:58.500000Z | +52.989,   +8.273 | 2.86 MLn | manual
2016-01-13T06:41:42.299999Z | +53.248,   +6.855 | 1.31 MLn | manual
2016-01-11T05:31:35.399999Z | +53.185,   +6.813 | 0.57 MLn | manual
2016-01-07T05:25:55.100000Z | +53.168,   +6.814 | 1.63 MLn | manual
2016-01-02T00:04:28.300000Z | +53.300,   +6.680 | 0.48 MLn | manual

TOTAL NUM OF PICKS
363

 Do you wanna get the noise? 1 - yes 0 - no: 0

 Do you wanna get the events? 1 - yes 0 - no: 1

 Do you wanna get the metadata from local csv? 1 - yes 0 - no: 0
YOU HAVE CHOSEN TO EXTRACT THE EVENTS WITH METADATA FROM OBSPY
```

FIGURE 4.7: Tool developed to create dataset in hdf5 format

This tool, allowed the creation of a dataset labeled with both seismic events and noise. The noise samples are selected by extracting streams between seismic events, with 15 minutes difference from any event. The seismic events were retrieved using the KNMI (Royal Netherlands Meteorological Institute) FDSN services, that is, services made available by the national institute to access standardized seismic data for research and monitoring purposes.
The collected data was recorded by 143 seismic stations with the following technical characteristics:

- Network - NL
- Region - Groningen
- Channels - HH; GH; EH

From the continuous waveforms downloaded, either for events and noise, each trace of the stream is preprocessed following these steps:

- ROTATION
- DETREND: Mean
- BANDPASS FILTER: 0.5Hz - 22Hz
- RESAMPLING: 100Hz (if $samp\_rate \neq 100Hz$)

First, the components were rotated to ensure that the signal was consistently represented by its East-West (E), North-South (N), and vertical (Z). Next, the mean value was removed from the signal (detrending) to eliminate any baseline offset. A bandpass filter in the range of 0.5 Hz to 22 Hz was then applied to retain the frequency content relevant to the target events while suppressing noise outside this band. Finally, traces were resampled at 100 Hz if the original sampling rate differed,

to ensure consistency across the dataset and improve computational efficiency for the model.
Compared to the regional approach (paragraph 4.2.2), where the model was pre-trained with a global dataset, in the local approach a training phase was performed. Figure 4.8 summarizes the subdivision of the dataset used to perform the training of the model. The dataset was equally split between Earthquakes and Noise: 9781 earthquake streams and 9781 noise streams, with a total of 19562 streams. Initially, the method performs a random shuffle to the entire dataset, then splits the dataset into Training, Validation, and Test set, respectively in: 80%, 10% and 10%.

