ANNALS OF GEOPHYSICS, 68, 2, V223, 2025; doi:10.4401/ag-9183
OPEN ACCESS

Combining Remote Sensing and AI to Explore and Classify Volcanic Patterns

Simona Cariello[*,1,2]

[(1)] University of Catania, Department of Electrical, Electronic and Computer Science Engineering, Catania, Italy
[(2)] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo, Catania, Italy

Article history: received October 16, 2024; accepted March 27, 2025

Abstract

Monitoring volcanic hazards is essential for understanding the behavior of active volcanoes, enhancing hazard forecasting, and mitigating potential impacts. The increasing availability of satellite sensors, particularly those providing thermal infrared data with varied spatial resolutions and revisit times, has significantly enhanced monitoring capabilities. However, the vast data volume requires advanced computational tools, such as Artificial Intelligence (AI) algorithms, to enable efficient, real-time processing and analysis. Recent AI advancements have shown considerable promise in enhancing satellite-based volcanic monitoring. Here, we introduce a cascading pipeline model designed to classify high-temperature volcanic features and quantify thermal anomalies using high-resolution Sentinel-2 Multi-Spectral Instrument (MSI) imagery. This automated approach allows for timely and detailed assessments of volcanic activity. We apply the model to three highly active volcanoes: Etna and Stromboli, in the Mediterranean, and Pacaya, in Central America. Etna and Stromboli, two of the most monitored volcanoes in the Mediterranean, offer complex eruption patterns ideal for testing AI-based models. Pacaya's frequent eruptions and lava flows provide valuable comparative data for testing the model's robustness across different volcanic systems. Our objective is to classify volcanic spatial patterns and explore significant changes in thermal anomalies during periods of unrest, ultimately broadening the model's applicability beyond the Mediterranean region.

Keywords: Artificial Intelligence; Cascading model; Thermal anomalies; Mount Etna; Stromboli Island; Pacaya volcano

1. Introduction

Monitoring active volcanoes is a complex, multidisciplinary activity that integrates various scientific disciplines and technologies, including observations from ground-based, aerial, and satellite systems. This integration is crucial for understanding, predicting, and mitigating the risks associated with volcanic activity. The importance of monitoring extends beyond scientific research, playing a key role in public safety and disaster management by forecasting eruptions, issuing timely warnings, and reducing the impacts of volcanic hazards on nearby communities.

Satellite-based monitoring systems, particularly those equipped with high-resolution sensors, have proven indispensable in capturing a wide range of thermal features related to volcanic activity. These systems enable consistent and detailed observation of phenomena such as changes in crater lake temperatures, emplacement of hot volcanic material, ash and gas emissions, and activities like lava lakes, flows, degassing, and explosions (Patrick et al., 2016; Abrams et al., 1991). However, the wealth of data generated by satellite systems requires innovative processing techniques to extract meaningful insights quickly and accurately. Traditional methods, such as fixed threshold-based approaches (Tramutoli et al., 2018; Higgins and Harris, 1997), often struggle to manage the complexity and scale of these datasets.

In response to this challenge, Artificial Intelligence (AI) techniques, particularly machine learning (ML) (Bonaccorso, 2017) and deep learning (DL) (Goodfellow et al., 2016), have emerged as transformative tools for analyzing large remote sensing datasets. These AI techniques have demonstrated remarkable capabilities in detecting and mapping thermal emissions, estimating lava field volumes, and characterizing volcanic plumes (Anantrasirichai et al., 2018; Lary et al., 2016; Corradino et al., 2020; Torrisi et al., 2024). Consequently, AI-powered models have become key components of modern volcanic hazard monitoring systems, providing automated, high-accuracy analyses that far surpass traditional methods, e.g. for volcanic anomaly detection task (Amato et al., 2023), for anomaly detection task (Wang et al., 2021), cloud detection task (Fabio et al., 2024), for active fire detection (Gargiulo et al., 2019).

The integration of advanced remote sensing techniques with AI-driven approaches holds immense potential for enhancing active volcano monitoring and improving our understanding of underlying volcanic processes. In this context, ensemble architectures, particularly cascading models (Gama and Brazdil, 2000; Aziz et al., 2019), have been applied to improve classification accuracy by sequentially combining multiple ML models. These models are typically deployed on cloud computing platforms (Ray and De Sarkar, 2012; Sether, 2016), which can handle the large computational and data processing demands, making real-time volcano monitoring accessible to a broader user base.

To address these needs, an ML cascading model based on a deep learning SqueezeNet (SN) scene classifier and a random forest (RF) model (Cariello et al., 2024; Corradino et al., 2022) was developed to analyze high-temperature volcanic features and quantify thermal anomalies using high-resolution satellite data, namely from the Sentinel-2 MSI, in near real-time. This model improves the mapping, monitoring, and characterization of volcanic thermal features, providing more accurate and automated assessments.

Here, we demonstrate the potential of adopting this pre-trained ML cascading pipeline (Cariello et al., 2024) to monitor the Etna (Italy), Stromboli (Italy), and Pacaya (Guatemala) volcanoes from two key perspectives. First, the model's high accuracy enables the detection of subtle thermal signals that are often difficult to capture with current detectors, revealing previously hidden volcanic dynamics. For example, this approach can track thermal changes that signal reactivation of vents or other shifts within the volcanic system. Second, we showcase the model's ability to classify volcanic spatial patterns promptly, recognizing new volcanic activity in near real-time.

2. Satellite data

The Copernicus Sentinel-2 mission consists of two polar-orbiting satellites, Sentinel-2A (S2A) and Sentinel-2B (S2B), launched in 2015 and 2017, respectively. Both satellites follow the same sun-synchronous orbit, phased 180 degrees apart. With a revisit frequency of one satellite every 10 days, the global revisit frequency is reduced to 5 days when considering both satellites. Equipped with the Multi-Spectral Instrument (MSI), they capture imagery across 13 spectral bands. These bands offer spatial resolutions of 10 meters in the visible and near-infrared spectrum, 20 meters in the red edge and shortwave infrared (SWIR) range, and 60 meters in the atmospheric bands.

