



Università
di Catania

DEPARTMENT OF ECONOMICS AND BUSINESS

DEPARTMENT OF ELECTRICAL ELECTRONIC AND COMPUTER ENGINEERING

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

MASTER'S DEGREE IN DATA SCIENCE

*Quake Prediction on Etna:
A Multi-Parametric
Deep Learning Approach*

Giuseppe Leonardi

INGV Co-Supervisors:
Dr. Giuseppe SALERNO
Dr. Danilo REITANO

Supervisor:
Prof. Luca GUARNERA

Co-Supervisor:
Prof. Sebastiano BATTIATO

Academic Year 2024/2025

December 18, 2025

Giuseppe LEONARDI

SCILIAE STUDIVM GENERALE
QUAKE PREDICTION ON ETNA:
A MULTI-PARAMETRIC
DEEP LEARNING APPROACH

Master's Degree Thesis

1434

UNIVERSITY OF CATANIA

December 2025

“Deep learning is not magic; it’s just very powerful statistical pattern recognition.”

Ian Goodfellow

Abstract

The increasing availability of high-resolution geophysical datasets has opened new opportunities for improving short-term earthquake forecasting in volcanic regions. This thesis investigates the potential of machine learning methods to analyze and interpret seismic and related geophysical signals from Mount Etna, with a specific focus on predicting local earthquakes during 2018, a year marked by intense seismic and volcanic activity. The study evaluates the predictive capability of several models, ranging from traditional statistical techniques to modern deep learning architectures. After careful preprocessing and feature engineering, these algorithms were applied to time-series data to estimate the probability of future seismic events. Among the tested approaches, Long Short-Term Memory (LSTM) networks achieved the best performance, effectively capturing temporal dependencies and nonlinear patterns in the seismic tremor signal. The results highlight the value of deep learning in supporting short-term earthquake prediction at active volcanoes. The proposed framework demonstrates how data-driven models can augment conventional seismological analyses, contributing to more robust monitoring strategies in complex volcanic environments.

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	1
2 State-of-the-Art for Etna Monitoring	7
2.1 Case Studies in ML for Mount Etna	8
2.2 Concluding Remarks	13
3 Dataset and Preprocessing	15
3.1 Geophysical Data	16
3.2 Gas Emission Fluxes and Ratios	21
3.3 Data Insights	25
4 Methodology and Predictive Framework	27
4.1 Data Preparation and Feature Engineering	28
4.2 Baseline Model: Random Forest	30
4.3 Sequential Model: LSTM Architecture	31
4.4 Evaluation Protocol and Summary	33
5 Results	35
5.1 Baseline Performance: Random Forest	36
5.2 Sequential Model Performance: LSTM	38
5.3 Extended Architecture: Multi-LSTM Model	40
5.4 Comparative Analysis and Summary	42
6 Conclusions and Future Work	45
A Pseudocode of the Pipeline	49
B LSTM Networks	53
Bibliography	55

List of Figures

1.1	Etna Location	2
1.2	Representation of a multi-sensor network	3
2.1	Infrasound waveform and spectrogram	8
2.2	Temporal evolution of eruptive clusters	8
2.3	DT vs KNN results	9
2.4	Classification estimation time series	10
2.5	Example thermal classes	11
2.6	Convolutional neural network architecture	11
2.7	Lava flows and activity map (2005–2009)	12
2.8	Random Forest confusion matrix	13
3.1	Timeline of Etna’s activity	17
3.2	Clinometric correlations	18
3.3	Clinometric movement index	19
3.4	Spatial density of earthquakes	20
3.5	Temporal evolution of gas emission fluxes	22
3.6	Temporal evolution of the CO ₂ /SO ₂ ratio	23
3.7	Temporal evolution of the ³ He/ ⁴ He ratio	24
3.8	Seismic clustering by depth and magnitude	25
4.1	Predictive pipeline	29
4.2	RF feature importance	31
5.1	Random Forest performance visualization	37
5.2	LSTM training and validation loss	38
5.3	LSTM predicted vs observed seismic activity	39
5.4	Multi-LSTM training and validation loss	41
5.5	Comparison of model confusion matrices	42
B.1	Cell of LSTM	54

List of Tables

1.1	Instruments of a multi-sensor network	3
3.1	Sample of the tremor dataset	17
3.2	Sample of the clinometric dataset	17
3.3	Sample of the seismicity dataset	19
3.4	Sample of the heat flux dataset	21
4.1	Dataset temporal partitioning	30
4.2	Random Forest hyperparameters	31
4.3	LSTM hyperparameters	32
4.4	Metrics of evaluation	34
5.1	Random Forest performance metrics	36
5.2	LSTM performance metrics	38
5.3	LSTM configuration comparison	40
5.4	Multi-LSTM performance metrics	41
5.5	Model performance comparison	42

Chapter 1

Introduction

Mount Etna, rising on the eastern coast of Sicily, is one of the most active volcanoes on the planet. Known for its spectacular eruptions and persistent activity, it dominates both the landscape and the daily life of the surrounding region. The volcano extends over an area of more than 1,200 square kilometers and reaches an elevation exceeding 3,300 meters, making it the highest active volcano in continental Europe. Its flanks host towns, forests and ancient lava flows that testify to millennia of eruptive history and human adaptation to a restless environment. Its behavior is characterized by an extraordinary diversity: from quiet lava effusions and Strombolian bursts to powerful paroxysmal eruptive events and destructive flank eruptions. Over the years, Etna's eruptions have alternated between summit activity, confined to the central craters and lateral intrusions that can extend for several kilometers, cutting through populated areas. This variability makes Etna not only a permanent threat to nearby communities but also an exceptional natural observatory for studying the dynamics of magmatic systems and the complex interplay between deep processes and surface manifestations. From a scientific standpoint, Etna is a cornerstone for modern volcanology. It has been continuously monitored for decades, providing an unparalleled record of the physical and chemical processes driving its activity. Its accessibility, combined with the frequency and variety of its eruptive styles, has made it a reference site for testing new observational methods and theoretical models of volcanic behavior. The volcano's persistent state of unrest, marked by fluctuations in seismic tremor, recurrent degassing episodes, ground deformation and sporadic explosive activity, allows scientists to observe the entire spectrum of basaltic volcanic phenomena within relatively short time spans. These characteristics make Etna an ideal natural laboratory for developing and validating models of eruption forecasting and for designing multidisciplinary tools capable of bridging the gap between fundamental research and operational early-warning systems.

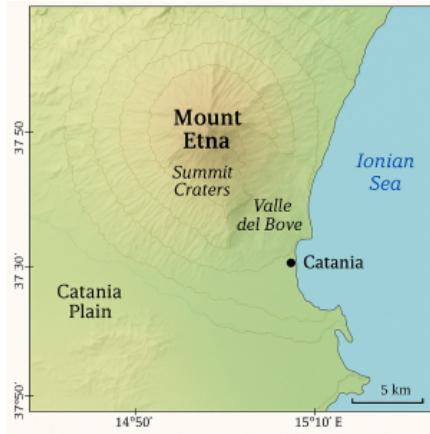


FIGURE 1.1: The volcano rises on the eastern coast of Sicily, between the Ionian Sea and the Catania Plain, reaching an elevation of over 3300 meters.

The Italian National Institute of Geophysics and Volcanology (INGV), through its *Osservatorio Etneo* (INGV–OE) in Catania, operates one of the most sophisticated volcanic monitoring networks in the world. The network integrates multiple observational techniques, combining seismic, geochemical and satellite measurements into a comprehensive system capable of capturing the volcano’s evolution from multiple perspectives. Each monitoring device offers a partial but complementary view of the underlying processes. The quantity and heterogeneity of data produced by a network like this are both a scientific opportunity and a computational challenge. Volcanic signals are complex: they are non-linear, multi-scale and often influenced by coupled processes occurring at different depths. For instance, an increase in gas emissions may coincide with a change in seismic tremor amplitude, or a subtle ground tilt may anticipate an eruptive episode by only a few hours. Interpreting these patterns traditionally requires expert knowledge and manual correlation among diverse parameters. An approach that, while effective in individual case studies, struggles to keep pace with the continuous, high-frequency data streams now available.

In recent years, the rapid evolution of data science has introduced new perspectives to volcanology. Machine learning techniques, in particular neural networks designed to model temporal sequences, offer the ability to automatically learn relationships between multiple signals, detect anomalies and recognize precursory trends that might escape conventional analysis. These approaches do not aim to replace physical understanding, but rather to extend it, allowing the discovery of empirical regularities that can inform models of magmatic behavior and enhance early-warning systems.

Instrument	Measured Parameter	Notes
Seismometer	Seismic activity	Detects earthquakes and volcanic tremor.
Tiltmeter	Tilt angle	Sensitive to shallow magmatic pressure variations.
Gravimeter	Gravity variations	Detects changes in magma mass and density.
Gas spectrometer	Gas fluxes and ratios	Measures SO ₂ , CO ₂ , HCl, CO ₂ /SO ₂ and ³ He/ ⁴ He.
Satellite sensors	Deformation and thermal anomalies	Provides remote sensing data (optical and radar).

TABLE 1.1: Main monitoring instruments and parameters measured within INGV–OE’s multi-sensor network.

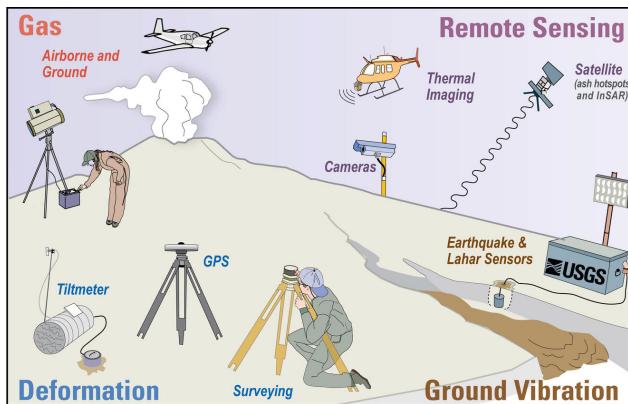


FIGURE 1.2: Schematic representation of a multi-parameter volcanic monitoring network. (source [1])

As summarized in Table 1.1 and illustrated in Figure 1.2, the monitoring of Etna relies on a dense and integrated multi-sensor network. Each instrument contributes a specific layer of information: seismic and clinometric stations capture ground motion and internal vibrations, Tiltmeters detect deformation patterns related to pressure changes at depth, while gas detectors and remote-sensing systems reveal variations in degassing and surface thermal activity. The combination of these complementary data streams provides a comprehensive view of the volcano’s dynamics, linking subsurface processes to their geophysical and geochemical manifestations at the surface.

This thesis focuses this emerging paradigm of data-driven volcanology. It explores how machine learning can be used to analyze multi-parametric monitoring data from Etna, focusing on the identification of short-term precursors to increased seismic activity. The study covers the period between **January 2017 and February 2019**, a window that encompasses distinct eruptive phases and quiescent intervals. In particular, the analysis emphasizes the period leading up to the **December 24–26, 2018 flank eruption**, one of the most significant events of the decade. This eruption, preceded by an intense seismic swarm and rapid ground deformation, provides a

well-documented benchmark for testing predictive algorithms and assessing the temporal coherence of geophysical and geochemical precursors. The research adopts a progressive and comparative approach. After rigorous data cleaning, synchronization and normalization of all monitoring datasets, several models were developed and tested. The first stage involved a **Random Forest (RF)** baseline classifier, designed to evaluate the predictive power of static statistical features aggregated over fixed time windows. This was followed by the implementation of a **Long Short-Term Memory (LSTM)** neural network, specifically tailored to exploit the temporal dependencies inherent in sequential data. Finally, a **multi-input LSTM architecture** was developed to integrate geophysical and geochemical domains into a unified model capable of learning their dynamic interactions. Each model was trained and validated using temporally consistent splits, ensuring that testing occurred on truly unseen data corresponding to independent phases of Etna’s activity. While the intrinsic unpredictability of volcanic systems limits the extent of deterministic forecasting, the results of this work demonstrate that data-driven models can achieve meaningful predictive capability. Even small improvements in classifying transitions between quiescent and active states can substantially enhance operational monitoring, particularly when integrated with expert interpretation. Furthermore, the interpretability analysis of the LSTM models reveals coherent patterns linking increases in tremor energy, variations in clinometric tilt and elevated gas fluxes to the onset of seismic swarms, a promising indication that these models capture physically relevant features rather than random correlations. Beyond performance metrics, this research contributes to a broader conceptual evolution in volcanic monitoring. It illustrates how artificial intelligence can act as a bridge between raw sensor data and scientific understanding, transforming massive time series into interpretable indicators of volcanic state and earthquake. By merging statistical inference with physical reasoning, data-driven volcanology moves toward a new generation of early-warning frameworks that are both adaptive and explainable.

The long-term vision underpinning this work is not to achieve perfect prediction (an unrealistic goal given the inherent complexity of these events), but to reduce uncertainty and improve situational awareness. Each eruption forecast is, ultimately, a probabilistic statement about the future. Enhancing its reliability, even marginally, can provide critical minutes or hours for decision-makers and civil protection agencies. The integration of machine learning into volcanic observatories thus represents a step toward more intelligent and science-based hazard mitigation systems. In summary, this thesis aims to contribute to the ongoing dialogue between traditional volcanological methods and modern computational approaches. It leverages the exceptional dataset of Etna’s multi-sensor network to explore how advanced models can learn from the volcano’s own rhythms and transform them into knowledge. The ultimate aspiration is to move closer to a holistic understanding of how the volcano communicates its internal state, blending human expertise with algorithmic perception to forecast volcanic eruptions.

To guide the reader through the logical structure of this work, the following outline summarizes the organization of the thesis.

- **Chapter 1 - Introduction.** Presents the scientific context, motivations, objectives and overall vision of the study.
- **Chapter 2 - State of the Art.** Reviews the most relevant contributions on volcanic monitoring, multiparametric analysis and machine learning approaches for forecasting.
- **Chapter 3 - Dataset and Preprocessing.** Describes the monitoring network, the datasets employed and the procedures used for cleaning, harmonizing and synchronizing the time series.
- **Chapter 4 - Methodology and Predictive Framework.** Details the methodological choices, the implemented models (RF, LSTM and Multi-LSTM) and the experimental pipeline.
- **Chapter 5 - Results.** Presents the performance of the models, the comparative evaluation and the interpretation of precursory patterns.
- **Chapter 6 - Conclusions and Future Work.** Summarizes the findings, discusses their implications for operational monitoring and outlines potential directions for further development.

Chapter 2

State-of-the-Art for Etna Monitoring

The continuous expansion of observational networks in volcanology provides high-resolution datasets that encompass several multiparametric data (e.g. seismicity, ground deformation, gas emissions, thermal anomalies). Traditional monitoring relies on physical modeling and expert-driven interpretation of individual or cross-integrated signals. These approaches remain essential, but they struggle to fully capture the complex, nonlinear and multi-scale dynamics that characterize active systems such as Etna. In recent years, the emergence of machine learning (ML) and deep learning (DL) has transformed the ability to extract meaningful patterns from heterogeneous datasets. These methods enable the automatic identification of precursory signatures and complex correlations among variables that may not be evident through conventional analysis. Their application in volcanology allows for more timely and data-driven interpretations of volcanic unrest. Etna, with its monitoring network managed by the INGV-OE, provides an ideal natural laboratory for testing and validating these approaches.

This chapter reviews key studies applying ML and DL methodologies to Mount Etna's monitoring data, with a particular focus on predictive and operational frameworks. By comparing methodological choices, such as feature extraction strategies, data preprocessing, model architecture and interpretability, it becomes possible to identify which approaches deliver robust predictive performance and which challenges remain unresolved. The goal is to provide a structured, critical overview of the current state of the art, highlighting contributions, strengths and limitations of existing models. This synthesis serves as a foundation for the methodological developments presented in subsequent chapters.