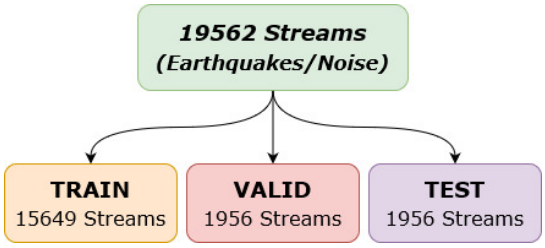



FIGURE 4.8: Dataset Subdivision for Training the model

For the training phase, data augmentation was performed to enrich the dataset. The augmentation, specifically, is a technique used to artificially increase the size and diversity of a dataset by applying transformations such as rotation, scaling, or adding random noise to the data. This helps improve the generalization and robustness of the model. These are the settings used for training the model:

- BATCH SIZE: 100
- EPOCHS: 100
- PATIENCE: 20
- add_event_r: 0.6
- add_gap_r: 0.2
- shift_event_r: 0.9
- add_noise_r: 0.5

"add_event_r", "add_gap_r", "shift_event_r" and "add_noise_r" are settings used to perform the data augmentation. To sum up them:

- add_event_r: Add secondary event randomly into the empty part of a Trace
- add_gap_r: Add an interval of zeros, simulating filled gaps
- shift_event_r: Randomly shifting the event within a Trace
- add_noise_r: Add gaussian noise

Each parameter takes value within [0,1], quantifying the extent to which the parameter affects the data, with higher values signifying a greater impact.
The performance of the training phase are shown by the F1-score and Loss curve, Figure 4.9 and 4.10.

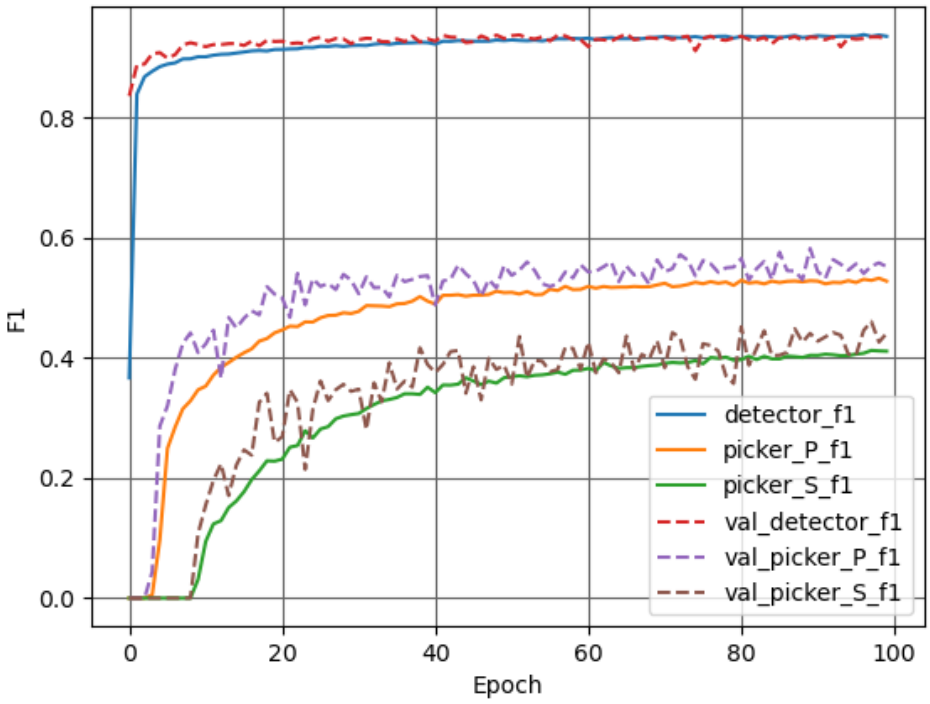



FIGURE 4.9: F1-Score from Training Phase of the EQTransformer model in the local approach

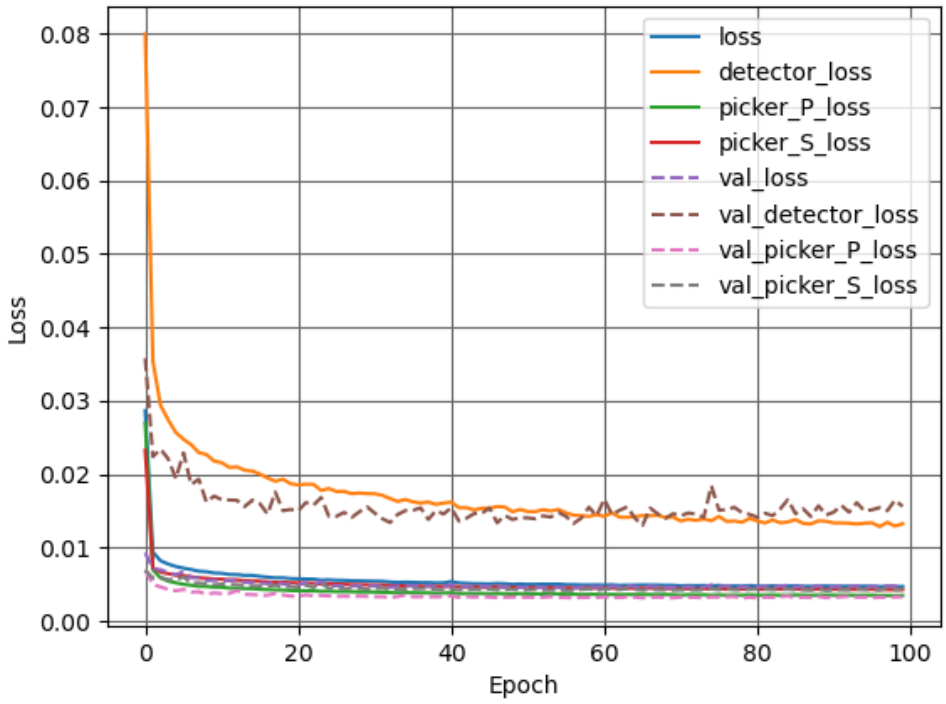



FIGURE 4.10: Loss curve from Training Phase of the EQTransformer model in the local approach

The training curves show that the detector component quickly achieves high performance, with F1 scores and losses stabilizing early and maintaining consistency between training and validation, indicating strong generalization. In contrast, the P-Phase and S-phase pickers improve more gradually, with lower F1 scores and higher losses, especially on the validation data, suggesting that these tasks are more complex, especially for the S-Phase, where the amount of noise content is higher. The persistent gap between training and validation in the pickers' losses may also suggest slight overfitting. Overall, the detector learns effectively, while the pickers could benefit from further tuning or better data quality.

After the model was trained, testing phase was performed on the Test set with the following settings:

- BATCH SIZE: 10
- Detection threshold: 0.2
- P threshold: 0.1
- S threshold: 0.1