For this study, we utilized Harmonized Sentinel-2 MSI data and processed it through the Google Earth Engine (GEE). We focused on bands ranging from visible to shortwave infrared wavelengths, including the visible bands VIS1 (B2: 0.496 μm for S2A/0.492 μm for S2B), VIS2 (B3: 0.56 μm for S2A/0.55 μm for S2B), VIS3 (B4: 0.66 μm for both S2A and S2B), near-infrared (NIR: B8: 0.84 μm for S2A/0.83 μm for S2B), and shortwave infrared (SWIR: B11: 1.61 μm for both S2A and S2B, and B12: 2.20 μm for S2A/2.19 μm for S2B).

The region of interest (ROI) for each volcano is a standard square area measuring 1 km^2 ([1 × 1] km), centered at the volcano summit. However, this ROI can be enlarged if needed based on specific volcanic activity. For each data acquisition, a false-color RGB image is generated as input for the image classifier. This is done by using

Top-of-Atmosphere (TOA) data from the Near-Infrared (NIR), Short-Wave Infrared 1 (SWIR1), and Short-Wave Infrared 2 (SWIR2) bands. The resulting image is normalized using a z-score method (Eesa and Arabo, 2017) and then converted into JPEG format with a size of [224, 224, 3], where pixel values are scaled within the 0-255 range.

3. Method

The cascading machine learning approach was adopted to automatically detect and classify volcanic thermal activity from Sentinel-2 MSI imagery, while also quantifying the spatial extent of thermal anomalies, following the model already trained in (Cariello et al., 2024). This cascading pipeline integrates two machine learning models: a deep learning-based SqueezeNet (SN) classifier and a random forest (RF) model, working together to enhance the detection and quantification of high-temperature volcanic features using high-resolution satellite data.

The SN classifier is responsible for identifying global volcanic activity classes by analyzing spatial and spectral features from satellite imagery. Following this, the RF model refines the classification and maps thermal anomalies at the pixel level based on spectral attributes. The efficacy of this cascading approach was validated through its application to scenes from ten volcanoes worldwide, using data from the Sentinel-2 MSI satellite imagery archive (Cariello et al., 2024).

The implementation was carried out using Google Earth Engine (GEE) and Google Colaboratory (Colab), platforms well-suited for large-scale data analysis and machine learning applications. This two-stage machine learning architecture, structured as a "top-down" cascading model, has proven to be highly effective, achieving an overall accuracy of 95% (Cariello et al., 2024). Once the process is completed, the model is capable of classifying scenes into four categories: Extended Volcanic Thermal Anomalies (ETA), Isolated Volcanic Thermal Anomalies (ITA), No Volcanic Activity (NVA), and Cloudy-Sky Condition (CSC), and in cases where thermal anomalies are detected, quantifies their spatial extent, as illustrated in Fig. 1.

The model's high accuracy enables it to detect thermal signals that may go unnoticed by conventional detection systems, providing valuable insights into volcanic phenomena. The cascading architecture allows for the efficient processing of large datasets, filtering out irrelevant scenes, such as those affected by cloud cover (CSC) or without thermal anomalies (NVA), and focusing on scenes that show significant volcanic activity. This optimization reduces the time required for analysis and directs attention to data with the highest potential for meaningful information.

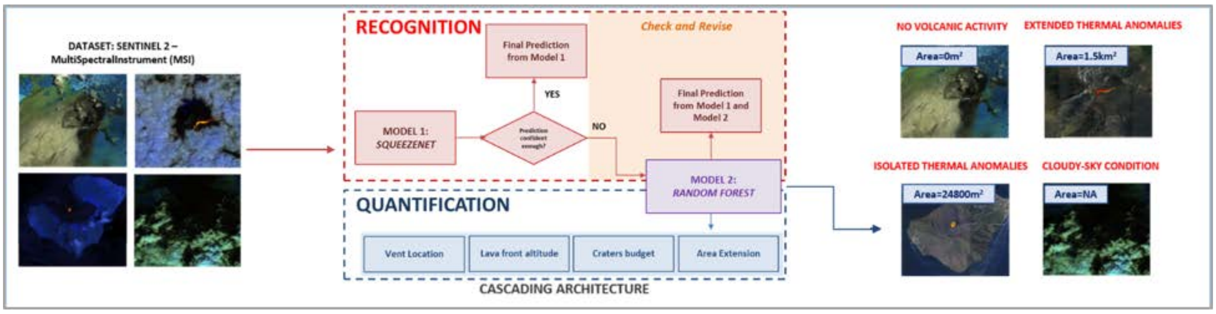

Figure 1. General scheme input-output cascading architecture.

After identifying scenes with isolated or extended thermal anomalies, an image segmentation technique (RF) is applied to quantify and further characterize the detected thermal features. This quantification allows for the analysis of patterns and trends in thermal anomalies, such as temporal temperature variations or distinct spatial distributions. By adopting this quantitative approach, we can gain deeper insights into volcanic activity dynamics, enabling the detection of subtle variations that could indicate potential eruptions.

Using Google Earth Engine (GEE), we outlined each crater region following the method described in Corradino et al. (2022), calculating the contribution of each crater to the overall detected thermal anomalies. This approach provides a detailed understanding of the spatial distribution of thermal features within each volcanic system, further enhancing our ability to monitor and predict volcanic activity.

The code for applying the cascading model can be obtained by sending an email to the author.

4. Volcano case studies

We present a detailed analysis of the results obtained by applying our cascading machine learning model to the monitoring of three highly active volcanoes: Etna, Stromboli, and Pacaya. These volcanoes, each characterized by distinct eruptive behaviors and geographical settings, serve as ideal case studies for evaluating the capabilities and robustness of the cascading model.

The primary objective of this case study analysis is to identify and map thermal anomalies detected in satellite imagery and to explore the underlying thermal dynamics of each volcanic system. Through this exploration, we aim to shed light on the unique volcanic activities at each site, providing valuable insights into the processes driving their behavior. This, in turn, contributes to a deeper understanding of how these complex systems evolve over time.