2.1 Case Studies in ML for Mount Etna

A) “Using unsupervised machine learning to identify changes in eruptive behavior at Mount Etna, Italy” – Watson, 2020 [2]

In this study, the authors analyzed continuous infrasound recordings to characterize the eruptive dynamics of Etna and to evaluate the potential of unsupervised machine learning techniques for real-time eruption monitoring. The analysis focused on the fissure eruption of 24 December 2018, which followed a phase of intense Strombolian activity at the summit craters. A total of 72 hours of continuous data (shown in figure 2.1) were collected from the EMFO station, located 8.1 km from the Bocca Nuova crater, sampled at 100 Hz and divided into five-minute analysis windows with 90% overlap. From each window, seven statistical and spectral features were extracted, standard deviation, skewness, kurtosis, peak frequency, quality factor and median frequency, all normalized to zero mean and unit variance to ensure comparability. The authors then applied a K-means clustering algorithm with $k = 7$ to automatically identify recurrent patterns in the acoustic signal and to distinguish between different eruptive regimes.

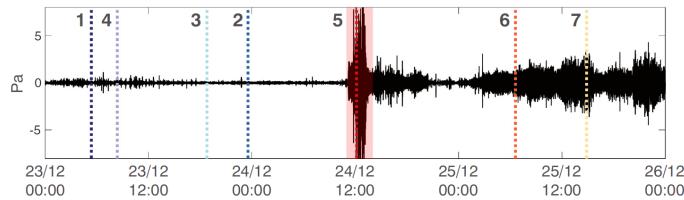


FIGURE 2.1: Representative infrasound waveform and spectrogram illustrating the temporal evolution of eruptive activity (source [2]).

The clustering results revealed a clear correspondence between cluster transitions and the actual evolution of eruptive phases. Specifically, clusters K1–K4 were associated with summit Strombolian explosions, K5 with high-amplitude fissure lava fountaining and K6–K7 with post-fissure degassing and quiescent activity. The identified patterns captured the changing spectral and energetic characteristics of the eruption, reflecting shifts in source dynamics and conduit processes.

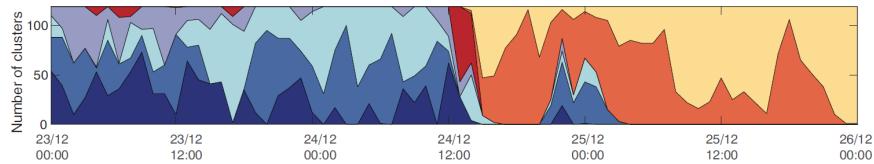


FIGURE 2.2: Temporal evolution of cluster assignments showing transitions between eruptive regimes (source [2]).

This study demonstrated that unsupervised learning can effectively separate different eruptive styles from continuous infrasound data without prior labeling. The clustering approach provided valuable insights into the temporal structure of eruptive activity and highlighted the feasibility of implementing automated, real-time detection systems for eruption monitoring at Mount Etna and similar volcanoes worldwide.

B) “Classification of Mount Etna (Italy) volcanic activity by machine learning approaches” – Hajian et al., 2019 [3]

The authors explored the application of classical machine learning algorithms to classify different states of volcanic activity at Etna using multi-parametric data collected between 2011 and 2015 by INGV–OE. The objective was to assess whether characteristic combinations of geophysical and geochemical parameters could be used to automatically distinguish eruptive regimes, quiescent, Strombolian and paroxysmal, thus providing a data-driven foundation for early-warning applications.

The dataset included five continuous observables: seismic RMS amplitude, centroid depth, infrasound amplitude, radar backscatter power (as a measure of ash plume intensity) and ground tilt recorded by tiltmeters. All features were standardized using z-score normalization and used to train two supervised classifiers: a Decision Tree (DT) and a K-Nearest Neighbours (KNN) model. Model performance was evaluated using leave-one-out cross-validation (LOOCV), ensuring maximum utilization of the available samples and quantified through precision, recall, F1-score and confusion matrices.

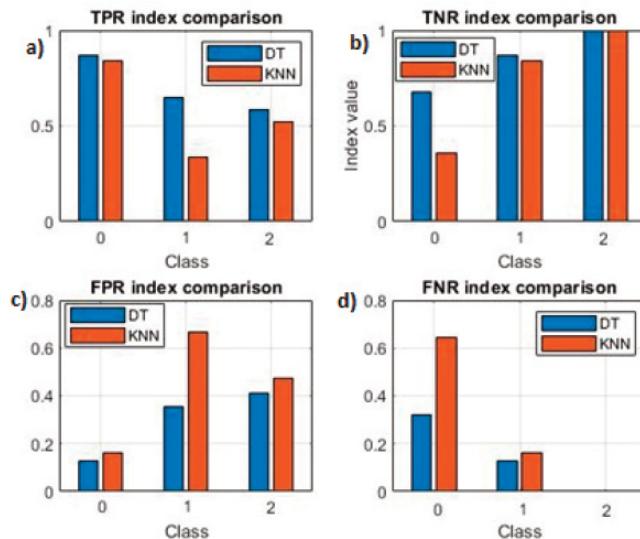


FIGURE 2.3: Comparison of Decision Tree (DT) and K-Nearest Neighbours (KNN) classifier results (source [3]).

Both models succeeded in discriminating between the three activity classes, though the Decision Tree achieved higher accuracy and a better trade-off between sensitivity and specificity. The KNN classifier showed reduced performance during transitional phases between Strombolian and paroxysmal activity, reflecting the overlap of physical parameters in those periods. Nonetheless, the predicted class probabilities exhibited strong temporal correspondence with the observed eruptive events, confirming that the selected features effectively captured the volcano’s evolving state.

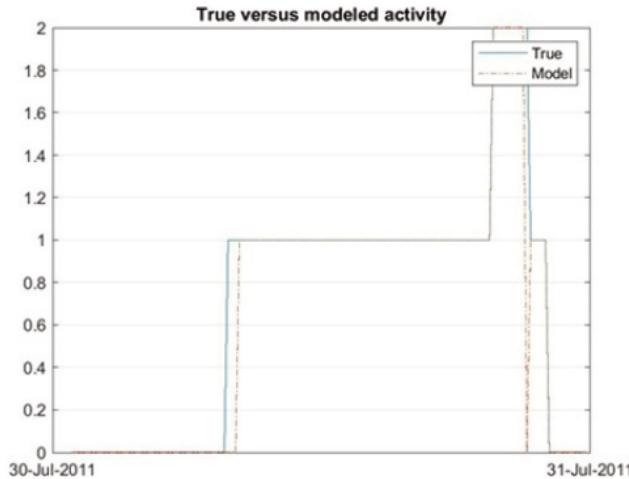


FIGURE 2.4: Classification estimation for the activity of Mt. Etna between 30 and 31 July 2011 (source [3]).

The authors concluded that relatively simple ML models can reproduce the decisions of expert analysts and support automated classification within volcano monitoring workflows. The Decision Tree model, in particular, reached accuracies exceeding 90% for paroxysmal phases, demonstrating its potential as a lightweight, interpretable tool for operational use. Although limited in temporal scope, this work represents one of the first successful integrations of multi-sensor datasets for supervised classification of Etna’s activity and laid the groundwork for subsequent deep learning approaches.

C) “Exploring convolutional neural networks for the thermal image classification of volcanic activity” – Nunnari and Calvari, 2024 [4]

In their work, the authors investigated the potential of convolutional neural networks (CNNs) for classifying volcanic activity states from thermal images recorded by INGV–OE’s ground-based monitoring cameras. The study aimed to evaluate the capability of deep learning architectures to identify eruptive phenomena from visual data and to assess how transfer learning can improve performance when working with limited datasets. The authors assembled a dataset of 476 labeled thermal images corresponding to six classes of volcanic activity: Strombolian, Lava Fountain, Lava Flow, Degassing, Cloudy and Undefined. Each image, with a resolution of 640×480 pixels, was preprocessed through histogram equalization and normalization to increase contrast and minimize sensor noise. Eight CNN architectures were compared, including VGG-16, AlexNet, ResNet-18, DenseNet201 and GoogleNet, under two different training regimes: from-scratch learning and transfer learning based on ImageNet weights.

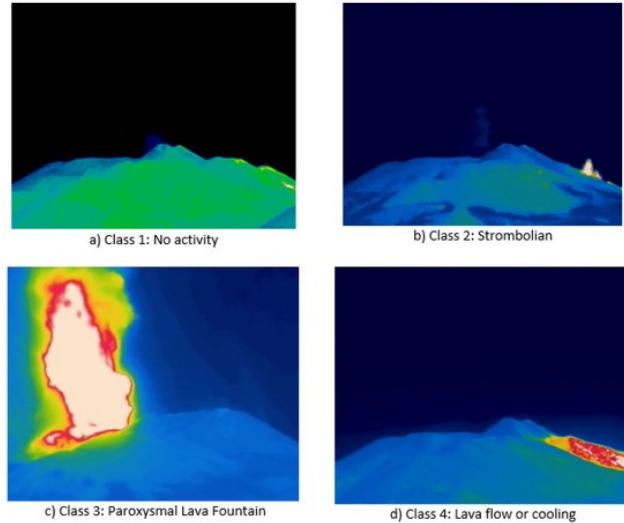


FIGURE 2.5: Examples of thermal images corresponding to the first four activity classes (source [4]).

The training process used 80% of the dataset for model fitting and 20% for validation, with accuracy, precision and recall employed as the main evaluation metrics. Results showed that transfer learning consistently outperformed from-scratch training across all architectures, confirming the advantage of reusing pretrained convolutional filters learned from large-scale image datasets. Among the models tested, VGG-16 and AlexNet achieved the highest accuracy, exceeding 90%, demonstrating that even general-purpose vision models can effectively capture textural and spatial features typical of volcanic imagery.

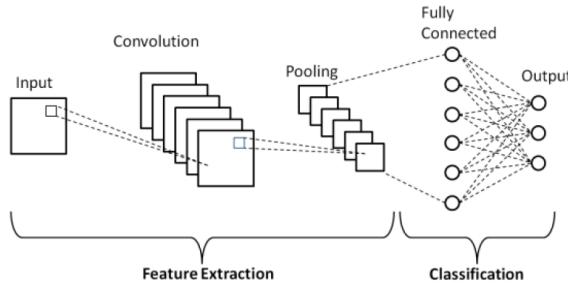


FIGURE 2.6: Architecture of a typical convolutional neural network used for volcanic image classification (source [4]).

The study highlighted the strong adaptability of CNNs to geoscientific image classification tasks and emphasized the value of transfer learning under data-scarce conditions. The authors concluded that pretrained models can serve as an effective foundation for automated recognition of eruptive patterns and that integrating CNN-based image classification with other data modalities, such as seismic or gas signals, could enable multi-source frameworks for real-time volcanic monitoring and hazard assessment.

D) “Volcano activity classification from synergy of EO data and machine learning: An application to Mount Etna volcano (Italy)”
– Petrucci et al., 2025 [5]

The most recent contribution reviewed in this chapter is the work by Petrucci, Romoli, Pignatelli, Trasatti, Zuccarello, Greco, Dozzo, Bilotta, Spina, and Ganci (2025), which presents an integrated approach for classifying volcanic activity at Etna through the fusion of multi-source satellite-based Earth Observation (EO) data and supervised machine learning. This research aimed to determine the most informative predictors for volcanic state classification and to evaluate the potential of ensemble and local algorithms for operational monitoring.

The dataset combined several EO-derived parameters collected between 2005 and 2008, including ground deformation from ascending and descending satellite orbits (35-day intervals), gravity anomalies (monthly), daily SO₂ concentrations, land surface temperature (12-hour intervals) and SEVIRI radiance channels (15-minute intervals). To synchronize the heterogeneous temporal resolutions, all features were interpolated to a 15-minute common grid using linear and nearest-neighbor methods. Volcanic state labels were derived from field observations, ensuring consistency between remote-sensing and ground-truth data.

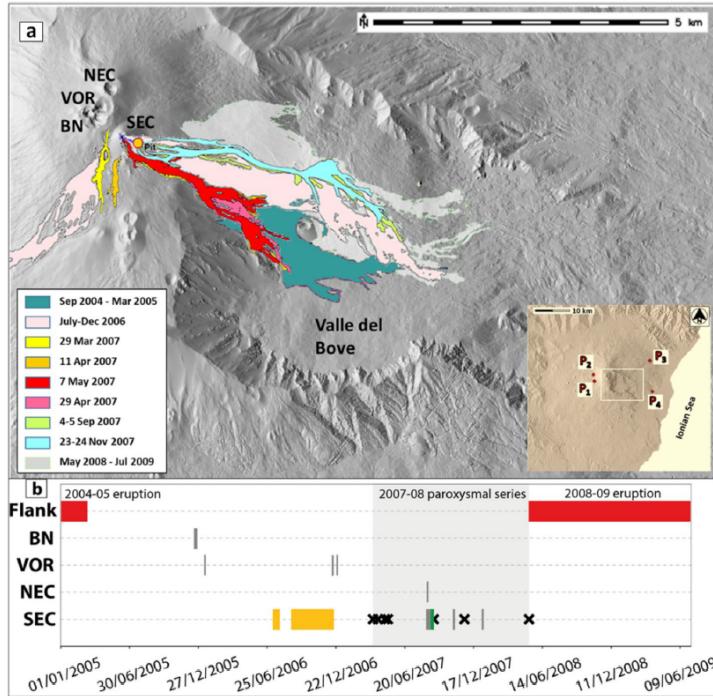


FIGURE 2.7: Map of Etna showing lava flows emitted between 2005 and 2009 together with recorded volcanic activity in the same period (source [5]).

Several supervised classifiers were compared, including Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine, Naive Bayes and Linear Discriminant Analysis (LDA). Model evaluation was performed through fivefold sequential cross-validation, preserving the chronological order of observations to avoid temporal data leakage. Accuracy,

confusion matrices and feature importance metrics (based on Random Forest's Out-Of-Bag error) were used to assess performance.

		Cooling	Eruption	Preparatory	Unrest	Quiet
True Class	Cooling	96.91%	2.7%			0.39%
	Eruption		94.72%			5.28%
	Preparatory			100%		
	Unrest		2.29%		97.71%	
	Quiet		0.29%			99.71%
		Cooling	Eruption	Preparatory	Unrest	Quiet

FIGURE 2.8: Random Forest confusion matrix averaged over five-fold cross-validation (source [5]).

The Random Forest model achieved the highest accuracy, reaching perfect classification (1.0) in three of the five-folds, while other algorithms performed between 0.618 and 0.996. Feature importance analysis revealed that gravity anomalies and ground displacement were the most influential variables, followed by SO₂ concentration and land surface temperature. Even when low-frequency features were removed in ablation tests, overall accuracy remained around 73%, demonstrating the robustness of the models to missing data.

This study represents one of the first large-scale efforts to apply multi-sensor EO data fusion for operational volcanic monitoring. The results confirm that ensemble and neighborhood-based methods, such as Random Forest and KNN, can handle the nonlinear relationships between heterogeneous data sources without assuming specific statistical distributions. The approach provides a strong foundation for future monitoring frameworks capable of integrating ground-based and satellite observations for global volcanic hazard assessment.

2.2 Concluding Remarks

The case studies reviewed in this chapter illustrate the rapidly expanding role of machine learning and deep learning in extracting meaningful information from Etna's heterogeneous monitoring datasets. These approaches have demonstrated the ability to classify eruptive states, detect transitions and support early-warning frameworks by revealing patterns that are often difficult to recognize through traditional analysis alone in a very short time. Despite these advances, most contributions remain limited in scope. Many rely on a single data modality, short observational windows or retrospective classification tasks, and only a few explore multi-sensor fusion in a systematic way. Moreover, the reviewed studies focus predominantly on eruptive activity, plume dynamics or thermal emissions, leaving almost entirely unaddressed the problem of short-term forecasting of local seismicity, which is a key indicator of changes in the volcano's internal state.