The model uses three threshold parameters to select the probabilities, in terms of accuracy, related to event detection, picking P-Phase and S-Phase. In this case, both for the testing phase and for the prediction phase, the default values were used (like batch size, loss curves, number of cpus, ecc ...). To show the performance of the model in this phase, the confusion matrix was calculated, as shown in Figure 4.11.

The confusion matrix shows 98% accuracy based on the test split (1956 streams); furthermore, there is a balance between the detected event streams and the noise streams, thanks to the balancing of the starting dataset.

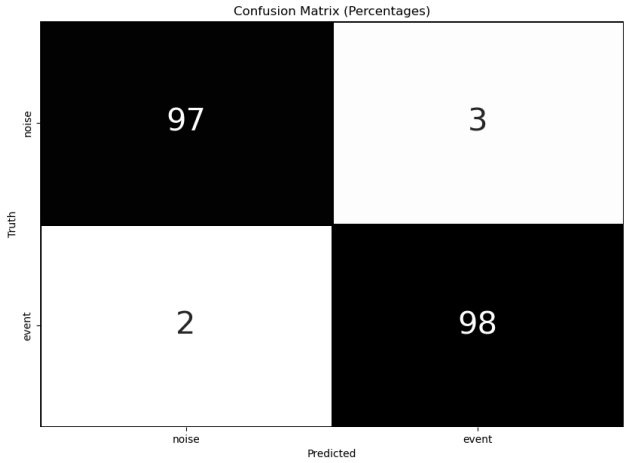



FIGURE 4.11: Confusion Matrix (percentages)

An example of event detected and phase picking in testing phase are shown in Figure 4.12 and 4.13.

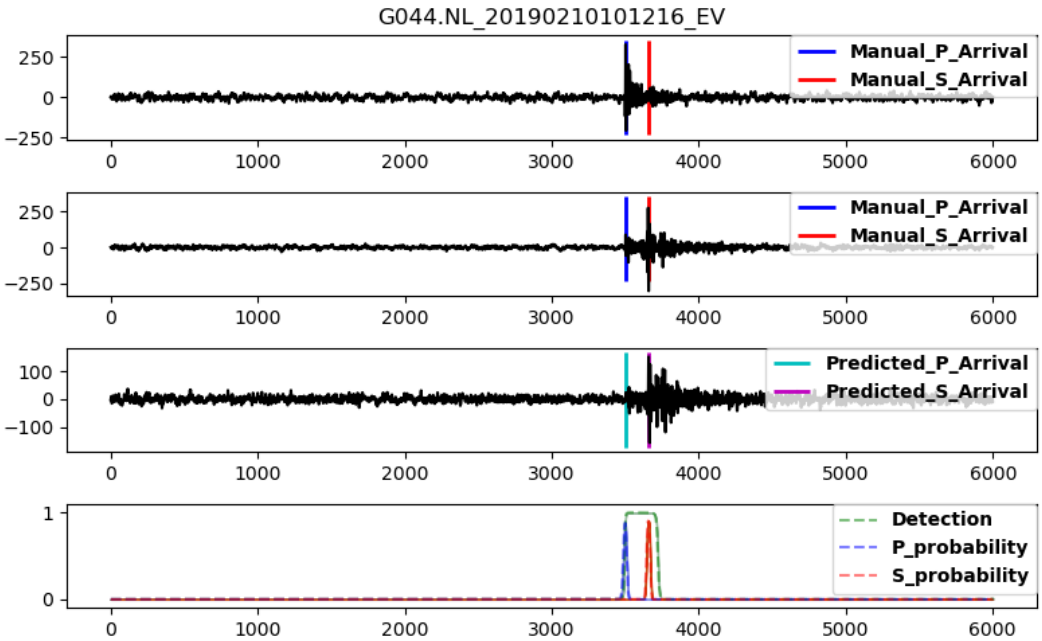



FIGURE 4.12: Example of event detection and phase picking in testing phase for G044 seismic station

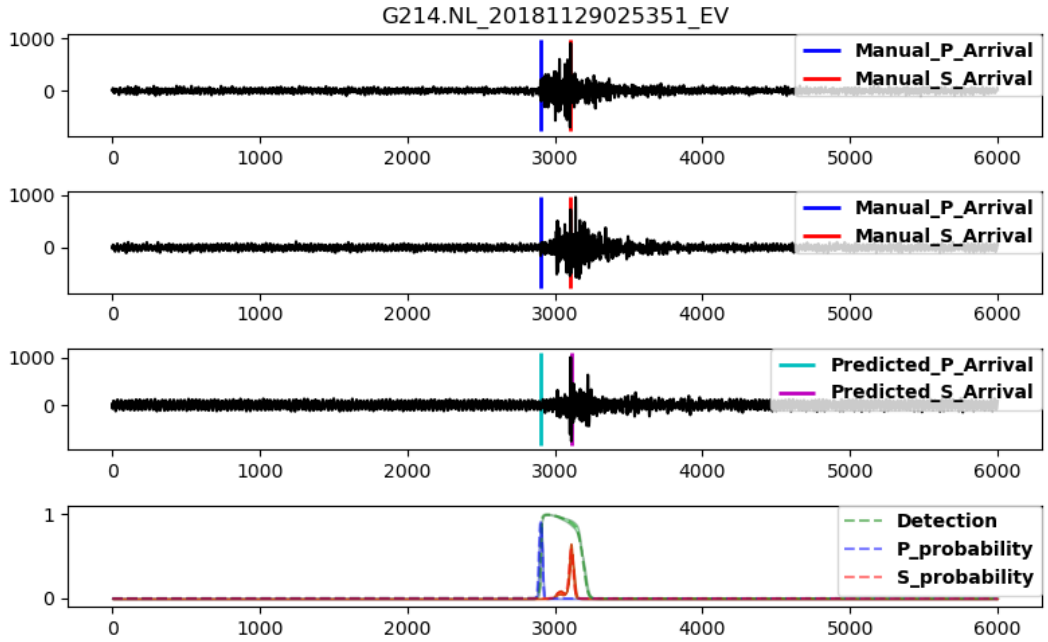



FIGURE 4.13: Example of event detection and phase picking in testing phase for G214 seismic station

After the testing phase, continuous waveforms between 2022-01-01 and 2022-01-04 in Groningen Area were downloaded to perform the prediction phase, that is, the phase in which the model is tested on data that has not been used either in the training phase or in the testing phase. In this case, data comes from three seismic stations belonging to the NL Network: G011, G021 and G041. The predicition phase was performed with the following settings:

- BATCH SIZE: 500
- Detection threshold: 0.3
- P threshold: 0.1
- S threshold: 0.1

The number of the detected seismic events for each seismic station are 74 events for G011, 54 events for G021 and 38 events for G041. In Figure 4.14 and 4.15, the detection and picking of one of the seismic events present among the waveforms selected for the seismic stations are shown.

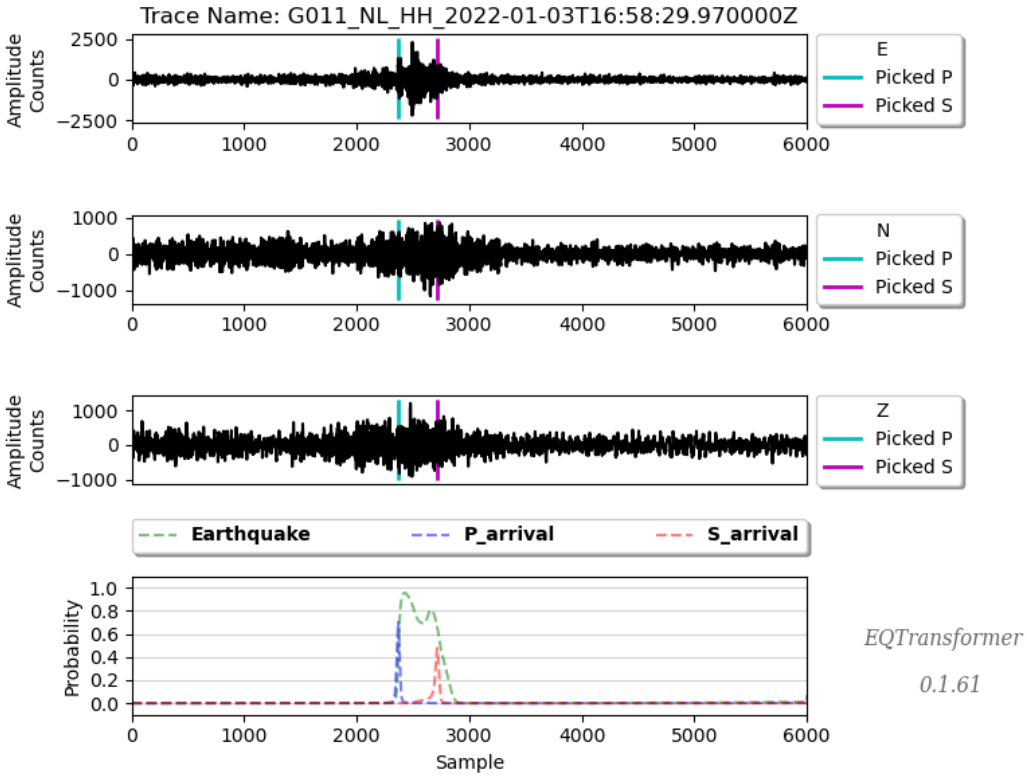



FIGURE 4.14: Example of event predicted in G011 seismic station belonging to the NL network in prediction phase

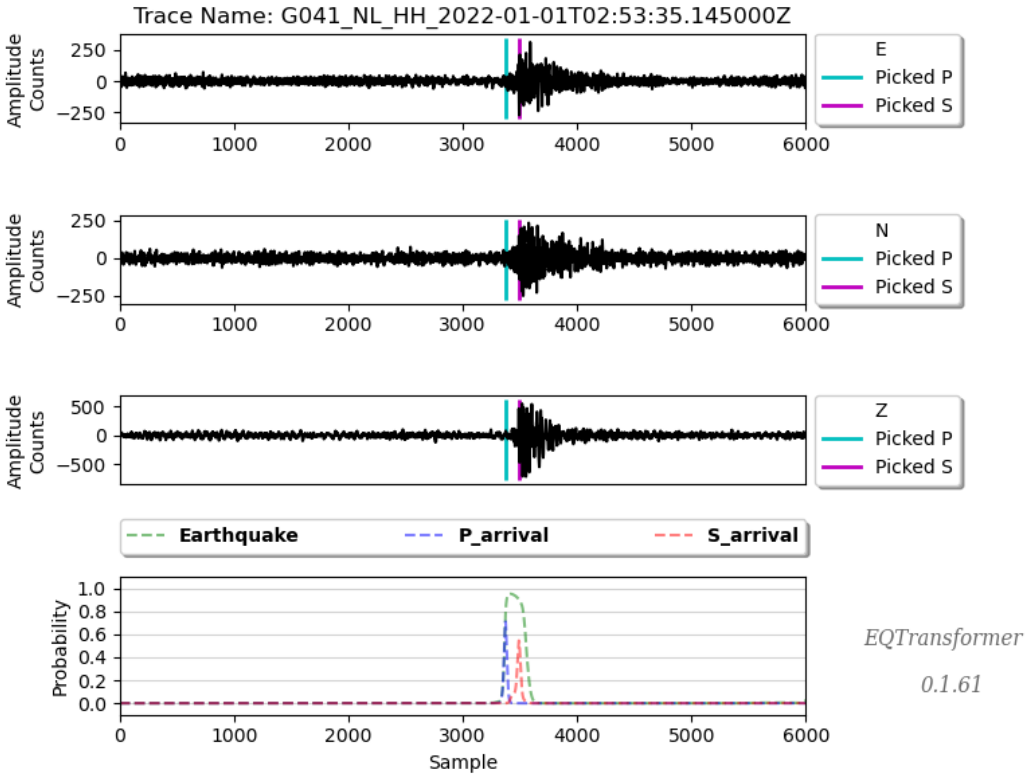



FIGURE 4.15: Example of event predicted in G041 seismic station belonging to the NL network in prediction phase

The performance of EQTransformer in the Testing and Prediction phases has shown that both the detection and picking problems can be addressed with a model based on a transformer architecture for monitoring and early warning purposes, it must certainly be considered that the picking of the S phase remains the major challenge.