Mount Etna (Sicily, Italy): Etna, one of the most active volcanoes in the world, is well-known for its frequent and varied eruptive patterns, including effusive and explosive activity. Its complex, multi-vent structure presents significant challenges for real-time monitoring and analysis. Using the cascading model, we were able to detect both extended and isolated volcanic thermal anomalies, providing a detailed thermal map of the ongoing volcanic processes. The application of the model revealed important thermal signals associated with reactivations of secondary vents and lava flows, offering insights into the spatial distribution of volcanic activity.

Stromboli (Aeolian Islands, Italy): Stromboli is characterized by persistent explosive activity, often referred to as "Strombolian" eruptions, marked by continuous degassing and frequent, small-scale explosions. The cascading model was particularly effective in capturing these intermittent thermal anomalies, allowing us to track changes in the intensity and frequency of the explosive events. The model's ability to classify thermal patterns in near real-time has proven valuable for monitoring Stromboli's dynamic and often unpredictable eruptive behavior, offering a more nuanced understanding of its thermal anomalies compared to conventional methods.

Pacaya (Guatemala): Pacaya, located in Central America, is another highly active volcano, with frequent lava flows and explosive eruptions. Unlike Etna and Stromboli, Pacaya's volcanic system is situated in a tropical environment, where cloud cover and atmospheric conditions often interfere with satellite-based observations. Despite these challenges, the cascading model demonstrated its robustness by effectively identifying and quantifying thermal anomalies during both effusive and explosive phases. The inclusion of Pacaya in this study enabled us to test the adaptability of the model across different geographical and environmental contexts, confirming its capability to generalize across diverse volcanic systems.

5. Results and discussion

Through these case studies, we validate the cascading model's ability to capture the unique thermal signatures of each volcano, offering valuable insights into their eruptive dynamics. By identifying thermal changes associated with volcanic unrest, our model not only enhances our understanding of these active systems but also provides a powerful tool for future volcanic hazard monitoring and risk mitigation efforts.

5.1 Mount Etna

Mount Etna has exhibited persistent degassing and frequent episodes of both explosive and effusive eruptions from its four summit craters over the past two decades (Calvari et al., 2022; Calvari and Nunnari, 2024). These craters, South-East (SE), North-East (NE), Bocca Nuova (BN), and Voragine (VOR), have contributed variably to Etna's overall volcanic activity. Accurate identification of active craters at any given time and quantification of their thermal emissions are crucial for understanding the dynamics within the volcanic system. In particular, the areal coverage of thermal anomalies provides valuable insights into how each crater's activity contributes to the overall behavior of the volcano.

We use the RF algorithm (Corradino et al., 2022) with the aim of evaluating the contribution of Mount Etna's four summit craters in terms of the areal extent of the thermal anomaly. To achieve this, a Region of Interest (ROI) was defined around each crater, the algorithm was applied, and the areal extent of the thermal anomaly associated with each crater was calculated. Figure 2 illustrates the temporal evolution of thermal activity at Etna's craters. The blue line represents the Cumulative Area, denoting the total areal coverage of thermal anomalies from all four craters. The red line, labeled BN-VOR Cumulative Area, isolates the thermal contribution from the Bocca Nuova and Voragine craters. This differentiation is essential for understanding the interaction between craters, especially during periods of heightened activity. Figure 2 shows the contribution of the two craters (BN and VOR) to the observed overall thermal activity, for instance when only BN and VOR are active, the blue and red lines overlap. In addition, Fig. 2 displays the results of our cascading model applied to satellite scenes, effectively filtering out samples classified as non-volcanic activity (NVA) or cloud/snow cover (CSC). Scenes identified as either Isolated Volcanic Thermal Anomalies (ITA) or Extended Volcanic Thermal Anomalies (ETA) are marked by gray dots, allowing us to pinpoint periods of significant volcanic thermal activity.

As illustrated in Fig. 2a, between September 12, 2019, and April 30, 2020, Voragine displayed heightened activity, with effusive eruptions originating from a vent on the eastern side of the intracrater cone. This activity resulted in the formation of a lava flow that advanced toward Bocca Nuova, crossing the saddle between the two craters. As shown in Figure 2a, the combined thermal output from BN and VOR during this period contributed significantly to the BN-VOR Cumulative Area, indicating their prominent role in Etna's eruptive dynamics.

Starting on December 18, 2020, Etna entered a phase of intense eruptive activity, characterized by 66 paroxysmal eruptions that produced numerous lava flows (Amato, 2022). Figure 2b reveals that prior to the first paroxysm, all four summit craters exhibited marked increases in thermal anomaly coverage. The NE and SE craters showed escalating thermal activity as early as June 2020, while BN and VOR began displaying a similar upward trend in November 2020, signaling the imminent paroxysmal sequence. These thermal anomalies acted as early indicators of volcanic unrest, emphasizing the value of continuous thermal monitoring for forecasting eruptive events.

On November 12, 2023, another significant paroxysm occurred. In the lead-up to this event, cascading model outcomes from October 14 to November 5, 2023 (Fig. 3), revealed persistent thermal anomalies at the southeast crater starting on October 24, 2023. These anomalies were detectable until November 5, 2023, the date of the last cloud-free satellite acquisition, just before the lava fountain of November 12, 2023. This paroxysmal event triggered several pyroclastic flows and continued with the emplacement of lava flows in multiple directions.

The cascading machine learning model, applied to Sentinel 2-MSI imagery, successfully classified the scene from November 12, 2023, as Extended Volcanic Thermal Anomalies (ETA), providing an estimated areal extension of 1.15 km^2. Additionally, the model identified the lowest altitude of the southwestern lava branch at 2430 meters above sea level (m a.s.l.). Figure 4 presents a false-color composite image of the lava field, created using the NIR, SWIR1, and SWIR2 bands from Sentinel 2-MSI. This image highlights the spatial distribution of the lava flows, offering a comprehensive view of the eruptive features.