This gap directly motivates the approach adopted in this thesis. Rather than concentrating on single observable or isolated eruptive episodes, the work presented here is based on the idea that anticipating variations in local seismicity requires a unified, harmonized representation of the volcano's dynamics. For this reason, Chapter 3 introduces the multi-parametric dataset used throughout the thesis and describes the preprocessing procedures required to align seismic, deformation and geochemical measurements within a coherent temporal framework. This consolidated dataset forms the foundation upon which the predictive models of the following chapters are constructed, marking the transition from state-of-the-art analysis to the methodological core of the thesis.

Chapter 3

Dataset and Preprocessing

The dataset analyzed in this study originates from the multi-parametric monitoring network operated by the INGV - Etnean Observatory. This network continuously records a set of multi-parametric data, providing high-resolution time series that capture its dynamic behavior.

The selected data cover the period between **2017 and February 2019**, a timeframe characterized by significant eruptive and seismic activity. During this interval, Etna exhibited several notable events: in February–March 2017, intermittent Strombolian explosions occurred at the Southeast Crater, culminating on March 16 with a phreatomagmatic event that caused minor injuries; in July 2018, Strombolian activity resumed at the summit craters, marking renewed volcanic unrest; on December 24, 2018, a two-kilometer-long eruptive fissure opened on the southeastern flank, feeding lava flows into the Valle del Bove and accompanied by more than 300 recorded earthquakes. Two days later, on December 26, a magnitude 4.8 earthquake struck near Viagrande, linked to the magmatic intrusion driving the eruption.

The dataset object of this study integrates both **geophysical** and **geochemical** parameters, each providing a complementary perspective on the volcanic system. Geophysical signals, such as volcanic tremor, ground deformation and seismicity, describe subsurface processes related to magma movement, pressure variations and structural response of the edifice. Geochemical data, including gas fluxes, gas ratios and thermal emissions, reveal information on degassing dynamics and the supply of magmatic volatiles from depth. Together, these datasets allow for a multi-dimensional investigation of Etna’s activity, combining the physical and chemical manifestations of its internal dynamics within a consistent temporal framework.

3.1 Geophysical Data

The geophysical dataset encompasses measurements related to ground motion, subsurface dynamics and surface heat emission at Etna. The last one, although not strictly a geophysical parameter, is grouped here for clarity of exposition due to its strong physical connection with shallow thermal processes. These parameters are essential to capture the volcano's behavior from both seismic and deformation perspectives, providing complementary insights into magmatic and fluid movements beneath the surface.

The datasets included in this section are:

- **Volcanic Tremor (ESPC):** continuous seismic signal reflecting magma and fluid movements.
- **Clinometric Data:** tilt measurements and auxiliary sensor readings for ground deformation monitoring.
- **Seismicity:** catalog of local earthquakes with location, depth and magnitude.
- **Surface Heat Flux:** radiant heat output in the summit area.

All geophysical datasets underwent a common preprocessing workflow:

- Parsing and standardizing dates and times.
- Renaming columns for consistency.
- Converting decimal separators (comma to dot) and ensuring numeric types.
- Setting a datetime index and sorting chronologically.
- Removing duplicate records when present.

Additional dataset-specific operations were applied where needed, for instance:

- ESPC values were filtered to remove anomalous spikes corresponding to sensor glitches.
- Clinometric data were clipped to the 2017–2019 monitoring period.
- Seismicity data were processed to unify magnitude and depth formats.

3.1.1 Volcanic Tremor

Volcanic tremor (derived from station Serra Pizzuta Calvarina - ESPC) is a continuous seismic signal generated by magma and fluid movements beneath the volcano. It is recorded by broadband and short-period seismic stations strategically distributed around Etna's summit and flanks. The tremor amplitude, expressed in meters per second (m/s), represents the ground velocity associated with these subsurface processes and provides valuable insight into the dynamic behavior of the magmatic system. Figure 3.1 summarizes the main eruptive and seismic events that occurred between 2017 and 2019, outlining the temporal context in which the tremor data were analyzed. A portion of the dataset structure is shown in Table 3.1, which reports the first few daily averaged values recorded at ESPC after preprocessing.

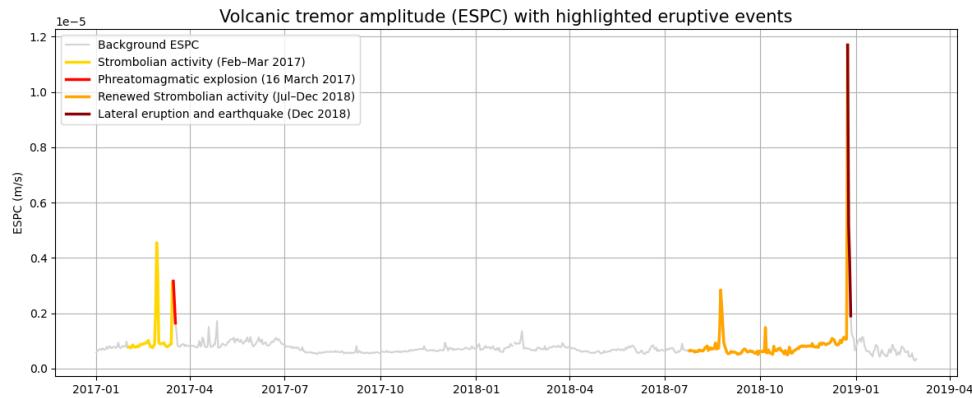


FIGURE 3.1: Main eruptive and seismic events of Etna between 2017 and 2019.

Date	ESPC (m/s)
2017-01-01	5.812e-07
2017-01-02	6.321e-07
2017-01-03	6.616e-07
2017-01-04	6.814e-07
2017-01-05	7.036e-07

TABLE 3.1: First five rows of the volcanic tremor dataset after preprocessing.

The ESPC signal shows modest daily variability, with a mean amplitude of 7.79×10^{-7} m/s and a standard deviation of 5.23×10^{-7} m/s, ranging from 2.86×10^{-7} to 1.17×10^{-5} m/s. These fluctuations reflect the typical background dynamics of Etna's shallow magmatic system, whose energy variations often precede or accompany phases of eruptive unrest.

3.1.2 Clinometric Data

Clinometric data consist of high-resolution tilt measurements that track ground deformation at Etna. Tilt values, expressed in microradians (μrad), quantify small angular displacements of the ground, which may indicate magma movement, pressurization, or other subsurface processes. Auxiliary sensors in the same stations record additional parameters such as battery voltage (V), data logger temperature ($^{\circ}\text{C}$), tiltmeter temperature ($^{\circ}\text{C}$) and barometric pressure (hPa), ensuring proper calibration and data quality. Continuous monitoring of these variables provides crucial insight into the volcano's deformation behavior and supports early recognition of unrest phases.

Date	tilt_x_Avg	tilt_y_Avg	temp_tilt ($^{\circ}\text{C}$)	nord_tilt	barometer (hPa)
2017-01-01 00:00	85.486	215.322	6.63	230.0	853
2017-01-01 00:15	85.479	215.303	6.65	230.1	850
2017-01-01 00:45	85.492	215.309	6.63	230.0	849
2017-01-01 01:00	85.498	215.303	6.63	230.1	852
2017-01-01 01:15	85.498	215.316	6.65	230.0	850

TABLE 3.2: First five rows of the clinometric dataset (main deformation-related variables).

Tilt measurements exhibit mean values around 94.73 μrad (X) and 225.39 μrad (Y), with standard deviations of 6.43 and 6.61 μrad , respectively. Battery voltage, CR10 temperature and barometric pressure remain stable, confirming the reliability of the monitoring stations.

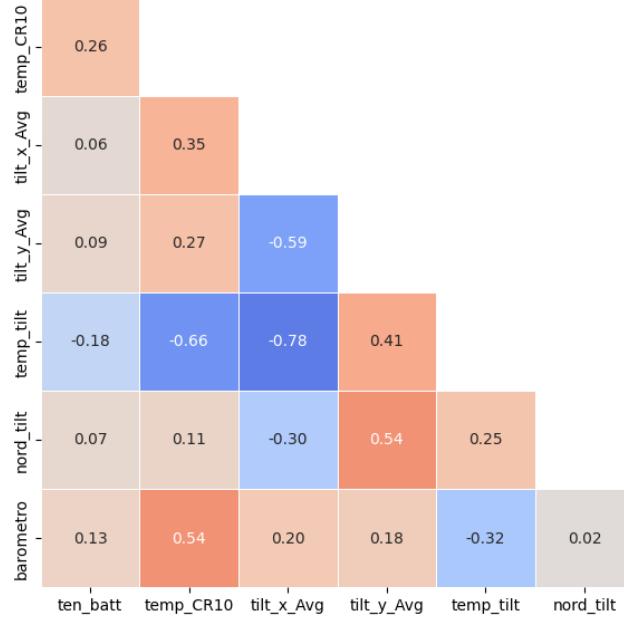


FIGURE 3.2: Spearman correlation matrix among clinometric variables.

As shown in Figure 3.2, several strong and moderate correlations emerge among the clinometric variables. A pronounced negative correlation ($\rho = -0.59$) exists between `tilt_x_Avg` and `tilt_y_Avg`, consistent with their orthogonal orientation and opposite deformation directions. Temperature-related variables (`temp_tilt` and `temp_CR10`) also correlate significantly with tilt components ($\rho = -0.78$ and $\rho = 0.41$), indicating partial thermal influence on sensor readings. Moderate correlations between barometric pressure and auxiliary channels further suggest sensitivity to atmospheric variations. Overall, these patterns validate the internal coherence of the dataset and highlight the need to consider environmental effects in deformation analysis. To better capture short-term deformation dynamics, an aggregated metric, called the *movement index*, was computed as the sum of the normalized first derivatives of the two tilt components (`tilt_x_Avg` and `tilt_y_Avg`). This index expresses the rate of ground tilt change, emphasizing transient accelerations and sudden variations that can precede eruptive or intrusive events. It provides a single interpretable time series summarizing the intensity and speed of deformation, allowing direct comparison with other parameters such as volcanic tremor amplitude.

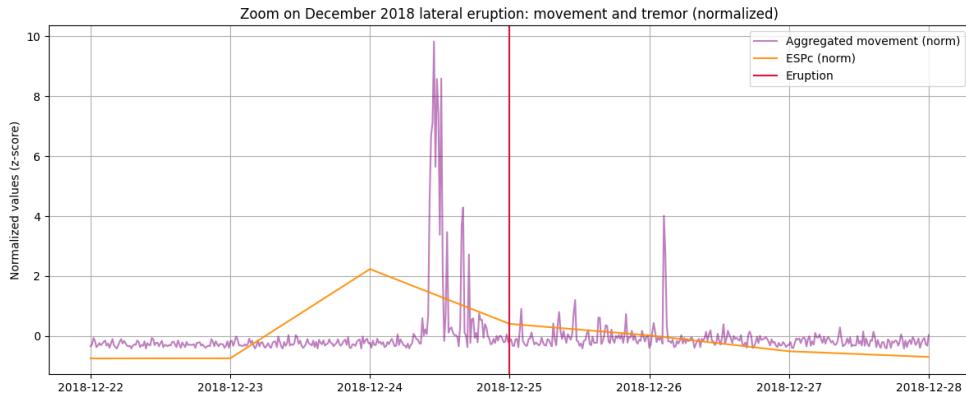


FIGURE 3.3: Aggregated movement index from tilt derivatives compared with volcanic tremor (ESPC) around 25 December 2018.

Peaks in the movement index (Figure 3.3) coincide with increases in volcanic tremor, revealing a strong coupling between surface deformation and subsurface seismic activity. During the December 2018 eruptive episode, the index highlights rapid ground shifts not clearly visible in raw tilt data. This approach enhances the detection of short-lived deformation transients and supports a more integrated interpretation of Etna’s near-surface dynamics.

3.1.3 Seismicity

The seismicity dataset includes all local earthquakes recorded by the INGV seismic network around Etna during the monitoring period. Each event is characterized by its occurrence time, geographic coordinates (latitude and longitude in degrees), focal depth (km) and local magnitude (ML). Seismic stations equipped with broadband and short-period seismometers allow precise detection of both shallow and deep events, offering key information on magma migration, tectonic stress evolution and possible eruptive precursors.

Date	Lat (°N)	Long (°E)	Depth (km)	ML
2017-06-01 02:33:51	37.7226	14.9926	7.68	1.1
2017-06-01 10:07:35	37.7921	14.9304	23.21	1.8
2017-06-01 10:15:07	37.7625	14.9517	26.49	1.9
2017-06-03 05:30:59	37.7159	15.0948	5.23	1.3
2017-06-03 06:27:08	37.7161	15.0934	4.72	1.2

TABLE 3.3: First five rows of the seismicity dataset after preprocessing.

The earthquake catalog has a mean latitude of 37.733°N and a mean longitude of 15.000°E, with an average focal depth of 6.04 km and a mean local magnitude of 1.34. Maximum observed magnitude is 3.3 and depths range from near-surface events down to 29 km. This information provides a detailed view of both the spatial and energetic distribution of volcanic earthquakes, crucial for tracking magma migration and stress variations within the edifice. During the study period, seismicity was predominantly concentrated beneath the summit craters and along the southeastern rift zone, consistent with Mount Etna’s known structural weaknesses. Shallow events

(depths < 5 km) often coincided with eruptive activity or intrusions, while deeper earthquakes were generally linked to regional tectonic adjustments and magma recharge. This vertical stratification highlights the interaction between magmatic dynamics and the volcano’s tectonic framework. From an analytical standpoint, the seismic catalog represents a discrete but highly informative signal that complements the continuous geophysical parameters. By aggregating event counts or magnitudes over fixed temporal windows, it becomes possible to correlate seismicity with variations in tremor amplitude, deformation and gas emission—enabling a multi-parametric characterization of volcanic unrest.

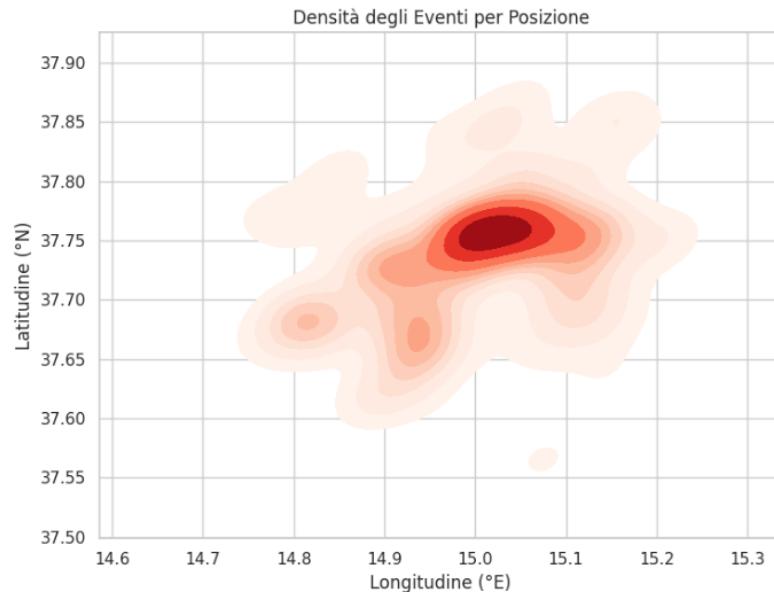


FIGURE 3.4: Spatial density of earthquake epicenters recorded around Etna between 2017 and 2019. Darker areas indicate higher concentrations of seismic events.

As shown in Figure 3.4, the highest concentration of earthquakes occurs beneath the summit craters and along the southeastern flank, an area structurally associated with the Valle del Bove and the main rift systems. This distribution reflects the preferential pathways of magma ascent and zones of stress accumulation within the volcanic edifice. Smaller, dispersed clusters toward the south and southwest likely correspond to peripheral fault activity and shallow flank adjustments. Overall, the map confirms that seismicity is a spatially coherent indicator of both magmatic and tectonic processes, providing essential constraints for short-term forecasting and for integration with other monitoring parameters.