### 4.2.2 Regional Approach with EQTransformer: February 2023 seismic sequence in Turkey

In the regional approach, EQTransformer was employed with the purpose of detecting earthquakes and seismic phases' arrival times in the Turkey area. On 6th February 2023, two strong earthquakes occur in the Turkey-Siria area. The first, a magnitude Mw 7.8 earthquake, occurred at 01:17 (UTC) northwest of Gaziantep, causing at least 50,000 deaths. It was followed by a Mw 7.7 earthquake at 10:24. There was extensive damage and tens of thousands of casualties. These earthquakes started a seismic sequence of approximately 8,000 earthquakes in February 2023 (*Disaster and Emergency Management Authority (AFAD)* 2025). Since this seismic sequence is rich in high-magnitude earthquakes, it was considered ideal to push the model to an unexplored magnitude range. This decision was made after noting that the authors of EQTransformer had tested the model using the Mw 6.6 aftershock sequence from the 2000 Tottori (Japan) earthquake, where the maximum recorded magnitude in the sequence was Mw 5.1. Since the area of Japan is not present in the dataset they used for training (Mousavi et al., 2019), it was considered ideal by them to test the model. In this approach, EQTransformer model is pre-trained with the STEAD dataset, the

same that the authors used for their work. It was decided not to retrain it, as it is already rich enough in seismic events (about 450k) for detection on local scale that occurred between January 1984 and August 2018. The magnitude range of STEAD dataset is about -0.5 to 7.9, where small-magnitude earthquakes (less than 2.5) are the majority of the dataset. Using FDSN services, metadata and waveforms were requested from KOERI (Turkish data center) and ORFEUS data centers from February 6th at 00:00:00 UTC to February 9th at 00:00:00 UTC in MSEED format, for the prediction and picking evaluation phases. In other words, the downloaded dataset was used with the main objective of accurately detecting earthquakes, with the model estimating the picking times (in seconds) compared to the manual annotations provided by expert operators. The collected data was recorded by one Turkish seismic station: ANDN. This seismic station is about 100km away from the epicenter of the earthquake at 01:17 UTC, making it ideal for our purposes, as it is situated neither too close nor too far from the area of interest. The following are the technical characteristics of ANDN:

- Network - TU
- Province - Kahramanmaras
- Lat - 37.58
- Lon - 36.3453
- Elevation - 1142
- Type - Velocity

From the continuous waveforms downloaded, each trace of the stream is preprocessed following these steps:

- DETREND: Mean
- BANDPASS FILTER: 1Hz - 45Hz
- TAPER: Cosine
- RESAMPLING: 100Hz (if $samp\_rate \neq 100Hz$)

For the preprocessing of regional seismic events, the same steps as in the local approach were applied, except for the bandpass filter and taper. In this case, a bandpass filter between 1 Hz and 45 Hz was used, which is better suited to the frequency content of regional events. Additionally, a cosine taper was applied to minimize edge effects during filtering.

Firstly, the default parameters of the pre-trained model were used to perform both detection and picking of the seismic phases. Figure 4.16 shows an example of one of the streams present in the collected data and correctly detected as an earthquake from the model. The earthquake shown belongs to the seismic sequence taken into consideration; specifically, the seismic event has ML 5.6 and occured at 01:26 UTC in Gaziantep, Turkey. This test was performed just to verify the potential of the model to perform detection and picking.

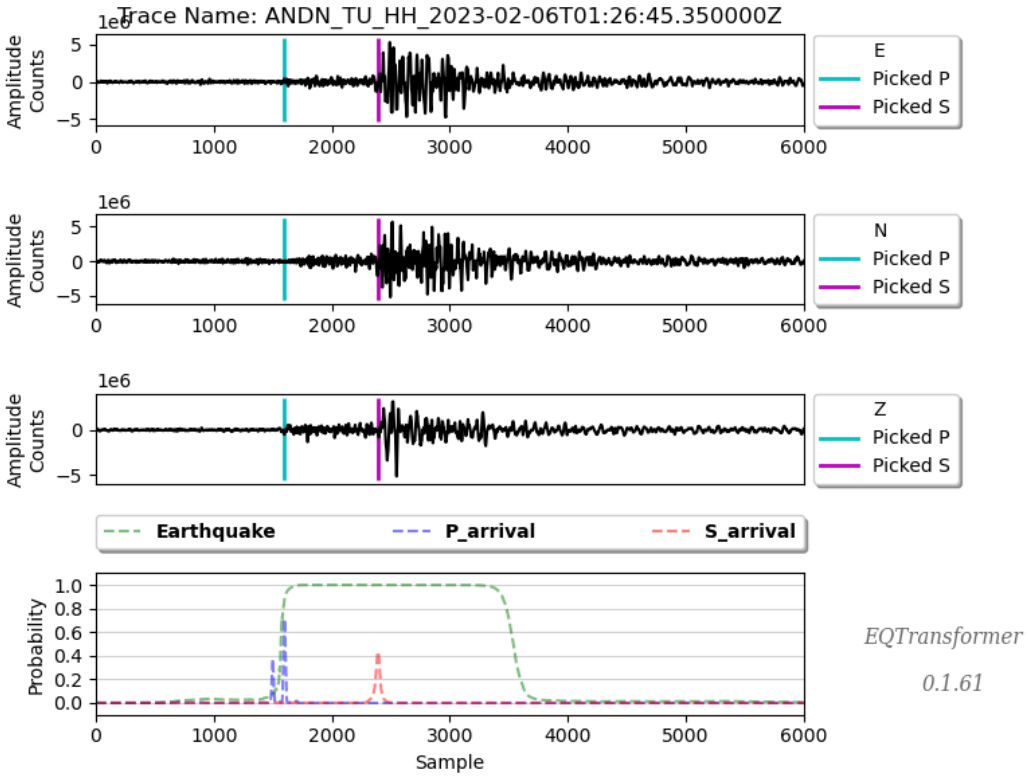



FIGURE 4.16: Earthquake detected by EQTransformer model using waveforms recorded by the ANDN seismic station. ML 5.6 at 01:26:45 UTC. Gaziantep, Turkey (*Disaster and Emergency Management Authority (AFAD)* 2025).