The cascading model's ability to accurately detect, classify, and quantify thermal anomalies on Mount Etna has proven invaluable for monitoring volcanic activity in near real-time. When referring to the concept of Near Real-Time, we specifically mean the ability to obtain data and information as soon as the Sentinel-2 satellite image

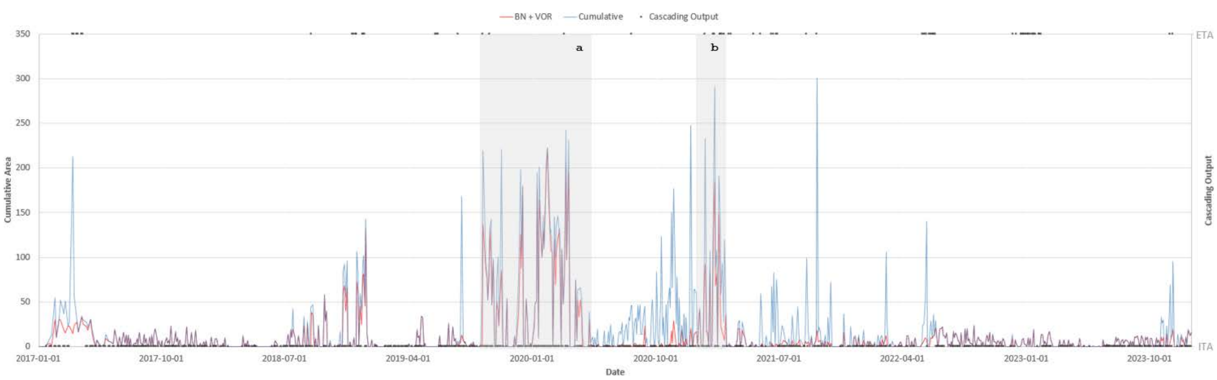

Figure 2. Thermal activity of the craters (North-East (NE), South-East (SE), Bocca Nuova (BN), and Voragine (VOR)), calculated through Random Forest Algorithm, from January 2017 to January 2023. The cumulative area of all four craters is represented by the blue line, while the contribution specifically related to BN and VOR is shown by the red line.

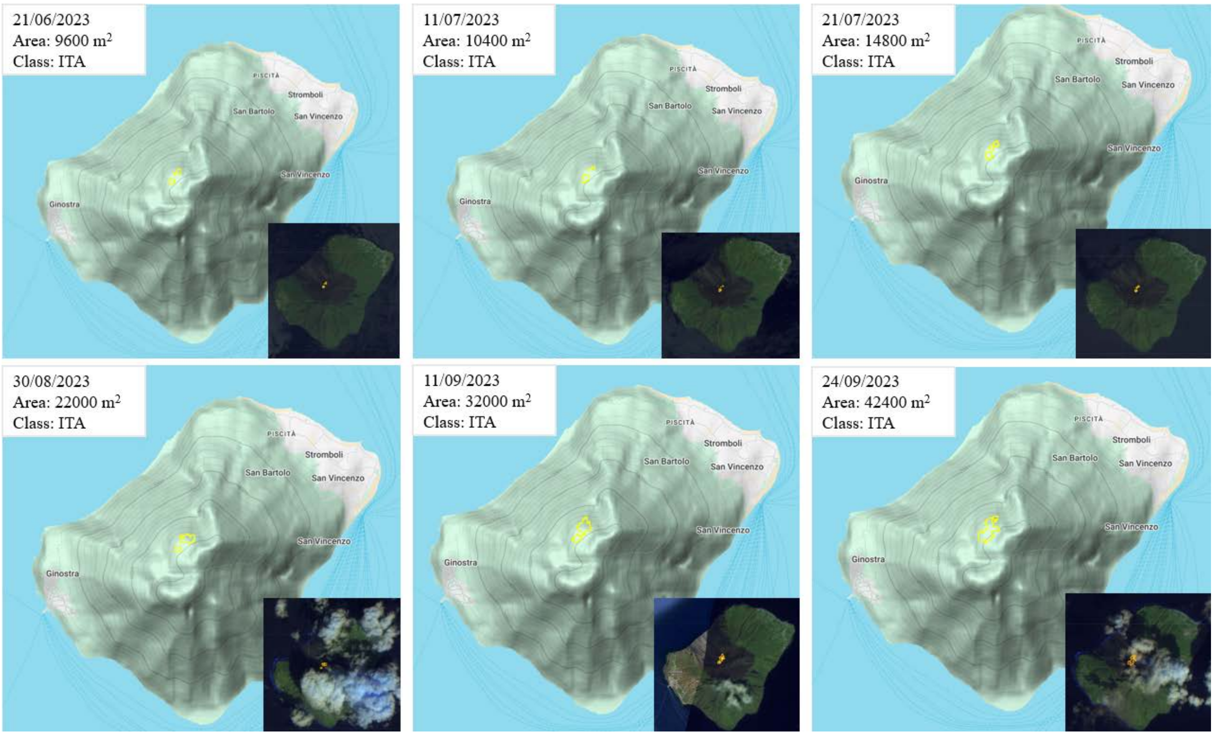

Figure 3. Volcanic thermal activity detected by the cascading model using S2-MSI data on Etna before the paroxysm on 12th November 2023. Reference system used WGS84.