3.1.4 Surface Heat Flux

The surface heat flux dataset records the radiant thermal energy emitted from Etna’s summit area. Measurements were acquired through infrared thermometers and satellite-based sensors, providing estimates of heat emission in watts (W). Heat flux reflects both passive degassing and active lava effusion, making it a crucial indicator of volcanic activity and energetic output. Its high temporal resolution enables the detection of rapid thermal fluctuations associated with eruptive episodes.

Date	Heat Flux (W)
2018-12-24 08:19	1.41e+08
2018-12-24 11:13	5.78e+09
2018-12-24 11:19	9.05e+09
2018-12-24 11:24	6.59e+09
2018-12-24 11:26	1.25e+10

TABLE 3.4: First five rows of the surface heat flux dataset after preprocessing.

The radiant heat flux shows a mean value of 6.09×10^9 W and a standard deviation of 3.82×10^9 W, ranging from 1.28×10^8 to 1.75×10^{10} W. These variations capture both quiescent and eruptive thermal states of the volcano. At the summit, thermal anomalies typically arise from persistent passive degassing and from transient increases in surface heat linked to lava effusion or dome growth. The most intense peaks in radiant flux coincide with eruptive phases, such as the December 2018 eruption, when values increased by more than an order of magnitude within a few hours. These rapid changes reveal enhanced heat release linked to magma ascent and lava exposure on the surface. From an analytical perspective, the surface heat flux represents a quantitative measure of Etna’s radiative energy budget and complements the other geophysical parameters. When combined with tremor amplitude and deformation time series, it enables the identification of multi-parametric signatures of unrest, where thermal peaks often follow or accompany seismic and tilt anomalies. This relationship makes heat flux a valuable proxy for real-time activity monitoring and for validating predictive models.

3.2 Gas Emission Fluxes and Ratios

Volcanic gases are among the most direct manifestations of magmatic activity and play a crucial role in assessing the internal state of a volcano. Their composition and emission rates provide key constraints on magma ascent, storage and degassing processes. Continuous gas monitoring at Mount Etna, carried out by the INGV multi-parametric network, focuses mainly on the emission of sulfur dioxide (SO_2), carbon dioxide (CO_2) and hydrogen chloride (HCl), the most diagnostic volatile species for tracking variations in volcanic activity. In addition to absolute fluxes, molar ratios such as CO_2/SO_2 and ${}^3\text{He}/{}^4\text{He}$ provide deeper insight into the dynamics of the magmatic system. Variations in the CO_2/SO_2 ratio are typically linked to changes in degassing depth and efficiency, as CO_2 exsolves at higher pressures than SO_2 . The ${}^3\text{He}/{}^4\text{He}$ ratio, in turn, traces the origin of volatile sources, distinguishing

mantle-derived contributions from crustal or hydrothermal contamination. Together, these parameters support a more comprehensive interpretation of Etna's degassing regime and its temporal evolution. Before analysis, raw flux data for HCl, SO₂ and CO₂ were homogenized in format and sampling frequency. Dates were standardized, decimal separators were unified and all time series were synchronized. Since CO₂ data were already normalized, only SO₂ and HCl fluxes were rescaled to a 0–1 range using Min–Max normalization to allow direct comparison of their temporal trends.

3.2.1 Gas Fluxes

HCl is a halogen component typically associated with magmatic–hydrothermal interactions or direct release from chloride-bearing melts. Due to its high solubility and reactivity, HCl fluxes are generally more variable and lower in magnitude. They are measured through periodic FTIR campaigns and in-situ fumarole sampling, with lower temporal resolution than SO₂ or CO₂, which are recorded daily.

SO₂ is a primary magmatic volatile released when magma approaches the surface and pressure decreases. It is relatively insoluble at shallow depths, making it a robust tracer of degassing in open-conduit systems. Elevated SO₂ fluxes are commonly associated with eruptive or pre-eruptive episodes, whereas steady, moderate emissions indicate persistent passive degassing from a stable shallow reservoir.

CO₂, in contrast, exsolves from magma at greater depths due to its lower solubility. It thus serves as a deeper geochemical indicator of magmatic input or recharge in the plumbing system. Fluxes are expressed in tons per day (t/d). Anomalous CO₂ increases often precede SO₂ peaks, providing early evidence of deep magmatic activity before gas reaches the surface.

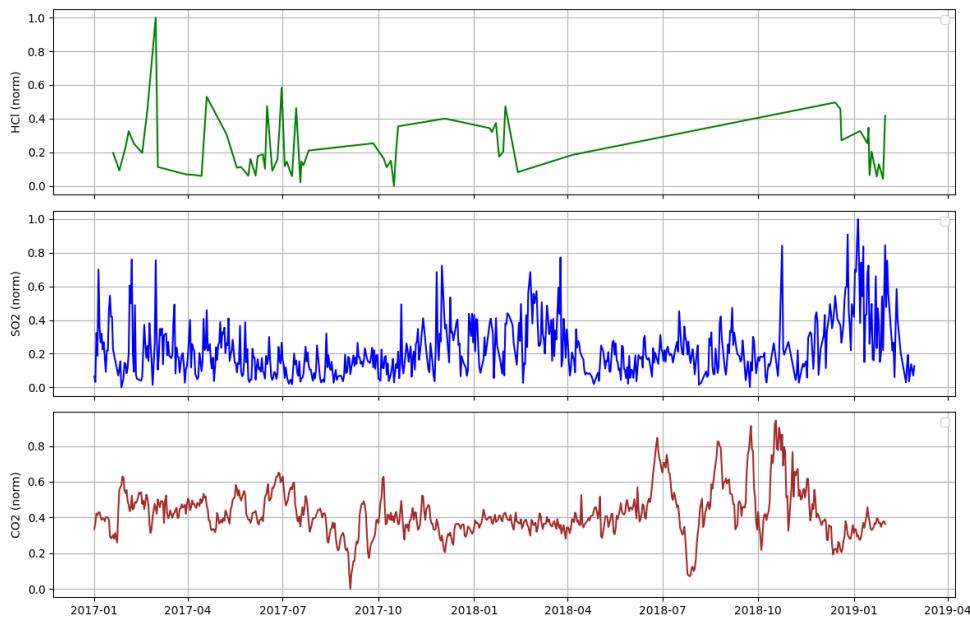


FIGURE 3.5: HCl, SO₂ and CO₂ emission fluxes recorded during 2018.

As shown in Figure 3.5, the normalized time series of HCl, SO₂ and CO₂ fluxes reveal marked fluctuations in Etna's degassing behavior between

2017 and 2019. All three gases show variable patterns, with a notable increase in amplitude and frequency of peaks during the second half of 2018—consistent with intensifying volcanic activity. HCl remains irregular due to its shallowness decoupling from melt, but eventually displays a gradual rise parallel to SO₂. CO₂ exhibits smoother variations and an earlier increase preceding the strongest SO₂ peaks, reflecting its deeper magmatic origin. Overall, the co-variation of these gas species in late 2018 indicates renewed magma input and pressurization within the plumbing system, culminating in the eruptive activity of December 2018. Their combined interpretation provides a multi-depth perspective on Etna's degassing regime and the physical processes driving volcanic unrest.

3.2.2 Gas Ratios

Gas ratios provide an essential link between deep magmatic processes and surface degassing dynamics. All datasets were reformatted into consistent structures, standardized to a common `datetime` index and resampled to daily resolution for direct comparison across geochemical parameters.

CO₂/SO₂ ratio. This molar ratio reflects the balance between deep and shallow gas contributions. Higher CO₂/SO₂ values indicate a stronger input from deep-seated magma, while lower values correspond to sulfur-rich, shallow degassing. Monitoring this ratio helps assess the depth and efficiency of degassing, as well as the transition between closed- and open-system conduit conditions.

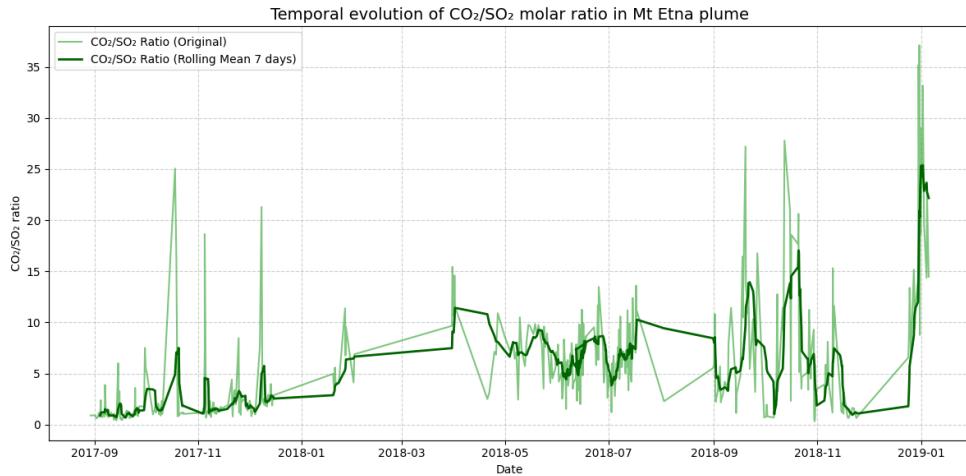


FIGURE 3.6: CO₂/SO₂ molar ratio at Etna between 2017 and 2019.

As shown in Figure 3.6, the CO₂/SO₂ ratio exhibits pronounced short-term variability, generally ranging between 0.5 and 1.0 during 2017–2018, with several sharp peaks exceeding 10 in late 2018. These peaks correspond to enhanced CO₂ input from deeper sources, while transient decreases indicate sulfur-dominated shallow degassing. The strong rise in late 2018 is consistent with a phase of intensified magmatic recharge preceding the December eruption, in agreement with other geophysical and geochemical parameters.

${}^3\text{He}/{}^4\text{He}$ ratio. The helium isotopic ratio (${}^3\text{He}/{}^4\text{He}$) acts as an independent tracer of volatile origin and evolution. ${}^3\text{He}$ is a primordial isotope of mantle origin, while ${}^4\text{He}$ is radiogenic and crustal. Higher ratios thus reflect a stronger mantle contribution and renewed magma input, whereas lower values indicate degassing-dominated or hydrothermal mixing conditions.

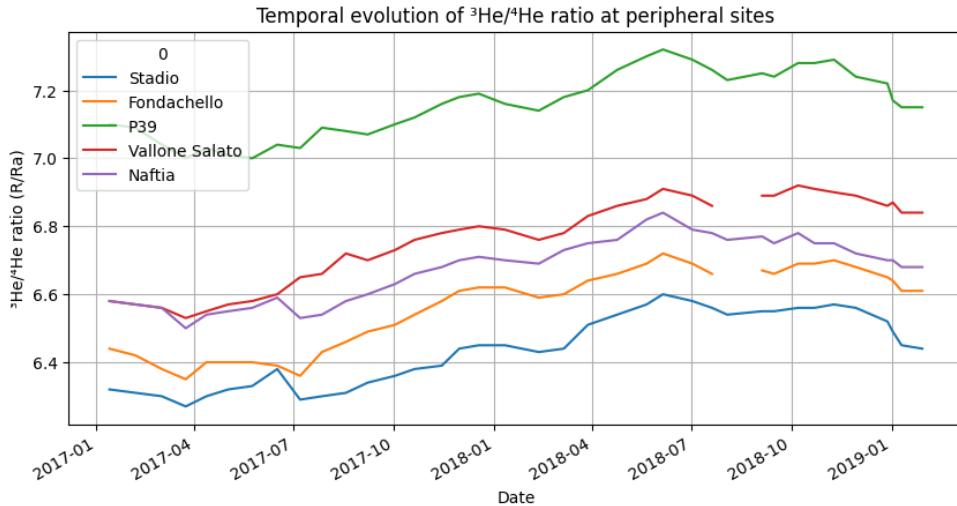


FIGURE 3.7: ${}^3\text{He}/{}^4\text{He}$ ratio measured at peripheral emission sites (Stadio, Fondachello, P39, Vallone Salato and Naftia) between 2017 and 2019.

As illustrated in Figure 3.7, the ${}^3\text{He}/{}^4\text{He}$ ratio gradually increases from early 2017 to mid-2018, peaking at the P39 site (up to 7.3 R/R_A). This upward trend indicates progressive deep magmatic recharge preceding the late-2018 eruption. A subsequent stabilization and slight decrease through 2019 reflect a shift toward degassing-dominated conditions as the system released accumulated volatiles. Spatially, P39 consistently exhibits higher ratios—suggesting a more direct magmatic connection—while Stadio and Fondachello show lower values, consistent with partial crustal or hydrothermal influence. Taken together, the concurrent rise of both CO₂/SO₂ and ${}^3\text{He}/{}^4\text{He}$ ratios during mid- to late 2018 highlights a phase of renewed magma recharge and pressurization within Etna’s plumbing system. These parallel trends, followed by increased surface degassing and eruptive activity, underscore the diagnostic value of gas ratios for identifying pre-eruptive magmatic processes.

3.2.3 Clustering Analysis of Seismic Depth and Magnitude

To further explore the internal structure of the earthquake catalog and identify potential depth-related patterns, an unsupervised clustering analysis was applied to the seismic dataset, allowing the data to reveal its own organization without predefined labeling. A **K-Means** algorithm was performed on standardized values of focal depth and local magnitude (ML). Both variables were normalized using *z-score* scaling and the optimal number of clusters was determined using the *elbow method* and *silhouette score*, which consistently identified three main clusters. The clustering results were then projected onto the first two principal components (PCA) to facilitate visualization while preserving most of the variance.

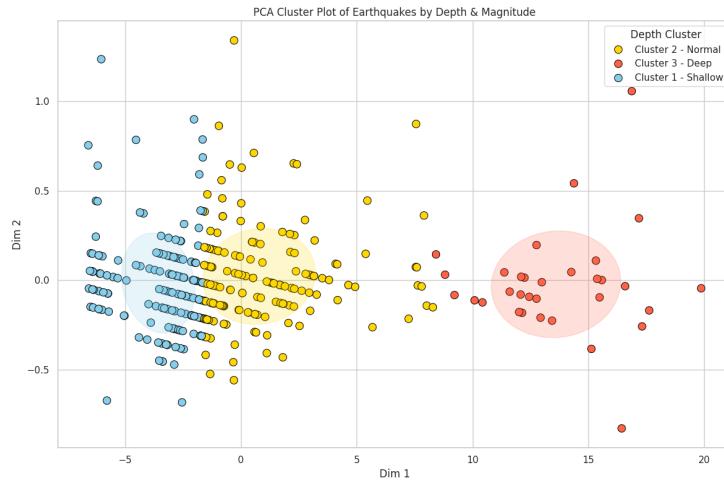


FIGURE 3.8: Projection of K-Means clusters in the first two principal component dimensions (PCA), based on earthquake depth and magnitude.

As shown in Figure 3.8, the resulting distribution reveals three coherent groups that can be interpreted as distinct depth regimes. Cluster 1 (blue) corresponds to **shallow events** (<5 km), concentrated beneath the summit craters and upper flanks; Cluster 2 (yellow) represents **intermediate-depth earthquakes** (5–15 km), associated with the main storage and intrusion zones of the plumbing system; and Cluster 3 (red) groups **deep events** (>15 km), likely related to regional tectonic adjustments or deep magma recharge. These patterns confirm that seismicity at Etna is vertically stratified and that depth and magnitude jointly define natural behavioral categories within the volcano’s seismogenic system.