Using a model trained on a dataset (STEAD, in this case) and reused for another similar task is defined as an application of Transfer Learning. This method is particularly effective in scenarios with limited labeled data, because allows the model to retrain and adapt high-level features learned during the initial training phase. For this reason, it was decided to explore more combinations of model parameters to observe how the best performances can be achieved for the picking of Turkey-Siria dataset. As already mentioned in the local approach for the testing phase, the model uses three threshold parameters also for the prediction phase to select the probabilities, in terms of accuracy, related to event detection, picking P-Phase and S-Phase. To explore the best combinations of all three parameters, it was decided to use the evaluation measures, *qi*, *ni* and *qni* (Di Benedetto et al., 2024b), to assess whether the model's picks are accurate, as descripted in detail in 2.3.1. To sum up briefly, *qi* evaluates the quality of each pick by verifying if the absolute residual value falls within a specified threshold k (e.g., k=2). *ni* measures the number of picks, ensuring that the predicted picks do not exceed twice the true picks. The overall metric, *qni*, is calculated as the product of *qi* and *ni*, providing a balanced assessment of both the quality and the numerosity of the picks.
The EQTransformer runs that have been performed, takes account of the values of Detection threshold and P Phase threshold to show the performance of the model for the picking of the P-Phase. Probability thresholds use values ranging from 0.1 to 0.9. S-Phase threshold was fixed with the default value used by the model: 0.1; This is because picking performance has been focused on the P phase. The Batch size was

left with the default value of the model, that is 500, the same value used in the local approach. To better illustrate the approach, Table 4.1 shows the EQTransformer runs performed with defined parameter thresholds:

TABLE 4.1: EQTransformer runs with the combination of the threshold parameters: Detection threshold, P-Phase threshold and S-Phase threshold

| Detection Threshold | P-Phase Threshold | S-Phase Threshold |
|---|---|---|
| 0.1 | 0.1 | 0.1 |
| 0.1 | 0.2 | 0.1 |
| 0.1 | 0.3 | 0.1 |
| 0.1 | ... | 0.1 |
| 0.1 | 0.9 | 0.1 |
| 0.2 | 0.1 | 0.1 |
| 0.2 | 0.2 | 0.1 |
| 0.2 | 0.3 | 0.1 |
| ... | ... | 0.1 |
| 0.2 | 0.9 | 0.1 |
| ... | ... | 0.1 |
| ... | ... | 0.1 |
| 0.9 | 0.1 | 0.1 |
| 0.9 | 0.2 | 0.1 |
| 0.9 | 0.3 | 0.1 |
| 0.9 | ... | 0.1 |
| 0.9 | 0.9 | 0.1 |

The results are shown in Figures 4.17 and 4.18 for *qi* and *ni*, respectively, and in Figure 4.19 for *qni*. Every plot has on x-axis the values of the Detection threshold and on y-axis the values of the P-Phase threshold. In the colorbar, the measures are represented as coloured circles: the darker the colour, the higher the value. Just to better show the results, Figure 4.17 represents *qi* values from the minimum value extracted, that is 0.77 to the maximum, 0.83; instead, Figure 4.18 and 4.19 from 0.0 to 1.0 (range of values for these measures).

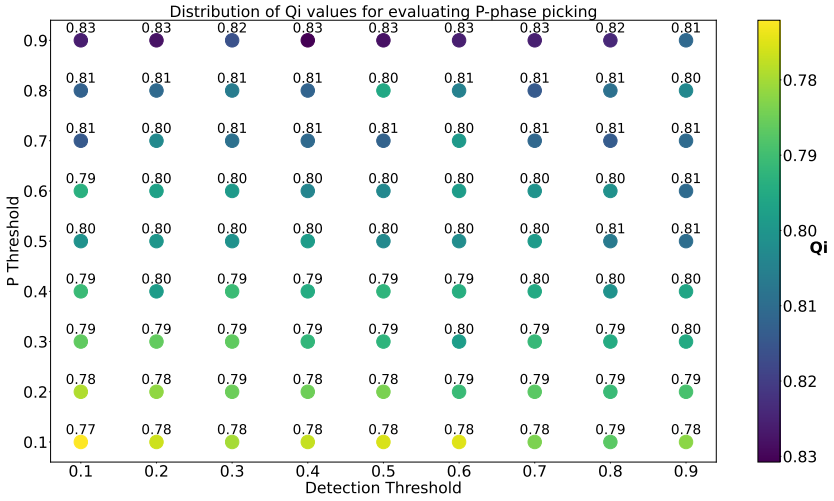



FIGURE 4.17: *qi* measure used for evaluating P-Phase picking performance of the model EQTransformer for a Transfer Learning problem

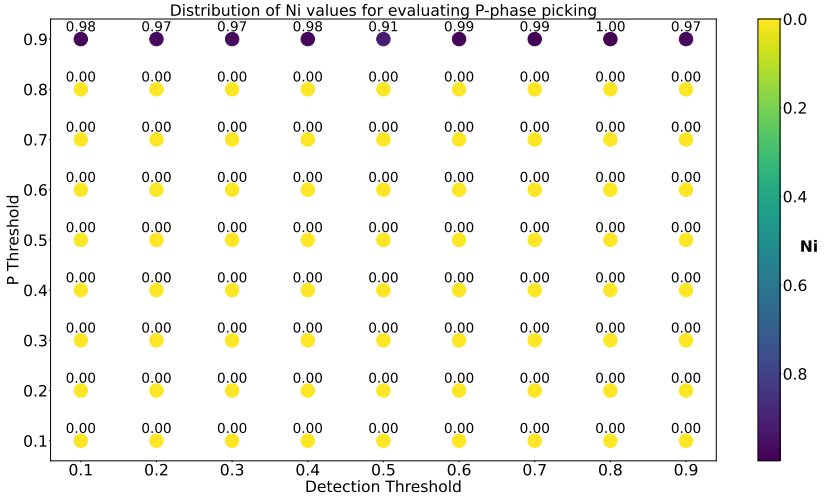



FIGURE 4.18: *ni* measure used for evaluating P-Phase picking performance of the model EQTransformer for a Transfer Learning problem

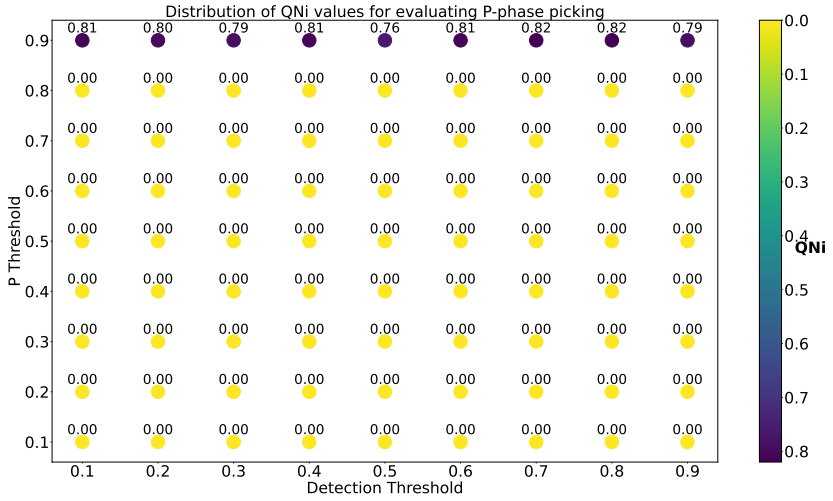



FIGURE 4.19: *qni* measure used for evaluating P-Phase picking performance of the model EQTransformer for a Transfer Learning problem

Each plot show that as the probability threshold increases, the values of the respective measures improve, as can be seen in the range of P threshold values between 0.6 and 0.9. This behaviour is particularly visible on the y-axis. For *qi*, this happens because by decreasing the number of seismic events as false positives, the number of deviations decreases and therefore the average of the latter increases. The number of false positive detected is strongly influenced by all the parameters. For instance, if the threshold for picking S Phase had assumed higher values in principle (such as greater than 0.5), the number of detected events would be significantly reduced, since for S Phase the uncertainty about picking increases (thereby decreasing the associated probabilities). A similar behavior occurs for *ni*, where the number of detected events significantly influences the result of this measure, and therefore also for *qni*; in fact, it is enough to note that by increasing the probability threshold associated with the P phase by only 0.1, *ni* assumes values between 0.91 and 1.0, as the number of detected events decreases drastically. *qni* simply shows the result of the product between the two measures *qi* and *ni*, in fact the values different from 0 correspond to the threshold for the P phase of 0.9; the variation of the values on the x-axis does not significantly affect the value assumed by the measure. In general, The figures show an increase in the values of *qi*, *ni* and *qni* as the detection threshold and P-Phase threshold increase.