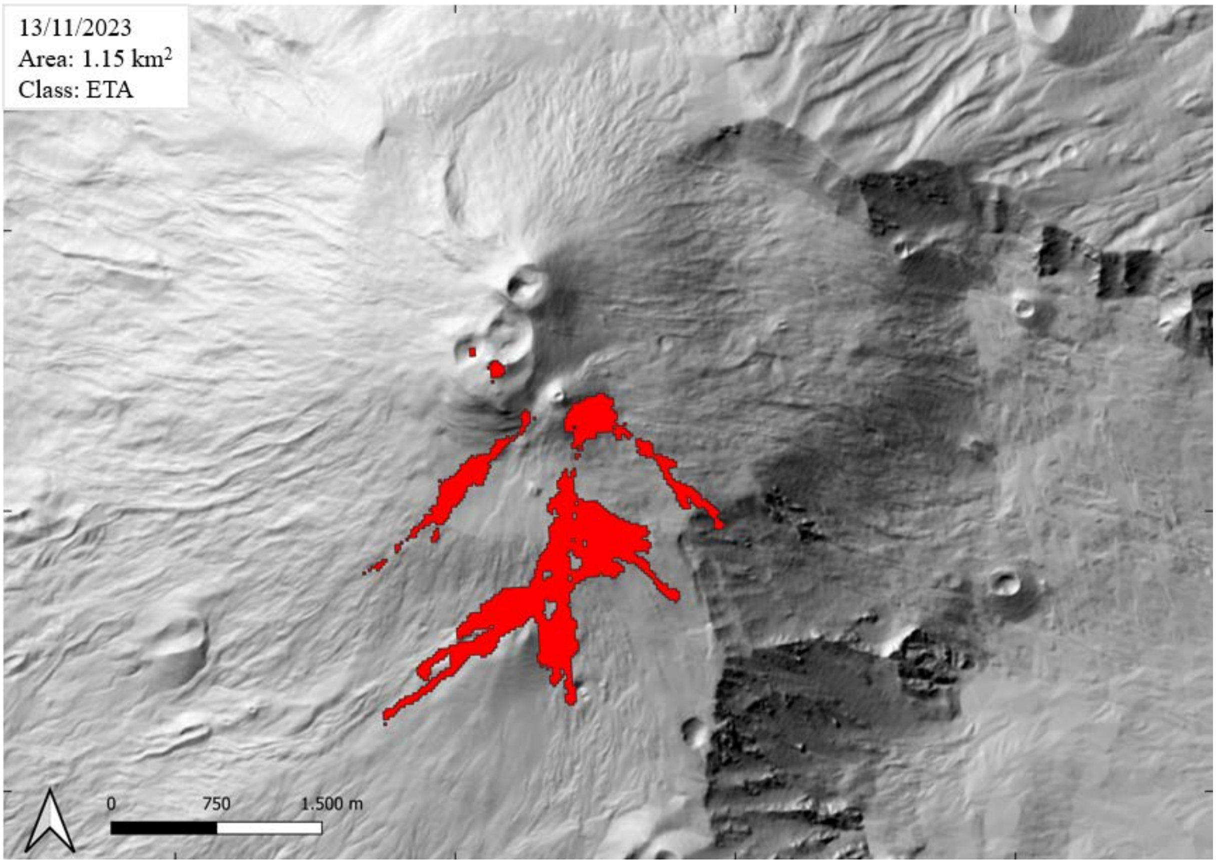

Figure 4. The lava field was mapped using the bands NIR, SWIR1, and SWIR2 from the image Sentinel 2-MSI 13/11/2023 09:53 UTC (resolution 20 m) and the trained cascading model ML. Its extension is 1.15 km^2 and the branch to the southwest reaches the lowest altitude of 2430 m a.s.l. Digital Elevation Model (DEM) (https://tinitaly.pi.ingv.it/). Reference system used WGS84.

becomes available. Once the image is accessible, the cascading model is promptly applied, requiring approximately 3 minutes for the download, processing, model application, and visualization of the results. By continuously tracking changes in thermal emissions across multiple craters, the model not only enhances our understanding of Etna's complex eruptive behavior but also provides critical data for improving hazard assessment and mitigation strategies.

5.2 Stromboli

Stromboli is renowned for its persistent explosive activity, which can occasionally escalate into more intense events, including major explosions or paroxysmal eruptions (Calvari et al., 2022; Giudicepietro et al., 2022). Such episodes may lead to the effusion of lava flows, marking significant phases of volcanic unrest. One such intense phase of activity began on September 25, 2023, with a major explosive event. This powerful explosion marked the onset of heightened volcanic activity, which culminated in a new eruptive phase starting on September 27, 2023, characterized by multiple lava overflow episodes.

This eruptive activity was preceded by a notable increase in thermal emissions, beginning in late June 2023 (Fig. 5). This rise in thermal anomaly coverage likely reflects the presence of magma within Stromboli's shallow plumbing system, as well as the accumulation of pyroclastic material. The cascading model was instrumental in monitoring and classifying this thermal activity, allowing for a timely and accurate assessment of evolving volcanic behavior. Fig. 6 shows the increase in thermal anomalies, in terms of areal extent, in the days leading up to and immediately following the explosion on September 27th, 2023.

The image available on September 27 allowed us to detect the first lava flow of the series, for a total of three events. On October 3, 6, 7, and 8, four lava overflows were emitted from sector N1 of the N crater area (ref: INGV bulletin 41/2023). On September 27, 2023, the cascading algorithm identified the first lava flow of this phase, classifying it as an Extended Volcanic Thermal Anomaly (ETA). The hot area was estimated at approximately 36,000 m^2, with the lava front positioned at an altitude of 300 meters above sea level. The overflow on October 3, which began in the early morning hours, was detected by our model the following day, on October 4 (during the cooling phase), as the data was not available on the same day. The hot area on this date was reduced to about 16,000 m^2, with the lava front advancing to 400 m a.s.l, classified as an Isolated Volcanic Thermal Anomaly (ITA). On October 6, 2023, the activity intensified again, the model detected the lava overflow on the same day as the event, and the algorithm classified the event as ETA, identifying a hot area of approximately 34,000 m^2, with the lava front at an altitude of 600 m a.s.l. Fig. 7 shows the three previously described overflow episodes, classified by the model.

The cascading model's ability to track the evolution of Stromboli's eruptive activity, particularly in relation to lava flows and thermal emissions, provides critical insights into the behavior of this highly active volcano. By monitoring thermal anomalies in near real-time, the model aids in the early detection of potential hazards, enhancing both scientific understanding and risk mitigation efforts for volcanic activity on Stromboli.