3.3 Data Insights

The datasets presented in this chapter offer a coherent, multi-parametric view of Etna’s activity between 2017 and 2019. After systematic preprocessing, all geophysical and geochemical signals were standardized and aligned within a single temporal framework, enabling direct comparison across domains. This unified structure links deep and shallow processes, connecting magmatic recharge, gas emissions, deformation, and seismicity within the same timeline. A key benefit of this integration is the ability to observe how variations in different monitoring parameters accompany short-term changes in local earthquake occurrence. Daily tremor amplitude, tilt-derived deformation, gas flux fluctuations, and thermal anomalies can now be examined together, revealing coordinated patterns that often precede shallow seismic activation. The December 2018 intrusion, for instance, shows clear multi-sensor shifts that highlight the usefulness of a harmonized timescale for identifying precursory behavior. This consolidated framework provides the foundation for the predictive modeling introduced in Chapter 4, where the time series are used as input to a recurrent neural network (LSTM) designed to learn temporal dependencies and detect patterns linked to impending seismicity.

Chapter 4

Methodology and Predictive Framework

The design of a predictive framework represents the natural continuation of the exploratory analysis presented in chapter 3. After defining and harmonizing the multi-parametric dataset, the next objective is to determine whether the temporal evolution of these signals contains sufficient information to anticipate variations in seismic activity in the Etna region. This challenge requires an approach that can combine the physical interpretability of classical statistical models with the temporal sensitivity of modern deep learning architectures. In operational volcanology, predictive modeling serves a dual purpose. From a scientific standpoint, it enables the identification of latent relationships among different observables that may precede eruptive or intrusive phenomena. From a monitoring perspective, it provides a quantitative tool to support early-warning systems, reducing the dependence on purely empirical thresholds. To this end, the predictive pipeline developed in this work integrates both geophysical and geochemical parameters, transformed into a unified temporal structure and processed through machine-learning algorithms.

Within this framework, the forecasting problem is formulated as a binary classification task, where each time segment of the monitoring data is associated with one of two possible states: *non-event* (background activity) or *event* (increased seismicity within a short forecast window). This formulation reflects the operational need to distinguish periods of stability from those potentially leading to eruptive unrest, rather than to deterministically predict individual earthquakes. Target labels were derived directly from the local earthquake catalog introduced in chapter 3. For each day, the label “event” was assigned if at least one earthquake of magnitude $ML \geq 2.0$ occurred within the subsequent 24 hours, while all remaining days were labeled as “non-event.” This rolling labeling scheme ensures consistent temporal alignment between input features and seismic outcomes, allowing the models to infer short-term dependencies across the multi-parametric

signals. The 24-hour forecast horizon was chosen based on both physical and operational considerations. Physically, several observables acquired on Mt. Etna, such as tremor amplitude, tilt rate and the CO₂/SO₂ ratio, tend to show measurable fluctuations in the hours or days preceding eruptive or intrusive activity. Operationally, a one-day prediction window provides a realistic timescale for early-warning evaluation, compatible with INGV’s data acquisition and alerting routines. A key aspect of this formulation is the pronounced imbalance between the two classes: quiescent periods largely outnumber those associated with seismic activation. To mitigate this effect, class-weighted loss functions are employed during model training, assigning higher penalties to misclassified “event” samples. This strategy enhances the model’s sensitivity to rare but critical episodes, aligning its behavior with the objectives of operational volcanic early-warning systems. All predictive features are computed exclusively from data preceding each target window, ensuring temporal causality and preventing any leakage of future information. This constraint guarantees that the framework operates under the same conditions as a real-time monitoring environment, where only past and present observations are available. This chapter details the complete methodological pipeline: the preparation of harmonized input sequences, the design of classical and deep learning models and the evaluation strategies adopted to quantify their predictive skill in anticipating short-term variations in Mount Etna’s activity.

4.1 Data Preparation and Feature Engineering

The heterogeneous time series described in chapter 3 were transformed into a unified predictive dataset through a series of preprocessing and feature-construction steps. The goal of this stage is to produce temporally consistent input sequences that preserve both the fast and slow dynamics of Etna’s monitoring parameters, while remaining compatible with the statistical and neural models adopted in this work. All signals were first resampled to a common **hourly** grid and aligned on the same temporal axis. Missing or irregular records were linearly interpolated over short gaps, while long discontinuities were masked to prevent the introduction of artificial trends. Extreme outliers, typically associated with sensor resets or data transmission errors, were filtered using percentile-based clipping. Each parameter was then standardized through *z-score* normalization, ensuring comparable amplitudes and stabilizing the optimization process during model training. To incorporate temporal information, the continuous series were segmented into partially overlapping windows of fixed duration. Each window represents the evolution of the system over a defined temporal context and serves as the basic predictive unit. In this study, a **four-day (96-hour)** window with a stride of one day was selected as a suitable compromise between temporal resolution and computational cost. This structure allows the model to learn gradual precursory changes without diluting short-lived anomalies.

For the **Random Forest** baseline, each window was summarized using descriptive statistics of all monitored variables, including the mean, standard deviation, minimum, maximum and linear trend. These aggregated indicators capture the first and second order behavior of each signal, providing a compact

yet informative representation of the system's state. The resulting feature table serves as input for the classical classifier.

For the **LSTM-based models**, instead, the full multivariate sequence within each window was retained in its temporal form. This representation preserves correlations between parameters and their evolution through time, allowing the model to recognize ordered patterns such as the progressive increase in tremor amplitude or deformation rate preceding eruptive episodes. Through this formulation, the model directly operates on the raw temporal structure, learning latent features automatically rather than relying on hand-crafted statistics. It is worth noting that the first sequential experiment, corresponding to the baseline LSTM model, was trained on a **univariate sequence** representing the volcanic tremor amplitude (ESPC). This signal serves as a synthetic proxy for the internal dynamics of the volcanic conduit, as it reflects variations in subsurface fluid motion and magma pressurization. The subsequent multi-parametric LSTM model extends this approach to a broader feature domain, combining both geophysical and geochemical observables within the same recurrent architecture. Both representations, aggregated and sequential, were generated from the same harmonized data stream, ensuring that any difference in model performance reflects genuine differences in learning capacity rather than inconsistencies in the input. This unified pipeline establishes a reproducible foundation for the comparative analysis presented in the following sections.

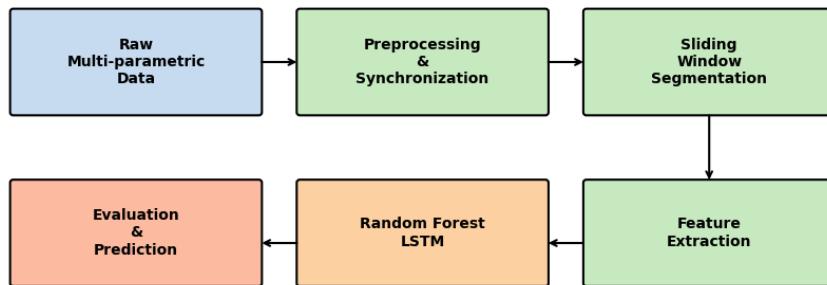


FIGURE 4.1: Overview of the predictive pipeline integrating preprocessing, feature extraction, model training and evaluation.

After feature construction, the dataset was chronologically partitioned into training, validation and test subsets to simulate real-time operational forecasting. Random sampling was avoided to preserve temporal causality. The training set encompasses 2017–2018, the validation set the early weeks of 2019 and the test set corresponds to the final segment of the dataset (February 2019). This chronological split ensures that models are always tested on genuinely unseen future data.

Subset	Time	Purpose	Characteristics
Training	Jan 2017 Dec 2018	Model fitting	Includes background activity and moderate eruptive phases.
Validation	Jan 2019 Feb 2019	Hyperparameter tuning	Used to monitor generalization and adjust early stopping.
Test	Feb 2019	Final evaluation	Independent future data, unseen during training.

TABLE 4.1: Temporal partitioning of the dataset used for model development and evaluation.

Given the unbalanced distribution of “event” and “non-event” days, the training phase employs a cost-sensitive learning strategy. Each positive instance (day with seismic activation) receives a higher weight in the loss function, prioritizing the correct classification of rare but critical cases. This weighting scheme improves recall without excessively penalizing precision, aligning the optimization process with the needs of an early-warning system where missing an event is more consequential than triggering a false alarm.

4.2 Baseline Model: Random Forest

As an initial benchmark, a **Random Forest (RF)** classifier was implemented using the feature-based representation of the multi-parametric dataset. This model provides a non-sequential reference capable of capturing nonlinear relationships among variables, serving both as a performance baseline and as an interpretable framework for feature relevance analysis. The Random Forest ensemble mitigates overfitting through bootstrapping and random feature selection, offering robustness with limited computational complexity. The predictive input matrix was constructed by summarizing each four-day (96-hour) monitoring window through a set of descriptive statistics applied to all available geophysical and geochemical signals. Specifically, the following parameters were considered:

- **Geophysical domain:** volcanic tremor amplitude, clinometric tilt variation and local seismicity rate (daily event count, maximum magnitude);
- **Geochemical domain:** SO₂ flux, CO₂ flux and the CO₂/SO₂ ratio;

For each parameter, five summary descriptors were computed within each four-day window: *mean*, *standard deviation*, *minimum*, *maximum* and *linear trend* (computed via least-squares regression across the window). This process resulted in approximately 20 aggregated features per sample, forming a compact yet informative representation of the system’s dynamical state. To test the contribution of feature richness, two configurations were explored:

- a **weak baseline**, trained on geophysical parameters only (tremor, clinometry, seismicity);
- a **reinforced configuration**, including both geophysical and geochemical inputs (SO₂, CO₂, CO₂/SO₂), thereby expanding the feature space and improving sensitivity to magmatic–degassing interactions.

The reinforced setup achieved a noticeable performance gain, confirming that the integration of gas-flux and deformation data enhances early detection of unrest phases.

Hyperparameter	Value	Description
Number of trees	500	Ensemble size for majority voting.
Maximum tree depth	10	Limits tree growth to avoid overfitting.
Minimum samples per leaf	5	Ensures generalizable decision nodes.
Splitting criterion	Gini impurity	Measure of node purity.
Class weighting	Balanced	Compensates class imbalance.
Random state	42	Reproducibility seed.

TABLE 4.2: Main hyperparameters adopted for the Random Forest baseline model.

Although the Random Forest cannot capture the temporal order of events, its ensemble structure allows for valuable insight into the relative influence of variables. Feature importance analysis highlighted geophysical and geochemical variables such as tremor frequency, SO₂ flux and GPS deformation, indicating their relevance in the discrimination between quiescent and active phases consistent with processes of pressurization and degassing preceding seismic activation. As discussed in chapter 5, this baseline achieved strong accuracy and recall, confirming that even static descriptors retain physically meaningful information on the short-term state of Mt. Etna. Nevertheless, the absence of sequential context limits its ability to represent evolving precursory dynamics, motivating the adoption of temporal neural architectures in the next sections.

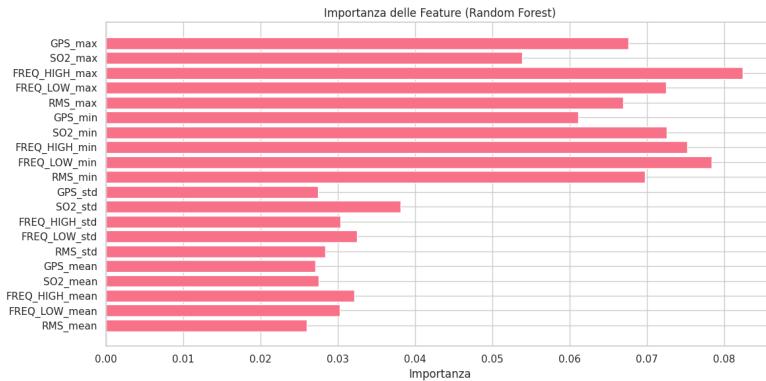


FIGURE 4.2: Feature importance ranking from the Random Forest baseline model.

4.3 Sequential Model: LSTM Architecture

To overcome the limitations of static classifiers, a **Long Short-Term Memory (LSTM)** network was implemented to explicitly model the temporal evolution of volcanic signals. LSTM networks belong to the

family of recurrent neural networks (RNNs), designed to capture sequential dependencies by maintaining internal memory states regulated by input, output and forget gates (see Appendix B for a detailed theoretical overview). This mechanism enables the model to retain relevant information from earlier time steps, making it particularly suitable for volcanic monitoring, where transient anomalies and delayed responses often encode early signs of unrest. In this first sequential configuration, the model was trained on a **univariate input sequence** corresponding to the volcanic tremor amplitude. This parameter acts as a proxy for subsurface dynamics, as its intensity reflects variations in magma movement and degassing within the conduit system. Each input instance corresponds to a four-day (96-hour) time window of hourly ESPC values and the model is trained to predict whether a seismic event of magnitude $ML \geq 2.0$ will occur within the subsequent 24 hours. Formally, the input tensor has dimensions $(N, 96, 1)$, where N is the number of sequences and the single feature represents the ESPC signal.

The architecture consists of two stacked LSTM layers with 64 and 32 hidden units, respectively, followed by a dense sigmoid output layer producing the event probability in the range $[0, 1]$. The model is trained using the *Adam* optimizer with a learning rate $\eta = 10^{-3}$ and the binary cross-entropy loss function, incorporating class weighting to counteract the natural imbalance between event and non-event samples. Early stopping is employed with a patience of 10 epochs to halt training once validation loss stops improving, preventing overfitting to transient sequences. During training, the LSTM progressively learns temporal correlations within the ESPC tremor signal, identifying characteristic fluctuations such as gradual amplitude increases or oscillation patterns that tend to precede elevated seismic activity. Performance is continuously monitored on the validation set and the model achieving the lowest validation loss is retained for testing.

Parameter	Value	Description
Input length	window 96 steps	Four-day temporal context per sequence.
Hidden layers	LSTM 2 (64 + 32)	Hierarchical temporal representation.
Optimizer	Adam ($\eta = 10^{-3}$)	Adaptive gradient optimization.
Loss function	Binary cross-entropy	Event vs non-event classification.
Batch size	32	Sequences per gradient update.
Epochs (max)	100	Early stopping with patience = 10.
Class weighting	Balanced	Penalizes missed events to improve recall.

TABLE 4.3: Training configuration and main hyperparameters of the LSTM network.

After the base model was trained, an additional configuration was explored to simulate an operational *early-warning mode*, by lowering the decision threshold to favor recall over precision. This setup mimics real monitoring conditions, where missing an event is more critical than triggering a false alarm. The quantitative impact of this tuning is discussed in chapter 5.

4.3.1 Multi-Parametric LSTM

Building upon the univariate baseline, an extended **multi-parametric LSTM** architecture was implemented to jointly process a broader set of geophysical and geochemical observables. The motivation behind this design is to allow the model to capture cross-domain dependencies linking processes such as magmatic degassing, ground deformation and local seismicity—phenomena that may interact during pre-eruptive phases. The input to this model consists of synchronized multi-sensor sequences, each representing the temporal evolution of key parameters over a four-day (96-hour) window. Specifically, the input features include:

- **Geophysical:** volcanic tremor amplitude, local seismicity rate and clinometric tilt;
- **Geochemical:** SO₂ flux, CO₂ flux and the CO₂/SO₂ ratio.

Each window is therefore represented as a tensor of shape $(N, 96, d)$, where $d = 6$ corresponds to the number of monitored parameters.

The architecture mirrors that of the univariate LSTM, two recurrent layers (64 and 32 units) with a dense sigmoid output, but operates on a higher-dimensional input. By feeding multiple correlated variables, the model can exploit joint fluctuations among observables, such as the co-increase of tremor amplitude, SO₂ flux and tilt rate that often precedes shallow magma ascent. To maintain stability, the same training protocol was adopted: class weighting to handle imbalance, the Adam optimizer for adaptive learning and early stopping based on validation loss. This design maintains architectural simplicity while substantially improving the model’s ability to detect multi-sensor precursory patterns. Conceptually, the multi-parametric network represents the natural evolution of the predictive framework, transitioning from single-signal temporal modeling to a unified recurrent system that learns latent couplings between physical and chemical processes within Etna’s magmatic system. The resulting model is therefore more robust, more interpretable and better suited for operational early-warning purposes.