To further investigate the results of the *qni* values for the explored range, it was decided to verify the variability of the qni values for the range from 0.80 for the detection probabilities to 1.0 for the probabilities associated with the P-Phase, at a step of 0.01. Figure 4.20 shows this result.

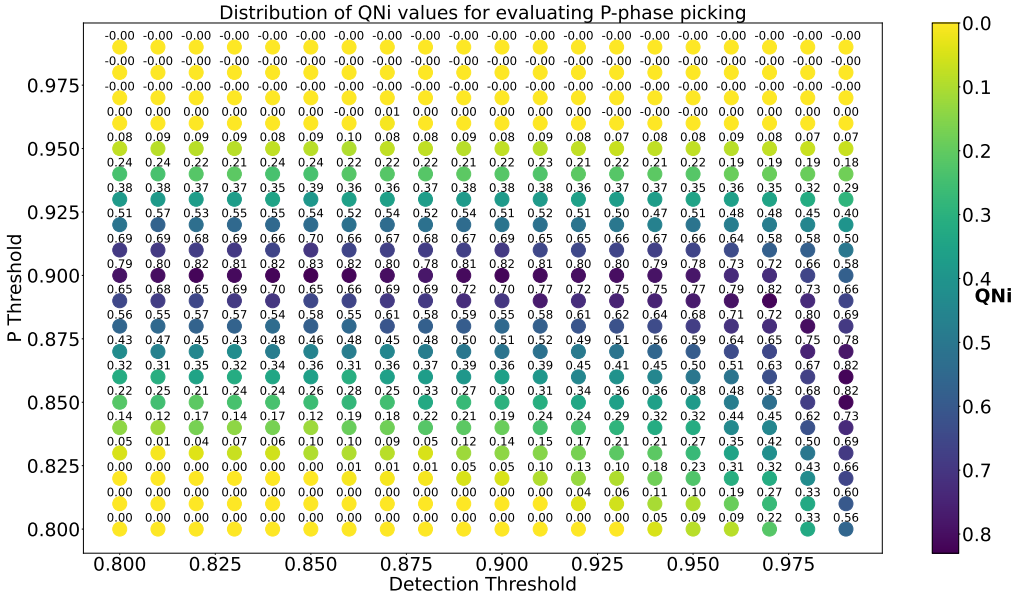



FIGURE 4.20: *qni* measure used for evaluating P-Phase picking performance of the model EQTransformer using finely spaced probability values

In this case, finely spaced probability values was used, showing more variability on the distribution. The grid here is 20x20, with a total of 400 *qni*s. The variability on the values shows higher *qni* values, concentrated in the central and upper-right regions, suggest that optimal performance is achieved when both detection and picking probabilities are moderately high. Conversely, lower values are found on the top, implying that poorer phase identification at those configuration parameters fails because above the 0.95 probability threshold, there are no events detected in that range.

# Chapter 5

# Discussion

Trying to solve the problem of earthquake detection and seismic phase picking is crucial to improve the field of earthquake monitoring and early warning. In this PhD Thesis, several methodologies and models were proposed for resolving these problems.
The grid-search trying to resolve the problem of the parameter's tuning for a method that, in this case, is STA/LTA (Allen, 1978). As the objective of this PhD thesis, this method, trying to give a resolution to the picking problem using parameters like STA, LTA, Trigger on and Trigger off. Among the parameter, the lengths for the STA and LTA windows depend on the frequency content of the seismogram. Long-period records require larger averaging windows than short-period records, which require shorter averaging windows. This means that the tuning of the parameters strongly depends on the nature of the data being analyzed; this is the reason why a grid-search helps in identifying the optimal configuration that best fits the specific requirements of the task.
Robust automatic techniques for seismo-volcanic events classification are of fundamental importance when studying active volcanoes, both for the importance of an accurate real-time seismic monitoring procedure and for the use of their results in modelling the dynamics of the volcanic system. The Active Learning approach proposed in this thesis, demonstrates high efficiency for the binary classification problem characterized by the Explosion Quakes and No Explosion Quakes classes. Moreover, this approach enables the visualization and analysis of waveforms, as well as their appropriate editing to enhance the EQs dataset. The creation and use of hybrid events with a specific method provided a real fingerprint to the dataset to demonstrate that expert operator intervention can increase the efficiency of the classifier.
The evaluation metrics used in the grid-search application, namely Quality Index (Qi), Numerosity Index (Ni) and their combined measure Quality Numerosity index (QNi), were found to be relevant to evaluate the accuracy of P-phase picking performed by the EQTransformer model. These metrics provided valuable insights into the model performance, helping to refine and optimize the detection process by parameter's fine-tuning.