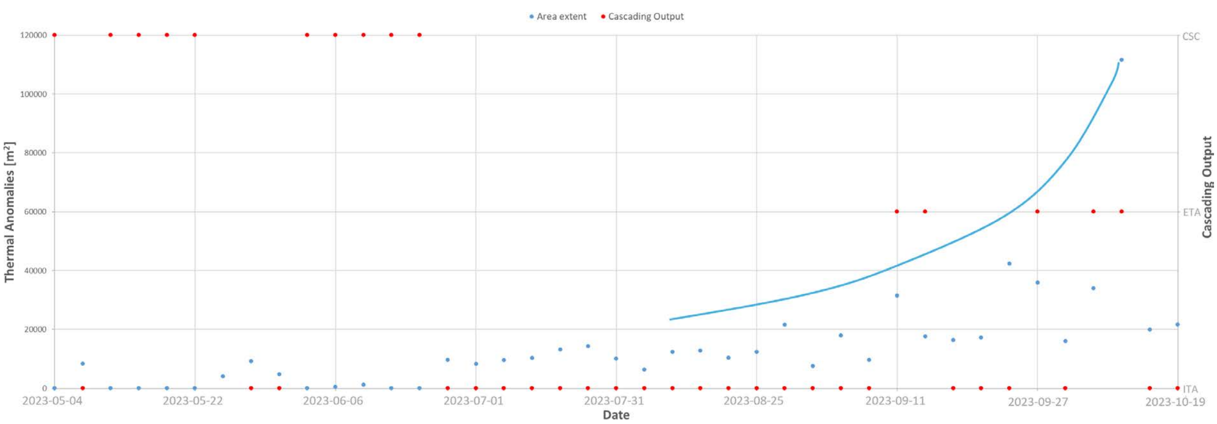

Figure 5. Volcanic thermal activity detected by the cascading model using S2-MSI data on Stromboli prior to the onset of the activity phase on September 27th, 2023. Red dots indicate the cascading model's output, the curved blue line represents the increasing trend of thermal anomalies over time, and blue dots show the thermal anomalies area extensions.

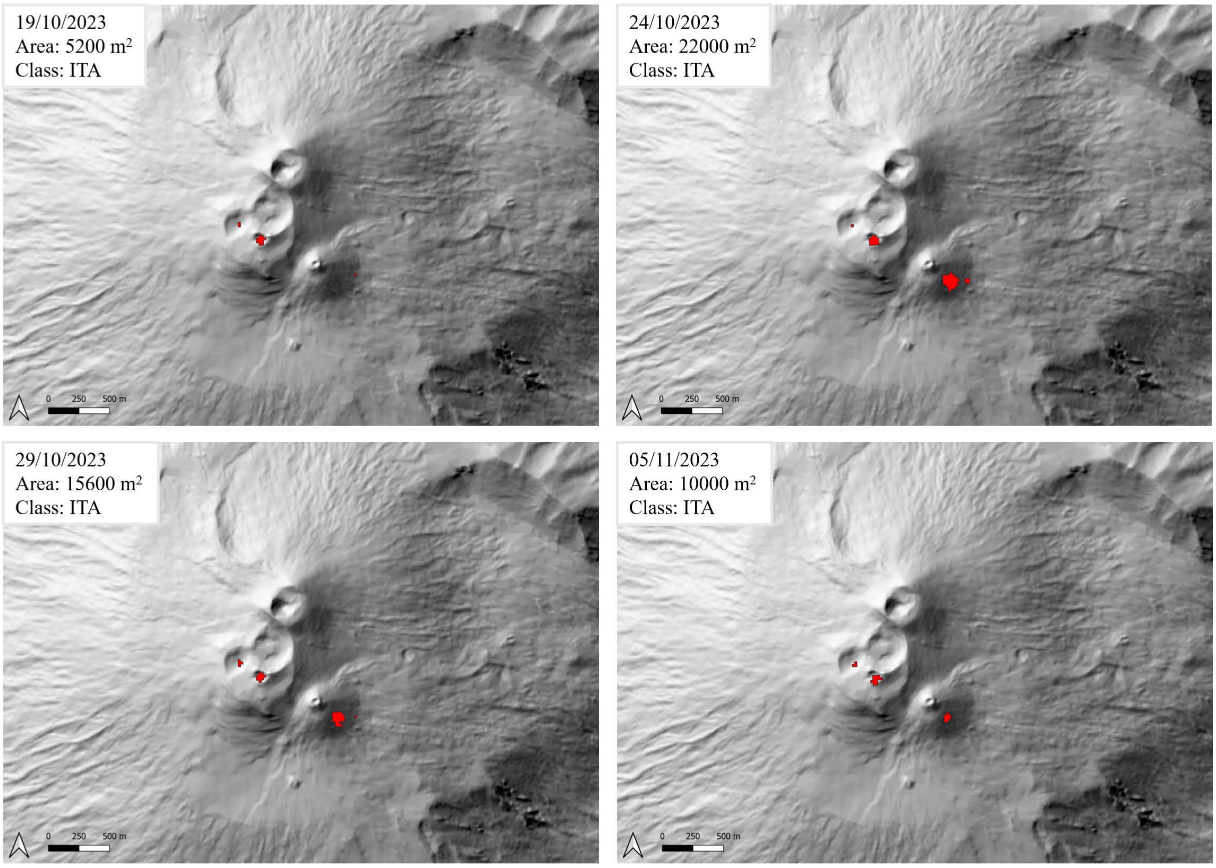

Figure 6. Increasing thermal activity on Stromboli before major explosion on September 27th 2023. Reference system used WGS84.