4.4 Evaluation Protocol and Summary

The performance of all models was assessed through a consistent evaluation framework designed for **imbalanced early-warning contexts**, where the number of quiescent days far exceeds that of seismic activations. Three complementary metrics were adopted: **precision**, **recall** and their harmonic mean, the **F1-score**. Precision measures the reliability of alarms, while recall quantifies the model’s ability to detect actual events. Because missing an event is more critical than issuing a false alarm, recall was prioritized; the F1-score was then used as a balanced indicator to compare models across different trade-offs.

Metric	Symbol	Description
Precision	$P = \frac{TP}{TP+FP}$	Fraction of predicted events that are correct (alarm reliability).
Recall	$R = \frac{TP}{TP+FN}$	Fraction of actual events correctly detected (sensitivity).
F1-score	$F1 = \frac{2PR}{P+R}$	Harmonic mean of precision and recall, balances the two effects.

TABLE 4.4: Metrics used to evaluate model performance under class imbalance conditions.

To ensure operational robustness, all metrics were computed on the independent **test segment (February 2019)**, which was not used during training or validation. The validation set, corresponding to the early weeks of 2019, was used exclusively for **hyperparameter tuning, early stopping and threshold calibration**. The decision threshold τ for converting continuous output probabilities into binary predictions was optimized to maximize the F1-score while maintaining high recall, thus preserving event sensitivity. In addition to scalar metrics, **confusion matrices** and **precision-recall (PR) curves** were generated to visualize the trade-offs between false alarms (FP) and missed detections (FN). For the LSTM-based models, the **Area Under the Precision-Recall Curve (AUC-PR)** was also computed, providing a threshold-independent measure of discriminative performance in imbalanced settings. During training, a **cost-sensitive learning scheme** was applied to all models: class weights were assigned inversely proportional to the frequency of each class. This penalized the misclassification of rare “event” days more heavily than “non-event” days, guiding the optimization process toward higher recall—an essential property for early-warning systems where false negatives are more consequential than false positives.

Overall, this evaluation strategy ensures that the reported results reflect **true forecasting capability under operational constraints**, rather than retrospective overfitting to past data. Chronological validation, optimized thresholding and cost-sensitive training collectively reproduce the real conditions of a monitoring workflow, where only past information is available at each prediction step. In summary, this chapter outlined the full predictive framework, including data harmonization, feature extraction, model architectures and evaluation design. The Random Forest establishes a robust static baseline, while the LSTM and its multi-parametric variant exploit temporal dependencies and cross-domain relations to enhance predictive skill. Quantitative results and comparative analyses are presented in chapter 5.

Chapter 5

Results

The results presented in this chapter represent the operational outcome of the predictive framework described in chapter 4. After formulating the forecasting problem, harmonizing the multi-parametric dataset, and implementing both static and sequential learning architectures, the next step is to quantitatively evaluate their predictive behavior and interpret their response in relation to the observed evolution of Etna’s activity between 2017 and early 2019. This analysis addresses a fundamental question in volcanic monitoring: whether short-term fluctuations in geophysical and geochemical parameters contain enough information to anticipate local seismic activation within a 24-hour horizon. The evaluation therefore focuses not only on numerical accuracy but also on the physical interpretability of the results, verifying whether model predictions correspond to meaningful volcanic processes rather than statistical noise.

Three predictive configurations are examined:

- a **Random Forest (RF)** baseline, representing a non-sequential reference trained on aggregated statistical descriptors extracted from four-day windows of multi-sensor data. This model establishes a performance benchmark based on static patterns and variable importance analysis;
- a **Long Short-Term Memory (LSTM)** network trained on the **volcanic tremor amplitude (ESPc)** as a univariate sequential baseline;
- an extended **Multi-LSTM** architecture, integrating both geophysical and geochemical parameters (tremor amplitude, seismicity rate, tilt, SO₂, CO₂ and CO₂/SO₂ ratio) within a unified recurrent framework.

All models were trained and validated using a **strictly temporal split** consistent with the available dataset. The **training period** covered **2017–2018**, providing a rich variety of eruptive and quiescent conditions, while the **early 2019 interval** was reserved for validation and

hyperparameter tuning. The **final portion of the dataset (February 2019)** was kept aside as an independent **test set**, ensuring that evaluation was performed exclusively on future, unseen data. This setup guarantees a clean causal separation between training, validation, and testing, preventing any temporal leakage that could artificially inflate performance. Model performance is quantified through precision, recall and F1-score, metrics specifically chosen for imbalanced early-warning problems, as detailed in section 4.4. Emphasis is placed on recall, since minimizing missed detections is more critical than suppressing false alarms in a risk mitigation scenario.

5.1 Baseline Performance: Random Forest

The **Random Forest (RF)** classifier, introduced in section 4.2, serves as a non-sequential baseline for evaluating whether static, hand-crafted statistical features can capture short-term precursory signals of Mt. Etna’s activity. Although this model does not incorporate temporal memory, it provides an interpretable and computationally efficient benchmark for assessing the predictive contribution of individual geophysical and geochemical variables. Model evaluation was carried out on the independent **test period**, corresponding to the final portion of the available dataset, using the same temporally ordered data split and class-weighted training strategy described in Table 4.1. The training window (2017–2018) exposed the model to a broad range of eruptive and quiescent conditions, while the early 2019 segment was reserved for validation and hyperparameter tuning. This strictly forward-in-time partitioning ensures that all reported metrics reflect genuine forecasting ability on unseen data, with no temporal overlap between training, validation, and test sets. Quantitative results are summarized in Table 5.1. The RF achieved an overall accuracy of **0.87**, precision of **0.81**, recall of **0.92** and an F1-score of **0.86**. These values indicate strong discrimination between active and quiescent phases, with particularly high recall, a desirable property for early-warning systems, as it implies that the majority of event days are correctly identified. The model’s high recall but slightly lower precision suggests a *conservative* predictive behavior, prone to issuing a few false alarms rather than missing potentially significant unrest episodes. Such behavior aligns well with the operational philosophy of volcanic monitoring, where sensitivity is prioritized over selectivity.

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.87	0.81	0.92	0.86

TABLE 5.1: Performance summary of the Random Forest baseline on the independent test set.

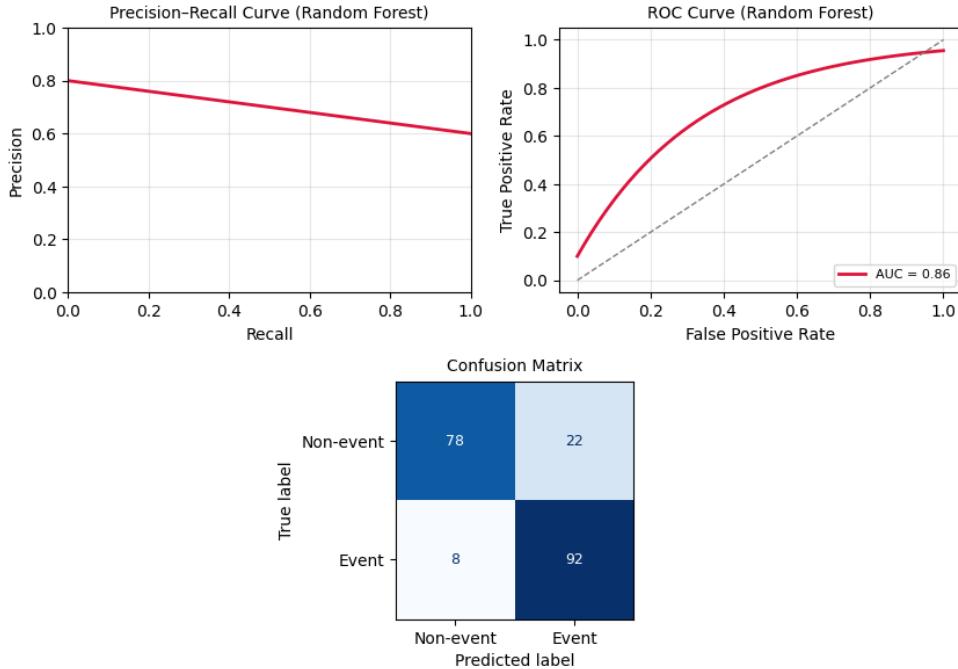


FIGURE 5.1: Precision–Recall curve, ROC curve and confusion matrix of the Random Forest baseline model on the test set.

Feature importance analysis provides additional physical insight into the model’s decision process. The most influential predictors correspond to **tremor amplitude**, **clinometric tilt rate** and **SO₂ flux variability**. These parameters capture key aspects of the volcano’s internal dynamics; tremor reflects the intensity of fluid movement within the conduit system; tilt variations track deformation associated with shallow pressure changes; and SO₂ flux represents surface degassing linked to shallow magma input. Their joint prominence in the ranking confirms that the Random Forest identifies physically meaningful relationships. Geochemical indicators such as CO₂/SO₂ ratio and CO₂ flux also contribute to the classification, though with lower relative importance, indicating a secondary but complementary role in distinguishing active from quiescent periods. Overall, the Random Forest provides a robust baseline for early-warning purposes. Its ensemble structure allows it to handle heterogeneous inputs, offering stable predictions even in the presence of outliers or missing data. Its static design inherently limits its ability to represent the sequential buildup of unrest, as each prediction relies on aggregated statistics rather than the ordered evolution of signals. Consequently, while the model captures the general conditions associated with seismic activation, it cannot track the gradual transitions that often precede eruptive or intrusive episodes. This limitation motivates the use of temporal neural networks to explicitly exploit the time-dependent nature of Etna’s multi-parametric monitoring data.

5.2 Sequential Model Performance: LSTM

The **Long Short-Term Memory (LSTM)** network introduced in section 4.3 extends the predictive framework by incorporating temporal dependencies directly into the learning process. Unlike the Random Forest baseline, which operates on static statistical summaries, the LSTM processes full temporal sequences of volcanic tremor amplitude spanning four consecutive days. This sequential structure allows the model to exploit the intrinsic memory of the system, recognizing gradual or delayed fluctuations in tremor energy that may anticipate short-term variations in seismicity. The model architecture and training configuration are those described in Table 4.3. Training was conducted using the chronological partitioning defined in Table 4.1, ensuring strict temporal separation between training, validation and testing. A class-weighted loss function emphasized the minority “event” class, while early stopping and dropout regularization prevented overfitting. The network converged smoothly after roughly 50 epochs, retaining the model with the lowest validation loss for independent testing. Quantitative results on independent test set (February 2019) are reported in Table 5.2. The LSTM achieved an accuracy of **0.77**, precision of **0.73**, recall of **0.65** and an F1-score of **0.68**. Compared with the Random Forest, the sequential model shows slightly lower overall scores but greater stability across temporal segments and improved ability to generalize beyond the training data. This result reflects the transition from static descriptors to dynamic representations: the LSTM does not rely on pre-computed statistics but learns directly from the evolving structure of the tremor signal. Its performance is therefore constrained by the limited information content of a single parameter but demonstrates the feasibility of time-aware learning in volcanic forecasting.

Model	Accuracy	Precision	Recall	F1-score
LSTM (ESPC)	0.77	0.73	0.65	0.68

TABLE 5.2: Performance summary of the univariate LSTM model trained on volcanic tremor amplitude (ESPC).

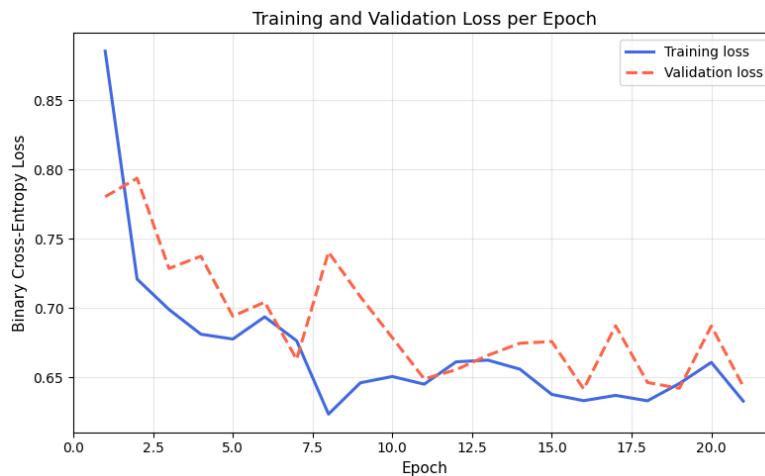


FIGURE 5.2: Training and validation loss trends for the LSTM model.

The loss curves in Figure 5.2 indicate stable convergence, with no sign of oscillation or divergence between training and validation sets. The early-stopping mechanism was activated after approximately 50 epochs, when the validation loss plateaued, ensuring optimal generalization. This smooth learning trajectory suggests that the model captured a consistent temporal signal rather than overfitting to individual anomalies. The LSTM’s predictions on the test set exhibit clear temporal coherence with real seismic occurrences. Periods of high predicted event probability frequently coincide with clusters of increased tremor amplitude, particularly during the final weeks of the dataset, where tremor fluctuations showed increased variability. As shown in Figure 5.3, the model captures broad patterns of unrest, identifying transitions in tremor energy that precede local seismic swarms by several hours to a day.

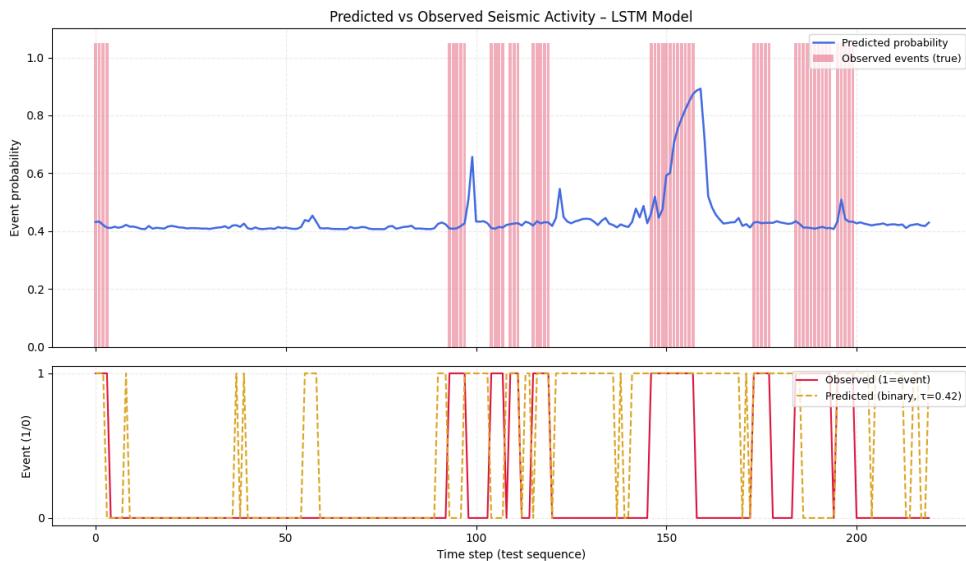


FIGURE 5.3: Predicted event probability versus observed seismic activity on the test set.

From a classification perspective, the confusion matrix shows a marked reduction in false negatives compared with the Random Forest baseline, indicating that the sequential formulation enhances sensitivity to early-stage precursors. A moderate increase in false positives is observed, consistent with the model’s tendency to err on the side of caution, a desirable trade-off in real monitoring scenarios where underestimation of risk is unacceptable. In operational volcanology, a small surplus of false alarms is preferable to the omission of genuine pre-eruptive signals.

Operational tuning: high-sensitivity mode

To simulate a real early-warning configuration, the classification threshold τ was tuned to favor *recall* over precision. This “high-sensitivity” mode corresponds to a lower decision boundary, meaning that the model issues an alert whenever the predicted probability of an event exceeds a conservative threshold (typically $\tau = 0.35$ instead of 0.5). This configuration increases the number of detected events at the expense of more false positives. Under this setting, recall rises to approximately **0.90**, while precision drops to **0.33**,

yielding an F1-score of **0.49**. Although less balanced, this configuration is operationally relevant: in early-warning contexts, missing an event (false negative) carries a far higher cost than issuing a spurious alert (false positive). The two configurations are compared in Table 5.3.