Two approaches were employed for the EQTransformer (Mousavi et al., 2020) model: a local approach applied to the Groningen area (the Netherlands), and a regional approach focused on the Turkey–Syria region. In the local approach, the model was trained to recognize seismic events with low magnitudes, i.e., below 3.0. This is particularly useful in improving seismic monitoring in that specific area, where low-magnitude seismicity is frequent and relevant for hazard assessment. Furthermore, a model trained on a local dataset can significantly reduce false positives and false negatives, leading to more reliable earthquake detection and phase picking. However, such models are often highly tuned to the specific seismic characteristics of the

area, which increases the risk of overfitting. As a result, their performance may degrade when applied to different regions with distinct geological and seismic properties. In contrast, the regional approach targets seismic events with wider magnitude range, where high magnitudes are present, i.e. between 2.0 and 7.7, and is trained on data covering a broader geographic area, specifically a global dataset (Mousavi et al., 2019). This allows the model to generalize better and detect earthquakes across multiple locations and fault zones. Although regional models may show slightly reduced accuracy for low-magnitude events compared to locally optimized models, they offer the advantage of greater robustness and applicability in heterogeneous seismic environments. Additionally, regional models are better suited for operational use in early warning systems and large-scale monitoring networks, where flexibility and adaptability are essential. From a practical perspective, the choice between a local and a regional approach has direct implications on the effectiveness and scalability of seismic monitoring and early warning systems. In the local approach, the finely tuned sensitivity allows for enhanced detection of low-magnitude events, which are often missed by broader models but are crucial in areas of anthropogenic seismicity, such as induced seismicity due to gas extraction or geothermal activity—as is the case in Groningen. The improved accuracy in both event detection and phase picking enhances the resolution of local seismic catalogs, allowing for better tracking of swarm activity, foreshock sequences, or evolving seismic patterns over time. However, the use of such a specialized model in an operational early warning context is limited: while it can support high-resolution local monitoring, it lacks the generalization capacity required for real-time application in unfamiliar regions or during large-scale emergencies. Furthermore, its reduced performance on non-local data may compromise the robustness needed for issuing timely alerts in multi-region systems. The regional approach, by contrast, favors generalization over specialization. Trained on a global dataset, it can operate effectively across different tectonic settings and seismic source characteristics. This makes it more suitable for national or transnational monitoring networks, particularly in seismically active regions with multiple fault systems like Turkey and Syria. In the context of early warning systems, a regional model's ability to detect moderate-to-large events rapidly and across a wide area is a crucial advantage. While its sensitivity to low-magnitude events may be lower than that of a local model, this is generally acceptable for early warning, where the primary goal is rapid identification of potentially damaging earthquakes. Additionally, regional models can be deployed more flexibly, providing a scalable solution for countries or regions with limited local training data or sparse instrumentation. In summary, local models are best suited for high-resolution, area-specific monitoring and scientific analysis, especially where small-magnitude seismicity is of interest. Regional models, on the other hand, are more practical for real-time seismic surveillance and early warning operations over wide geographical areas. An ideal strategy may involve a hybrid approach, where local models are embedded within a broader regional framework, allowing systems to dynamically switch between models based on geographic context and the type of seismic event.

# Chapter 6

# Conclusions

This PhD thesis highlights the importance of exploring different methodological approaches and models to effectively address the challenges of earthquake detection and seismic phase picking. By integrating a grid-search optimization strategy for STA/LTA parameter tuning, adapted to the frequency content of the analyzed waveforms, allowed for the exploration of parameter spaces to maximize detection and picking accuracy. The introduction of the Quality Index (Qi), Numerosity Index (Ni), and the overall measure Quality Numerosity index (QNi) provided a quantitative basis for evaluating picking performance both in a volcanic context, specifically on Explosion Quakes (EQs) and in a tectonic context for local and regional earthquakes, using innovative AI models like EQTransformer. Still focusing on volcanic context, the Active Learning approach enabled efficient binary classification between Explosion and Non-Explosion Quakes by reducing the intervention of manual labeling by the human operator while maintaining high performance. It also allowed a more precise cutting of waveforms, improving dataset quality and consistency. Furthermore, the inclusion of hybrid events provided a controlled benchmark, simulating a more realistic dataset where EQs and NEQs were merged.
The comparative analysis between local and regional approaches with the EQTransformer model, highlights the balance between sensitivity and generalization. While local models offer high-resolution insights for specific regions with low-magnitude seismicity, regional models provide broader applicability and robustness for large-scale monitoring and early warning systems. Ultimately, this work suggests that the most effective seismic monitoring strategy is a hybrid one—combining the precision of local models with the scalability of regional ones to create a flexible and reliable system for earthquake surveillance, adaptable to both scientific research and operational needs.

# Bibliography


Allen, Rex V. (Oct. 1978). "Automatic earthquake recognition and timing from single traces". In: *Bulletin of the Seismological Society of America* 68.5, pp. 1521–1532. ISSN: 0037-1106. DOI: 10.1785/BSSA0680051521. URL: https://doi.org/10.1785/BSSA0680051521.

Andronico, Daniele et al. (July 2021). "Uncovering the eruptive patterns of the 2019 double paroxysm eruption crisis of Stromboli volcano". In: *Nature Communications* 12, p. 4213. DOI: 10.1038/s41467-021-24420-1.

Bai, Chao-ying and Brian Kennett (Apr. 2001). "Phase identification and attribute analysis of broadband seismograms at far-regional distances". In: *Journal of Seismology - J SEISMOL* 5, pp. 217–231. DOI: 10.1023/A:1011436421196.

Belete, Daniel and Manjaiah D H (Sept. 2021). "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results". In: *International Journal of Computers and Applications* 44, pp. 1–12. DOI: 10.1080/1206212X.2021.1974663.

Beyreuther, Moritz et al. (May 2010). "ObsPy: A Python Toolbox for Seismology". In: *Seismological Research Letters* 81.3, pp. 530–533. ISSN: 0895-0695. DOI: 10.1785/gssrl.81.3.530. eprint: https://pubs.geoscienceworld.org/ssa/srl/article-pdf/81/3/530/2762059/530.pdf. URL: https://doi.org/10.1785/gssrl.81.3.530.

Bischl, Bernd et al. (2023). "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges". In: *WIREs Data Mining and Knowledge Discovery* 13.2, e1484. DOI: https://doi.org/10.1002/widm.1484. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1484. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1484.

Bishop, C.M. (2006). *Pattern recognition and machine learning*. Vol. 4. Springer New York. URL: http://scholar.google.com/scholar.bib?q=info:jYxggZ6Ag1YJ:scholar.google.com/&output=citation&hl=en&as_sdt=0,5&as_vis=1&ct=citation&cd=0.

Breiman, Leo (Oct. 2001). "Random Forests". English. In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324.

Breiman, Leo et al. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

Chambefort, Mathieu et al. (2022). "Improving the quality control of seismic data through active learning". In.

Di Benedetto, Andrea et al. (2024a). *Earthquake Detection and Phase Picking using Deep Learning method: a case study on local events (Groningen Region, Netherlands)*. URL: https://www.cs.ingv.it/collane/index.php/mi/article/view/678.

Di Benedetto, Andrea et al. (2024b). "Grid-search method for short-term over long-term average parameter tuning: an application to Stromboli explosion quakes". In: *Frontiers in Earth Science* 12. ISSN: 2296-6463. DOI: 10.3389/feart.2024.1440967. URL: https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2024.1440967.

Diaz-Uriarte, Ramon and Sara Alvarez de Andrés (Feb. 2006). "Gene Selection and Classification of Microarray Data Using Random Forest". In: *BMC bioinformatics* 7, p. 3. DOI: 10.1186/1471-2105-7-3.

*Disaster and Emergency Management Authority (AFAD)* (2025). Accessed: 2025-02-04. URL: \url{https://deprem.afad.gov.tr}.