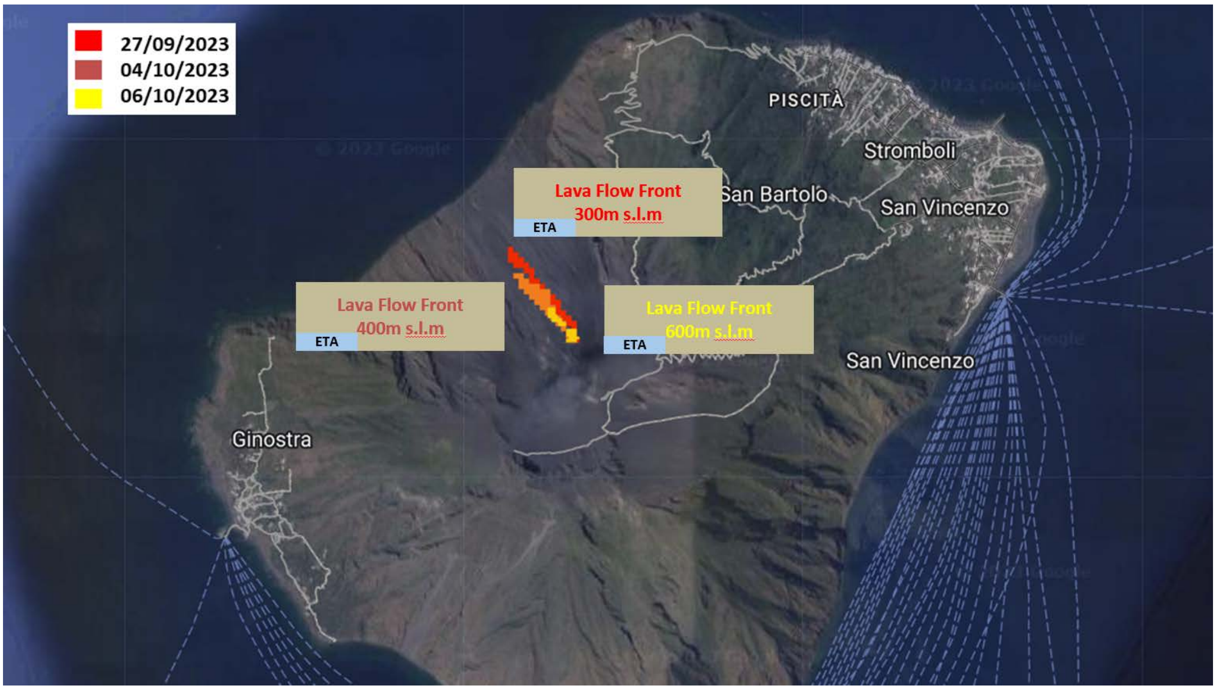

Figure 7. Overflow episodes, classified by the model, occurred on September 27, 2023, October 4, 2023, and October 6, 2023. Reference system used WGS84.

5.3 Pacaya volcano

Pacaya, a highly active stratovolcano in Guatemala, Central America, is renowned for its frequent eruptions and impressive lava displays. While it typically exhibits Strombolian activity, Pacaya occasionally experiences more violent Plinian eruptions, making it a critical focus for volcanic monitoring in the region (Lechner and Rouleau, 2019; Rose et al., 2013). The volcano entered a heightened phase of activity in early 2021, characterized by both effusive (lava flows) and explosive eruptions. This phase had been building up since mid-2015 when the volcano began showing signs of increasing unrest, marked by intensified lava effusion and ash production. The growing volcanic hazard prompted evacuations in affected areas to protect the local population (Lechner and Rouleau, 2019; and Bulletin INSIVUMEH).

The first indicators of heightened activity were detected in February 2015, with an uptick in seismicity, ash plume emissions, and increased degassing (Bulletin INSIVUMEH). These signs of volcanic unrest persisted throughout 2016, as Pacaya exhibited intermittent Strombolian eruptions, accompanied by ongoing lava flows and ash emissions, signaling the persistence of elevated volcanic activity.

The cascading model was applied to the 2017-2022 period to assess its effectiveness in monitoring and evaluating Pacaya's activity (Fig. 8). Utilizing advanced machine learning techniques, the cascading pipeline effectively classified satellite images of Pacaya's thermal behavior. The SN model (Cariello et al., 2024) within the cascading framework was particularly adapted at filtering out environmental noise, such as cloud cover and solar radiance on local slopes in the daytime images, ensuring accurate detection of volcanic activity. The RF model (Corradino et al., 2022) further reconstructed a continuous time series of thermal anomaly maps, providing a comprehensive view of the volcano's thermal dynamics.

A clear upward trend in volcanic activity is evident from August 2016 onward, coinciding with the availability of Sentinel-2 Harmonized data, which aligns well with field observations (Gonzalez-Santana et al., 2024; Bulletin INSIVUMEH, Global Volcanism Program, 2017).

In February 2017, increased activity led to the formation of lava flows extending approximately 300 meters along Pacaya's northwest flank. The cascading model identifies approximately 55% of the entire dataset as cloudy scenes, thus the system successfully filtered out much of this environmental interference. However, due to the frequent and dense cloud cover, thermal anomalies may sometimes be obscured.

From 2018 to 2020, Pacaya's volcanic activity remained sustained, with multiple lava flows, persistent Strombolian eruptions, and continuous gas emissions. Figure 9 presents a series of Sentinel-2 Harmonized images alongside the anomaly maps generated by the cascading algorithm, depicting these key phases of activity. Notably, there was a brief pause in effusive activity during July and August 2020, followed by a renewed and intensified eruptive phase in early 2021.

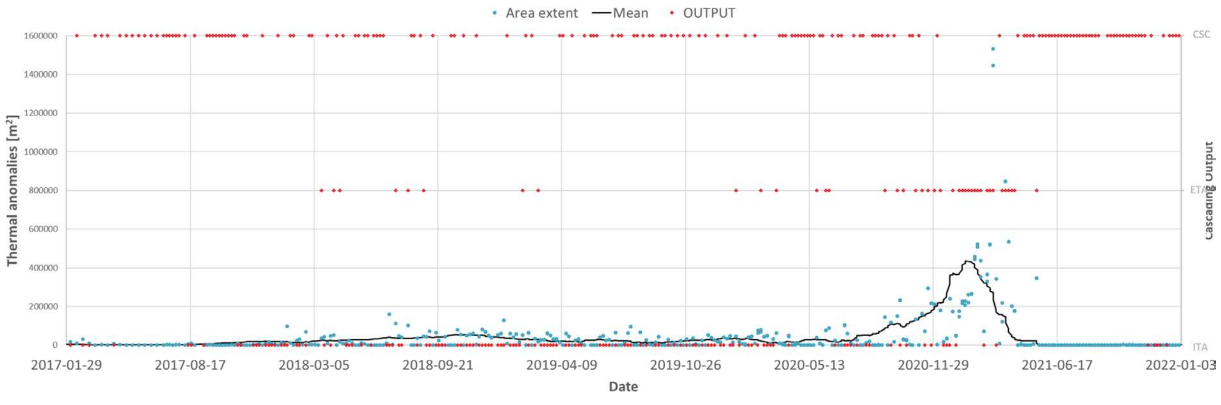

Figure 8. Temporal analysis of Pacaya's volcanic activity from August 2016 to the present, utilizing the cascading model for thermal anomaly classification. The orange line represents the mean thermal anomaly over time, highlighting the overall increasing trend in volcanic activity. The red output indicates the classifications derived from the cascading model: Isolated Volcanic Thermal Anomalies (ITA), Extended Volcanic Thermal Anomalies (ETA), and Cloudy-Sky Condition (CSC). Non-volcanic activity (NVA) has been filtered out to enhance visualization, allowing for a clearer interpretation of Pacaya's eruptive behavior and thermal dynamics throughout this period.