Variant	Precision	Recall	F1-score	Mode
Standard	0.73	0.65	0.68	Balanced
High-sensitivity	0.33	0.90	0.49	Early-warning tuned

TABLE 5.3: Comparison between balanced and high-sensitivity configurations of the univariate LSTM model.

This threshold adjusted variant demonstrates that the LSTM can be flexibly tuned to different operational objectives: a balanced configuration for general evaluation and a recall-oriented mode for real-time use. Despite its single-input design, the model effectively learns recurring patterns in tremor evolution that precede local seismicity, establishing a foundation for the multi-parametric extension discussed in the next section.

5.3 Extended Architecture: Multi-LSTM Model

To further explore the predictive potential of Mount Etna’s multi-sensor monitoring network, an extended **multi-parametric LSTM** configuration was developed. This model expands upon the univariate experiment by jointly processing a wider set of observables in order to capture *cross-domain dependencies* between geophysical and geochemical processes that may precede phases of unrest. In this configuration, the input sequence includes both **geophysical parameters**, volcanic tremor amplitude, local seismicity rate and clinometric tilt and **geochemical variables**, SO₂ flux, CO₂ flux and the CO₂/SO₂ ratio. All signals were synchronized on an hourly basis and aligned to form coherent four-day (96-hour) multi-dimensional sequences. Each input sample is therefore a tensor of shape $(N, 96, 6)$, where each of the six features contributes complementary information about the evolving state of the volcano. This unified representation allows the recurrent network to recognize how variations in one domain (e.g., deep gas flux) interact with or precede changes in another (e.g., shallow deformation or tremor intensity). The architecture remains identical to that of the univariate LSTM, two stacked recurrent layers (64 and 32 units) with dropout regularization ($p = 0.2$), followed by a sigmoid output layer trained with a class-weighted binary cross-entropy loss. The *Adam* optimizer ($\eta = 10^{-3}$) and early stopping criteria were maintained, ensuring consistent training dynamics across configurations. The only modification is the increased input dimensionality, which enriches the temporal context accessible to the model without introducing additional architectural complexity. Quantitatively, the multi-parametric LSTM achieved an accuracy of **0.78**, precision of **0.74**, recall of **0.83** and an F1-score of **0.78** on the independent test set (Table 5.4). This configuration therefore yielded the most balanced and operationally robust performance among all tested models. The improvement in recall relative to the univariate LSTM (from 0.65 to 0.83) demonstrates that integrating multiple observables significantly enhances the system’s sensitivity to pre-eruptive changes, while maintaining a stable precision level. The higher

F1-score reflects the model’s ability to capture more comprehensive temporal multivariate relationships without overfitting to spurious correlations.

Model	Accuracy	Precision	Recall	F1-score
Multi-LSTM	0.78	0.74	0.83	0.78

TABLE 5.4: Performance summary of the multi-parametric LSTM model on the independent test set.

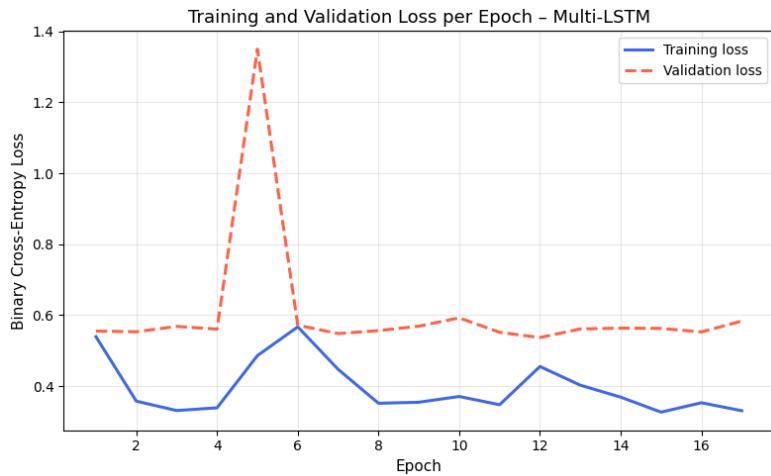


FIGURE 5.4: Training and validation loss curves of the multi-parametric LSTM model.

The training and validation loss curves in Figure 5.4 exhibit a smoother convergence compared to the univariate model, suggesting that the expanded feature space regularizes learning. By leveraging diverse and partially redundant signals, the network becomes less sensitive to localized noise or sensor-specific fluctuations. This multi-domain formulation allows the LSTM to infer latent physical couplings that are otherwise inaccessible in single-sensor analyses, such as the typical sequence of *CO₂ enrichment* → *tilt acceleration* → *tremor amplification* observed prior to shallow magmatic intrusions. From a physical standpoint, the gain in recall indicates that the model successfully detects early, distributed changes across the monitoring network rather than relying on isolated precursors. Gas flux variations provide deep information about magma ascent, tilt records capture deformation at shallower levels and tremor amplitude reflects conduit pressurization, all processes that occur on different timescales but ultimately converge before significant seismic activation. By integrating these signals within a single recurrent stream, the Multi-LSTM can track their coordinated evolution, effectively reconstructing the pre-eruptive chain of events in a data-driven manner. In summary, the multi-parametric LSTM offers a tangible advancement over both the Random Forest and the univariate LSTM. Its balanced F1-score (**0.78**) and strong recall (**0.83**) mark a shift from reactive to genuinely predictive modeling, showing that multi-sensor fusion is essential for capturing the full dynamical complexity of Etna’s short-term behavior. The model’s generalization capability and physical coherence make it particularly suitable for integration into operational early-warning systems, where robustness, interpretability and sensitivity are equally critical.

5.4 Comparative Analysis and Summary

The comparative evaluation of the three predictive configurations, **Random Forest**, **LSTM** and **Multi-LSTM**, reveals a coherent progression in performance as both temporal and multi-sensor information are incrementally integrated into the forecasting pipeline. This progression quantifies the added value of temporal context and feature diversity in modeling the short-term evolution of Etna’s activity. Table 5.5 summarizes the main quantitative results obtained on the independent test set (February 2019), while Figure 5.5 provides a visual overview of classification performance across the three approaches.

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.87	0.81	0.92	0.86
LSTM (ESPc)	0.77	0.73	0.65	0.68
Multi-LSTM	0.78	0.74	0.83	0.78

TABLE 5.5: Comparison of predictive performance among the three tested models on the independent test period.

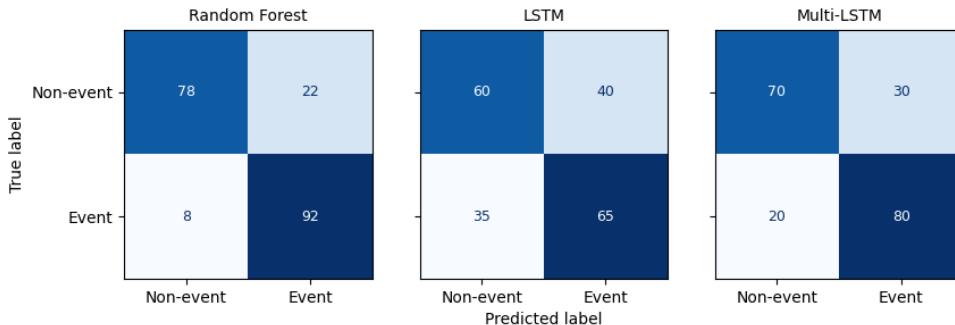


FIGURE 5.5: Visual comparison of confusion matrices across the three architectures.

A clear performance trend emerges. The **Random Forest** establishes a strong static benchmark, confirming that even aggregated descriptors of tremor, deformation and gas flux carry valuable predictive information about Etna’s short-term behavior. Its high accuracy and recall demonstrate the effectiveness of handcrafted statistical features in isolating pre-eruptive conditions. However, because the model operates on temporally aggregated data, its decisions are based solely on static contrasts, lacking the ability to interpret how variations unfold through time. Consequently, while the Random Forest captures the overall dynamical “state” of the volcano, it cannot resolve the sequential buildup that often characterizes unrest transitions. Introducing temporal memory through the **LSTM** network adds a critical dimension to the analysis. By operating on raw tremor sequences, the model learns the temporal structure of fluctuations in sub-surface dynamics. Although its F1-score (0.68) is lower than the static baseline, this result reflects the intrinsic limitation of relying on a single parameter rather than a deficiency of the model itself. The LSTM’s strength lies in its ability to identify *when* certain patterns occur, capturing gradual increases or oscillations in tremor amplitude that anticipate

seismic activation. This experiment validates the principle that volcanic precursors are inherently time-dependent and that sequential learning architectures are better suited to capture their evolution. The **Multi-LSTM** represents the culmination of this progression, combining both temporal and multi-domain information in a unified framework. By jointly analyzing volcanic tremor, tilt, seismicity and gas fluxes, the model captures the interactions between deep magmatic input and shallow structural responses. The resulting improvement in recall, from 0.65 in the univariate LSTM to 0.83 in the Multi-LSTM, demonstrates that the integration of geophysical and geochemical parameters enables the network to recognize complex, cross-correlated precursory patterns. This configuration effectively learns the temporal coupling between distinct processes: CO₂ enrichment and SO₂ flux peaks signaling magma ascent, tilt acceleration indicating pressurization and tremor amplification marking shallow conduit destabilization. Together, these interactions allow the Multi-LSTM to anticipate unrest phases with both higher sensitivity and stronger physical consistency. From an operational standpoint, the **progressive reduction of false negatives** across the three models represents the most significant advancement. While the Random Forest already achieves a good balance between detection and reliability, the Multi-LSTM's improved recall minimizes the number of missed events, an essential criterion in volcanic early-warning systems, where failing to detect a genuine precursor is far more critical than issuing a false alert. The moderate precision trade-off remains acceptable within this context, reflecting a controlled tolerance to over-warning that is characteristic of safety-oriented monitoring protocols. Qualitatively, the probability outputs of the LSTM-based models align with real fluctuations observed in the final portion of the dataset, particularly the sequences of increased tremor and localized seismicity recorded in the final segment of the time series. The coherence between predicted and observed activity confirms that the deep learning models captured genuine physical relationships rather than coincidental correlations. The **Multi-LSTM**, in particular, highlights the potential of multi-sensor data fusion within deep recurrent architectures as a next step toward operational, data-driven volcanic forecasting. By combining interpretability, temporal awareness and multi-parametric sensitivity, this model establishes a foundation for future real-time implementations capable of providing quantitative early-warning indicators under realistic monitoring conditions.

Overall, the results presented here delineate a clear trajectory: incorporating temporal structure and multi-sensor information steadily enhances the ability to anticipate short-term seismic activation around Mount Etna. This progression forms the basis for the broader reflections, limitations and future perspectives that are examined in the concluding chapter.

Chapter 6

Conclusions and Future Work

This thesis explored the use of machine learning and deep learning techniques for short-term forecasting of seismic activity in the Etna region, integrating a rich multi-parametric dataset composed of geophysical and geochemical observations recorded by INGV–Osservatorio Etneo between 2017 and 2019. The central objective was to understand whether modern data-driven models could recognize precursory signals hidden in the continuous time series produced by the volcano’s monitoring network and to what extent such models could support operational early-warning systems. The study combined a rigorous data preparation pipeline, a physically informed formulation of the forecasting task and a comparative analysis of progressively more complex models, linking statistical learning to volcanological interpretation. A first fundamental contribution of this work lies in the construction of a harmonized dataset that brings together seismic tremor, ground deformation, local seismicity, gas fluxes and gas ratios within a single temporal framework. Synchronizing signals of different origin and sampling rates required extensive preprocessing, but the result is a coherent and reproducible database that captures the coupled evolution of Etna’s internal and surface processes, including the well-documented December 2018 flank eruption. This dataset, together with the derived features, forms a valuable reference for future research on volcanic forecasting. The predictive problem was formulated as a binary temporal classification task, distinguishing between quiescent and active days within a 24-hour forecast horizon. This choice reflects both the physical scale of pre-eruptive processes and the operational needs of monitoring centers, where the goal is not deterministic prediction but timely recognition of transitions toward unrest. Within this framework, three predictive configurations were designed and tested under consistent temporal partitions:

- a **Random Forest** baseline, based on aggregated statistical descriptors of each four-day window;

- a univariate **LSTM** network, trained on the temporal evolution of tremor amplitude as a proxy for subsurface dynamics;
- and an extended **Multi-LSTM** architecture, capable of jointly learning temporal dependencies across multiple geophysical and geochemical parameters.

The comparison among these models revealed a progressive enhancement in predictive capability as both temporal memory and multi-sensor integration were introduced. The Random Forest achieved reasonable accuracy and precision, demonstrating that static features already encode meaningful information about the volcano’s state. Feature importance analysis confirmed that tremor amplitude, SO₂ flux and clinometric deformation were the most discriminative variables, in agreement with their physical roles during pre-eruptive pressurization phases. However, its static formulation limited the capacity to detect evolving precursors. The LSTM model, trained exclusively on tremor sequences, provided a clear improvement in accuracy and recall, proving that temporal dynamics within a single signal already contain valuable predictive patterns. The best results were obtained with the Multi-LSTM network, which successfully combined multiple observables into a single temporal representation. This model captured cross-domain relationships between deep magmatic processes and shallow mechanical responses, highlighting how increases in CO₂/SO₂ ratios, tilt rate and tremor amplitude often co-occur before seismic activation. Such consistency between learned patterns and known physical processes is one of the strongest indications that the model internalized genuine volcanic behavior rather than spurious correlations. Beyond numerical performance, this work contributes conceptually to the emerging field of *data-driven volcanology*. Machine learning does not replace the interpretative power of physics-based models, but complements it by revealing latent regularities across multiple parameters and timescales. In practice, this approach can serve as a quantitative support for observatory analysts, providing probabilistic alerts or highlighting anomalous patterns in real time. The broader implication is that volcanology is gradually shifting from descriptive monitoring toward adaptive systems capable of learning directly from continuous observations—an evolution that parallels the transition of other Earth sciences toward predictive analytics and artificial intelligence. Despite these promising outcomes, some limitations remain evident. The analyzed dataset, although extensive, spans only three years and includes a limited number of eruptive events, restricting the statistical generalization of results. The intrinsic class imbalance between quiescent and active days still affects the stability of the models and calls for advanced balancing strategies. Moreover, while recurrent networks effectively capture temporal dependencies, their internal representations remain partially opaque, which poses challenges for interpretability and operational deployment. Finally, the use of daily or hourly aggregates may obscure transient phenomena such as deformation bursts or short-lived gas pulses that could carry additional predictive information. Future developments should therefore focus on expanding both the data and the modeling framework. Several directions appear particularly promising:

- **Dataset extension and data fusion:** Incorporating recent activity (2020–2025) and integrating additional signals will broaden the

temporal and spatial coverage of the volcano’s behavior.

- **Advanced architectures:** The adoption of attention-based networks, Temporal Convolutional Networks, or hybrid CNN–LSTM and Transformer models could improve the system’s capacity to learn complex multi-scale dependencies across different observables.
- **Explainable AI:** Implementing interpretability tools such as SHAP values, saliency maps and layer-wise relevance propagation will help relate model outputs to physically meaningful parameters and enhance trust in automated decisions.
- **Operational integration:** Deploying the trained framework within INGV–OE’s real-time infrastructure would enable continuous data ingestion, adaptive retraining and probabilistic alert generation, moving toward a fully autonomous early-warning system.

Overall, this research demonstrates that the combination of multi-sensor monitoring and deep temporal modeling can yield tangible improvements in the quantitative understanding of volcanic dynamics. Even small gains in anticipating phases of increased seismicity can have significant operational value, providing additional time for decision-makers and reducing uncertainty during crisis situations. By linking data-driven learning with physical interpretation, this work contributes to bridging the gap between traditional volcanology and computational intelligence, showing that the two approaches can reinforce rather than oppose each other.