Dost, Bernard, Elmer Ruigrok, and Jesper Spetzler (2017). "Development of seismicity and probabilistic hazard assessment for the Groningen gas field". In: *Netherlands Journal of Geosciences* 96.5, s235–s245. DOI: 10.1017/njg.2017.20.

D'Alessandro, Antonino et al. (2022). "An Active Learning Approach for Classifying Explosion Quakes". In: *2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pp. 1–6. DOI: 10.1109/EAIS51927.2022.9787519.

Earle, Paul S. and Peter M. Shearer (Apr. 1994). "Characterization of global seismograms using an automatic-picking algorithm". In: *Bulletin of the Seismological Society of America* 84.2, pp. 366–376. ISSN: 0037-1106. DOI: 10.1785/BSSA0840020366. URL: https://doi.org/10.1785/BSSA0840020366.

Fenner, Darius et al. (2021). "AWESAM: A Python Module for Automated Volcanic Event Detection Applied to Stromboli". In: arXiv: 2111.01513 [physics.geo-ph]. URL: https://arxiv.org/abs/2111.01513.

Fenner, Darius et al. (2022). "Automated Seismo-Volcanic Event Detection Applied to Stromboli (Italy)". In: *Frontiers in Earth Science* 10. ISSN: 2296-6463. DOI: 10.3389/feart.2022.809037. URL: https://www.frontiersin.org/articles/10.3389/feart.2022.809037.

Friedman, Jerome H (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics*, pp. 1189–1232.

Gentili, S. and A. Michelini (Jan. 2006). "Automatic picking of P and S phases using a neural tree". In: *Journal of Seismology* 10. ISSN: 1573-157X. DOI: 10.1007/s10950-006-2296-6. URL: https://doi.org/10.1007/s10950-006-2296-6.

Jones, Joshua and Mirko Baan (May 2015). "Adaptive STA–LTA with outlier statistics". In: *Bulletin of the Seismological Society of America* 105. DOI: 10.1785/0120140203.

Küperkoch, Ludger et al. (May 2010). "Automated determination of P-phase arrival times at regional and local distances using higher order statistics". In: *Geophysical Journal International* 181, pp. 1159–1170. DOI: 10.1111/j.1365-246X.2010.04570.x.

Li, Ang et al. (2019). "A Generalized Framework for Population Based Training". In: KDD '19, 1791–1799. DOI: 10.1145/3292500.3330649. URL: https://doi.org/10.1145/3292500.3330649.

Maggi, Alessia et al. (July 2009). "An automated time-window algorithm for seismic tomography". In: *Geophysical Journal International* 178. DOI: 10.1111/j.1365-246X.2009.04099.x.

Manley, Grace F. et al. (2022). "A Deep Active Learning Approach to the Automatic Classification of Volcano-Seismic Events". In: *Frontiers in Earth Science* Volume 10 - 2022. ISSN: 2296-6463. DOI: 10.3389/feart.2022.807926. URL: https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2022.807926.

Mousavi, S.Mostafa et al. (Oct. 2019). "STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI". In: *IEEE Access* PP, pp. 1–1. DOI: 10.1109/ACCESS.2019.2947848.

Mousavi, S.Mostafa et al. (Aug. 2020). "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking". In: *Nature Communications* 11, p. 3952. DOI: 10.1038/s41467-020-17591-w.

Münchmeyer, Jannes et al. (Jan. 2022). "Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers". In: *Journal of Geophysical Research: Solid Earth* 127.1. ISSN: 2169-9356. DOI: 10.1029/2021jb023499. URL: http://dx.doi.org/10.1029/2021JB023499.

Pal, Mahesh (Jan. 2005). "Random forest classifier for remote sensing classification". In: *International Journal of Remote Sensing - INT J REMOTE SENS* 26, pp. 217–222. DOI: 10.1080/01431160412331269698.

Pedregosa, Fabian et al. (Nov. 2011). "Scikit-Learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12.null, 2825–2830. ISSN: 1532-4435.

QGIS Development Team (2024). "QGIS Geographic Information System". In: URL: https://www.qgis.org.

Richardson, Alan (2021). "Active learning for seismic processing parameterisation, with an application to first break picking". In.

Ross, Zachary E. et al. (Aug. 2018). "Generalized Seismic Phase Detection with Deep Learning". In: *Bulletin of the Seismological Society of America* 108.5A, 2894–2901. ISSN: 1943-3573. DOI: 10.1785/0120180080. URL: http://dx.doi.org/10.1785/0120180080.

Schapire, Robert E (2013). "Explaining adaboost". In: *Empirical inference*. Springer, pp. 37–52.

Settles, Burr (2012). *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.

Titos, Manuel et al. (2018). "A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.5, pp. 1533–1544. DOI: 10.1109/JSTARS.2018.2803198.

Unzueta, Diego, ed. (2022). *Transformers: An Overview of the Most Novel AI Architecture*. URL: \url{https://medium.com/towards-data-science/transformers-an-overview-of-the-most-novel-ai-architecture-cdd7961eef84}.

Vaswani, Ashish et al. (2023). "Attention Is All You Need". In: arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

Wang, Hongtao et al. (Apr. 2024). "MSSPN: Automatic first-arrival picking using a multistage segmentation picking network". In: *GEOPHYSICS* 89.3, U53–U70. ISSN: 1942-2156. DOI: 10.1190/geo2023-0110.1. URL: http://dx.doi.org/10.1190/geo2023-0110.1.

Wassermann, Joachim (Jan. 2012). "Volcano Seismology, IASPEI New manual of seismological observatory practice 2 (NMSOP-2)". In: pp. 1–77.

Wenner, M. et al. (2021). "Near-real-time automated classification of seismic signals of slope failures with continuous random forests". In: *Natural Hazards and Earth System Sciences* 21.1, pp. 339–361. DOI: 10.5194/nhess-21-339-2021. URL: https://nhess.copernicus.org/articles/21/339/2021/.

Withers, Mitch et al. (Feb. 1998). "A comparison of select trigger algorithms for automated global seismic phase and event detection". In: *Bulletin of the Seismological Society of America* 88, pp. 95–106. DOI: 10.1785/BSSA0880010095.

Yoma, Nestor Becerra et al. (2021). "End-to-end LSTM based estimation of volcano event epicenter localization". In: arXiv: 2110.14594 [eess.SP]. URL: https://arxiv.org/abs/2110.14594.

Zhu, Weiqiang and Gregory C Beroza (Oct. 2018). "PhaseNet: a deep-neural-network-based seismic arrival-time picking method". In: *Geophysical Journal International* 216.1, pp. 261–273. ISSN: 0956-540X. DOI: 10.1093/gji/ggy423. eprint: https://academic.oup.com/gji/article-pdf/216/1/261/26329430/ggy423.pdf. URL: https://doi.org/10.1093/gji/ggy423.

Zhu, Weiqiang et al. (2022). "An End-To-End Earthquake Detection Method for Joint Phase Picking and Association Using Deep Learning". In: *Journal of Geophysical Research: Solid Earth* 127.3. e2021JB023283 2021JB023283, e2021JB023283. DOI: https://doi.org/10.1029/2021JB023283.