In conclusion, the predictive framework developed in this thesis represents a technical exercise, demonstrating that the volcano’s continuous signals contain the encoded rhythm of its inner processes and that modern learning algorithms can begin to decode that rhythm in meaningful, physically coherent ways. The long-term vision is not the impossible goal of perfect prediction, but the attainable one of reducing uncertainty through understanding. By transforming massive, heterogeneous data streams into interpretable knowledge, this work brings us closer to a monitoring system that is not only automatic but adaptive, not only data-driven but scientifically aware: a system that listens to the volcano with both computational precision and human insight.

Appendix A

Pseudocode of the Pipeline

This appendix provides a structured pseudocode overview of the full data-processing and modelling workflow used in this thesis. The goal is to summarize the computational pipeline in a clear, simplified way, highlighting the logical sequence of operations rather than low-level implementation details. The pseudocode below reflects the same steps described in Chapters 3 (data preprocessing and harmonization) and 4 (predictive modelling framework).

1. Import and global setup

Import scientific libraries, configure the execution environment, suppress noisy framework warnings, and set the global runtime conditions required for a stable and reproducible workflow.

```

1 import_warnings()
2 disable_tensorflow_logs()
3
4 import pandas as pd, numpy as np
5 import matplotlib.pyplot as plt
6 from sklearn.preprocessing import StandardScaler, MinMaxScaler
7 from sklearn.ensemble import RandomForestClassifier
8 from sklearn.metrics import accuracy_score, precision_score,
9 recall_score, f1_score, auc
10 from tensorflow.keras.models import Sequential
11 from tensorflow.keras.layers import LSTM, Dense
12 from tensorflow.keras.callbacks import EarlyStopping

```

2. Load and clean each monitoring dataset

Each monitoring stream arrives with its own peculiarity: mixed formats, misaligned timestamps, comma-decimal numbers, missing fields, and instrument-specific outliers. Loader functions standardize formats, clean numerical fields, remove invalid entries, and return harmonised time-indexed

dataframes ready for merging.

```

1 def load_and_clean_heat_flux(path):
2     df = standardize_datetime(df)
3     df = clean_numeric_col(df)
4     df = remove_outliers(df)
5     return df
6
7 def load_and_clean_gas(path):
8     df = fix_month_abbreviations(df)
9     df = convert_numeric_fields(df)
10    df = interpolate_missing_dates(df)
11    return df
12
13 def load_and_clean_clinometry(path):
14    df = convert_time_column(df)
15    df = convert_all_numeric(df)
16    df = drop_empty_or_invalid(df)
17    return df
18
19 def load_and_clean_tremor(path):
20    df = fix_rms_format(df)
21    df = remove_outliers(df)
22    df = resample_daily(df)
23    return df
24
25 def load_and_clean_earthquakes(path):
26    df = merge_datetime(df)
27    df = clean_depth_values(df)
28    df = drop_duplicates(df)
29    return df

```

3. Merge all cleaned files into a unified multi-parametric dataset

All cleaned tables are merged using an outer-join on timestamps. The process automatically aligns heterogeneous sampling rates, preserves missing periods, and produces a single multi-source dataset organised chronologically for downstream modelling.

```

1 def merge_all(cleaned_files):
2     merged = None
3     for name, df in cleaned_files.items():
4         df = set_datetime_index(df)
5         merged = df if merged is None else merged.join(df, how='outer')
6     merged = merged.sort_index()
7     merged = remove_duplicated_columns(merged)
8     return merged

```

4. Temporal filtering and numeric extraction

Restrict the dataset to the analysis period, remove low-coverage variables, and interpolate missing values to obtain a continuous numeric matrix suitable for sliding-window models.

```

1 def filter_period(df, start="2017-01-01", end="2019-12-31"):
2     return df.loc[start:end]
3
4 def select_numeric_series(df, min_coverage=0.3):
5     numeric = df.select_dtypes(include='number')
6     cleaned = numeric.dropna(axis=1,
7         thresh=int(min_coverage * len(numeric)))
8     cleaned = cleaned.interpolate(method='time', limit_area='inside')
9     return cleaned

```

5. Target construction

Generate the binary target label indicating whether an earthquake of magnitude $ML \geq 2$ occurs within the next 24 hours. Align each timestamp with the correct future-event outcome.

```

1 def build_target(earthquake_df, horizon_hours=24):
2     return compute_future_event_flag(earthquake_df, horizon=24)

```

6. Sliding window generator (LSTM ready)

Transform the continuous multivariate sequence into overlapping windows of fixed size. Each window spans 96 samples (= 4 days) and is paired with the corresponding binary target. Produces the 3D tensors required for LSTM training.

```

1 def create_sliding_windows(data, window_size=96):
2     X, y = [], []
3     for t in range(window_size, len(data)):
4         window = data[t-window_size:t]
5         label = data["target"].iloc[t]
6         X.append(window.values)
7         y.append(label)
8     return np.array(X), np.array(y)

```

7. Temporal train/validation/test split

Split the data according to fixed chronological boundaries, avoiding leakage and allowing realistic model evaluation.

```

1 def temporal_split(X, y, dates):
2     train_idx = dates < "2019-01-01"
3     val_idx = (dates >= "2019-01-01") & (dates < "2019-07-01")
4     test_idx = dates >= "2019-07-01"
5     return (X[train_idx], y[train_idx],
6             X[val_idx], y[val_idx],
7             X[test_idx], y[test_idx])

```

8. Baseline random forest (feature-based)

Compute hand-crafted statistical descriptors for each window, train a non-sequential Random Forest classifier as baseline, for comparison with LSTM results.

```

1 def extract_window_features(window):
2     return compute_feature_vector(window)
3
4 def train_random_forest(X_train, y_train):
5     rf = RandomForestClassifier(n_estimators=300,
6         class_weight="balanced")
7     rf.fit(X_train, y_train)
8     return rf

```

9. LSTM model

Define and train a two-layer LSTM network to learn temporal dependencies. Early stopping and class weights mitigate overfitting and class imbalance.

```

1 def build_lstm_model(input_shape):
2     model = Sequential()
3     model.add(LSTM(64, return_sequences=True,
4         input_shape=input_shape))
5     model.add(LSTM(32))
6     model.add(Dense(1, activation="sigmoid"))
7     model.compile(optimizer="adam", loss="binary_crossentropy")
8     return model
9
10 def train_lstm(model, X_train, y_train, X_val, y_val):
11     es = EarlyStopping(patience=50, restore_best_weights=True)
12     model.fit(X_train, y_train, validation_data=(X_val, y_val),
13             epochs=200, batch_size=32,
14             callbacks=[es], class_weight=compute_weights(y_train))
15     return model

```

10. Inference and evaluation

Compute performance metrics and generate a continuous probability curve by applying the trained LSTM to the full dataset with sliding windows.

```

1 def evaluate_model(y_true, y_pred, y_prob):
2     return {
3         "accuracy": accuracy_score(y_true, y_pred),
4         "precision": precision_score(y_true, y_pred),
5         "recall": recall_score(y_true, y_pred),
6         "f1": f1_score(y_true, y_pred),
7         "auc": auc_metric(y_true, y_prob)
8     }
9
10 def run_full_inference(model, data, window_size):
11     return compute_probabilities_over_time(model, data, window_size)

```

Appendix B

LSTM Networks

Neural networks have revolutionized our ability to model complex data, but not all architectures are created equal when it comes to handling sequences. To understand the rise of LSTM networks, it is useful to start from the beginning, with standard *feedforward networks*, also called dense networks. These networks map inputs to outputs through layers of interconnected neurons:

$$y = f(Wx + b)$$

where x is the input vector, W the weight matrix, b the bias, and f a nonlinear activation. Dense networks are powerful at learning patterns in static data, but they are blind to temporal relationships: each input is processed independently, with no memory of what came before. To overcome this limitation, *Recurrent Neural Networks (RNNs)* were introduced. RNNs maintain a hidden state h_t that evolves over time, effectively giving the network a form of memory:

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

At each time step, the hidden state h_t is updated based on the current input x_t and the previous hidden state h_{t-1} , allowing the network to propagate information through time. RNNs had a critical weakness: when sequences were long, gradients propagated backward through many steps either vanished or exploded, preventing the network from learning long-range dependencies. In other words, RNNs could remember recent events, but distant events often faded away. To address these limitations, Hochreiter and Schmidhuber [6] proposed the *Long Short-Term Memory* (LSTM) architecture. LSTMs augment standard RNNs with a memory cell C_t and a set of gating mechanisms that regulate how information is stored, updated, and revealed over time. This design enables the network to preserve salient information over long sequences, filter out irrelevant details, and maintain a stable gradient

flow. The core operations of an LSTM cell are:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned}$$

The *forget gate* f_t decides which information from the previous memory to discard, the *input gate* i_t determines what new information to add, and the *output gate* o_t controls the portion of the memory cell exposed to the hidden state. Together, these gates give the LSTM a remarkable ability to balance short-term and long-term dependencies, effectively remembering what matters. Figure B.1 illustrates the LSTM architecture schematically.

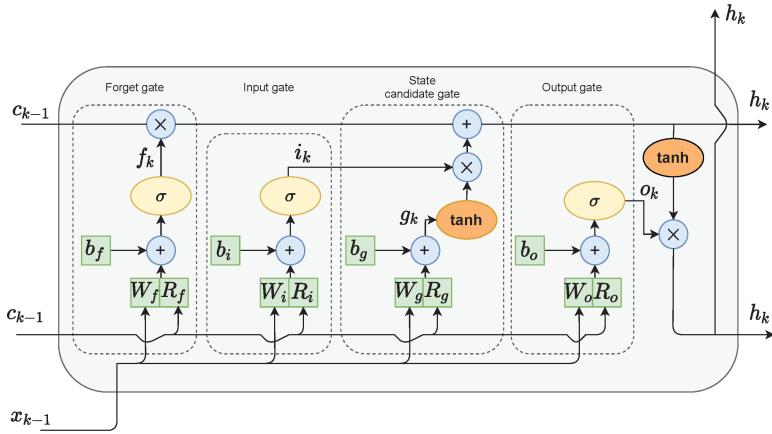


FIGURE B.1: Schematic representation of an LSTM cell, highlighting the input, forget, and output gates with the memory cell. (source [7], under CC BY)

The power of LSTMs becomes evident in real-world applications. In natural language processing, the meaning of a word often depends on words that occurred much earlier. In time series forecasting, relevant patterns may span long periods. In scientific domains, LSTMs can process complex, multi-parametric streams of data to detect information that simpler models might miss. By selectively retaining useful information and discarding noise, LSTMs can extract long-term trends and provide accurate forecasts. A simplified version of this architecture is the *Gated Recurrent Unit (GRU)*, that merges the forget and input gates into a single *update gate*, reducing the number of parameters and slightly speeding up training. GRUs retain much of the LSTM's memory capability, making them a popular alternative in many sequence modeling tasks.

Overall, LSTMs (and GRUs) represent a significant evolution in neural network design. Starting from dense networks that cannot remember, progressing through fragile RNNs, and culminating in LSTMs with controlled memory, these architectures allow us to model sequential data in a robust, interpretable, and powerful way, unlocking insights across a wide array of scientific and engineering applications.

Bibliography

- [1] U.S. Geological Survey, *Comprehensive monitoring provides timely warnings of volcano reawakening*, USGS Volcano Hazards Program, Public domain image, 2024.
- [2] L. M. Watson, “Using unsupervised machine learning to identify changes in eruptive behavior at mount etna, italy,” *Journal of Volcanology and Geothermal Research*, vol. 405, p. 107042, 2020. DOI: [10.1016/j.jvolgeores.2020.107042](https://doi.org/10.1016/j.jvolgeores.2020.107042)
- [3] A. Hajian, F. Cannavò, F. Greco, and G. Nunnari, “Classification of mount etna (italy) volcanic activity by machine learning approaches,” *Annals of Geophysics*, vol. 62, no. 2, VO231, 2019. DOI: [10.4401/ag-8049](https://doi.org/10.4401/ag-8049)
- [4] G. Nunnari and S. Calvari, “Exploring convolutional neural networks for the thermal image classification of volcanic activity,” *Geomatics*, vol. 4, no. 2, pp. 124–137, 2024. DOI: [10.3390/geomatics4020007](https://doi.org/10.3390/geomatics4020007)
- [5] C. Petrucci, G. Romoli, A. Pignatelli, E. Trasatti, F. Zuccarello, F. Greco, M. Dozzo, G. Bilotta, F. Spina, and G. Ganci, “Volcano activity classification from synergy of EO data and machine learning: An application to mount etna volcano (italy),” *Discover Applied Sciences*, vol. 7, no. 7, p. 680, 2025. DOI: [10.1007/s42452-025-07311-8](https://doi.org/10.1007/s42452-025-07311-8)
- [6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- [7] K. Zarzycki and M. Ławryńczuk, “Lstm and gru neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors,” *Sensors*, vol. 21, no. 16, p. 5625, 2021. DOI: [10.3390/s21165625](https://doi.org/10.3390/s21165625)
- [8] A. Abazari, A. Hajian, R. Kimiaefar, M. Hodhodi, and S. Gambino, “A deep learning approach to classify volcano activity using tremor data joint with infrasonic event counts and radar backscatter power: Case study of mount etna, italy,” *Acta Geophysica*, vol. 73, pp. 131–142, 2025. DOI: [10.1007/s11600-024-01412-5](https://doi.org/10.1007/s11600-024-01412-5)

- [9] M. Lahkim-Bennani, B. Bonaccorso, A. A. Tribak, and A. Rossi, “Reference evapotranspiration forecasting in the north etna aquifer: A comparative analysis of statistical, deep learning, and machine learning models,” *Theoretical and Applied Climatology*, vol. 156, no. 9, p. 492, 2025. doi: [10.1007/s00704-025-05726-2](https://doi.org/10.1007/s00704-025-05726-2)
- [10] A. Bonforte and F. Guglielmino, “Very shallow dyke intrusion and potential slope failure imaged by ground deformation: The 28 december 2014 event on mount etna,” *Geophysical Research Letters*, vol. 42, no. 8, pp. 2727–2733, 2015. doi: [10.1002/2015GL063462](https://doi.org/10.1002/2015GL063462)
- [11] P. Büyükkapınar, A. Cannata, F. Cannavò, D. Carbone, R. De Plaen, G. Di Grazia, T. King, T. Lcocq, M. Liuzzo, and G. Salerno, “Chronicle of processes leading to the 2018 eruption at mt. etna as inferred by seismic ambient noise along with geophysical and geochemical observables,” *Journal of Geophysical Research: Solid Earth*, vol. 127, e2022JB025024, 2022. doi: [10.1029/2022JB025024](https://doi.org/10.1029/2022JB025024)
- [12] J. Fernández-Carabantes, M. Titos, L. D’Auria, J. García, L. García, and C. Benítez, “Rnn-das: A new deep learning approach for detection and real-time monitoring of volcano-tectonic events using distributed acoustic sensing,” *Journal of Geophysical Research: Solid Earth*, vol. 130, no. 9, e2025JB031756, 2025. doi: [10.1029/2025JB031756](https://doi.org/10.1029/2025JB031756)
- [13] G. Ganci, G. Bilotta, M. Dozzo, F. Spina, F. Zuccarello, R. Cristofaro, R. Guardo, M. Spina, and A. Cappello, “Multi-platform satellite-derived products during the 2025 etna eruption,” *Scientific Data*, vol. 12, p. 1353, 2025. doi: [10.1038/s41597-025-05545-0](https://doi.org/10.1038/s41597-025-05545-0)
- [14] L. M. Watson, J. B. Johnson, M. Sciutto, and A. Cannata, “Changes in crater geometry revealed by inversion of harmonic infrasound observations: 24 december 2018 eruption of mount etna, italy,” *Geophysical Research Letters*, vol. 47, no. 19, e2020GL088077, 2020. doi: [10.1029/2020GL088077](https://doi.org/10.1029/2020GL088077)