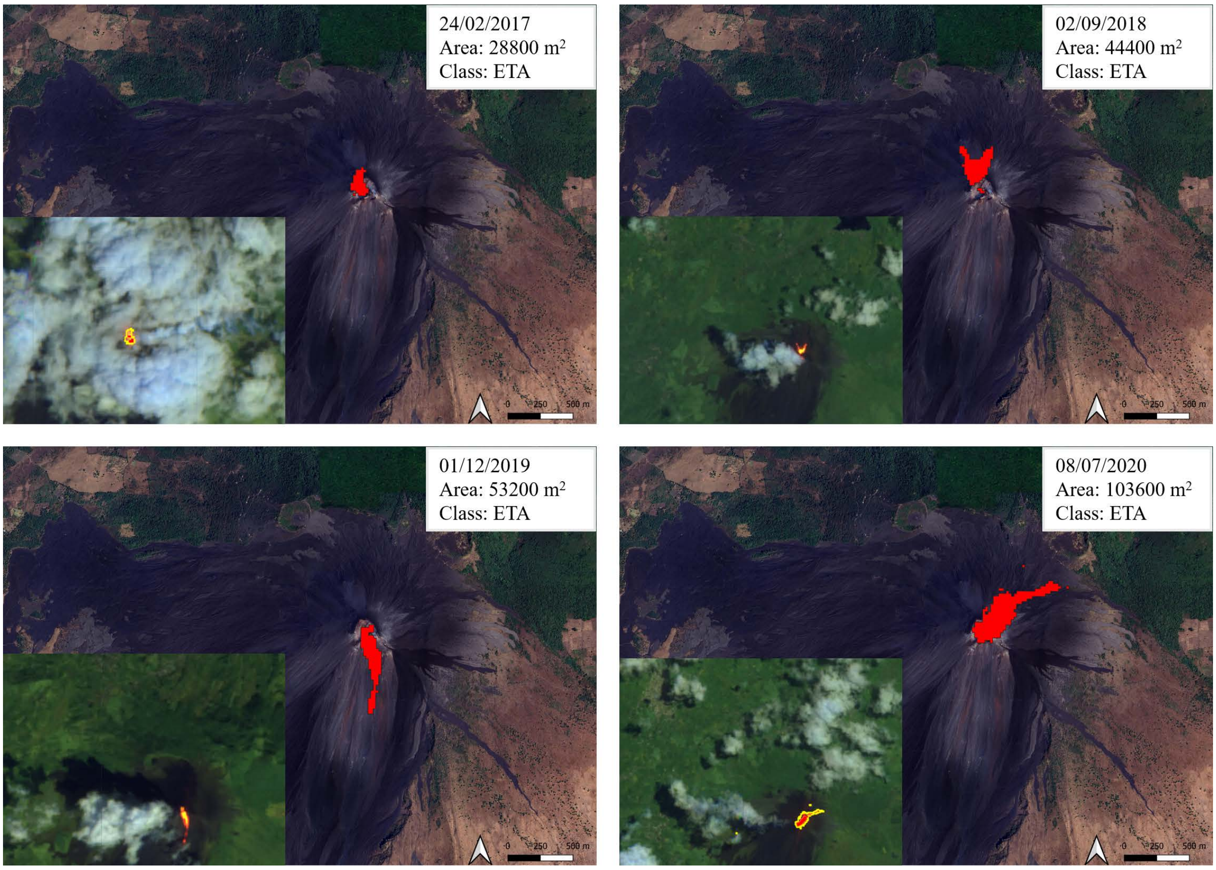

Figure 9. Lava flow episodes occur on February 24, 2017, September 2, 2018, December 1, 2019, and July 8, 2020. Reference system used WGS84.

The early months of 2021 marked some of the most intense volcanic activity observed at Pacaya. Frequent explosions and significant lava emissions resulted in lava flows that extended into surrounding areas, necessitating road closures and evacuations to ensure public safety. Authorities issued warnings to nearby communities, highlighting the growing volcanic hazard (Bulletin INSIVUMEH).

Figure 10 presents a sequence of Sentinel-2 images highlighting significant increases in activity as identified by the cascading model. Some images exhibit disturbing effects (i.e. clouds coverage, diffraction spikes, blurring and thermal halo (Massimetti et al., 2020). These optical phenomena occur due to the interaction of light with atmospheric particles or the sensor's optical components, causing light to scatter and produce halo effects or refract, leading to image distortions. Such effects can artificially enlarge or intensify the appearance of thermal anomalies, potentially misrepresenting the actual extent of volcanic activity. Massimetti et al. (2020) discusses how halo effects and cloud refraction can affect thermal detection and outlines methods to effectively isolate genuine thermal anomalies from such interferences.

The effusive activity reached its peak in March 2021, with three lava flows, each approximately 1 km in length, advancing along multiple branching paths on different flanks of the volcano. This heightened activity is corroborated by MIROVA radiative power data (Coppola et al., 2020), indicating a substantial increase in intensity through mid-April. Another spike in activity occurred in late April and early May, characterized by additional explosions and lava flows. However, by the third week of May, thermal activity began to decline significantly, suggesting a shift in Pacaya's eruptive behavior.

Our work directly addresses some of the future challenges highlighted by (Coppola et al., 2020), particularly regarding cloud filtering with the MIROVA system. The cascading model we propose incorporates a robust cloud classification step that effectively isolates cloudy scenes, identifying approximately 55% of the dataset as cloud-covered for Pacaya volcano. This is especially relevant considering the complex challenges posed by tropical volcanoes, which often experience frequent and dense cloud cover. Our model not only detects cloud-affected scenes, regardless of cloud density, but also provides a probabilistic assessment for each prediction. This feature enables the

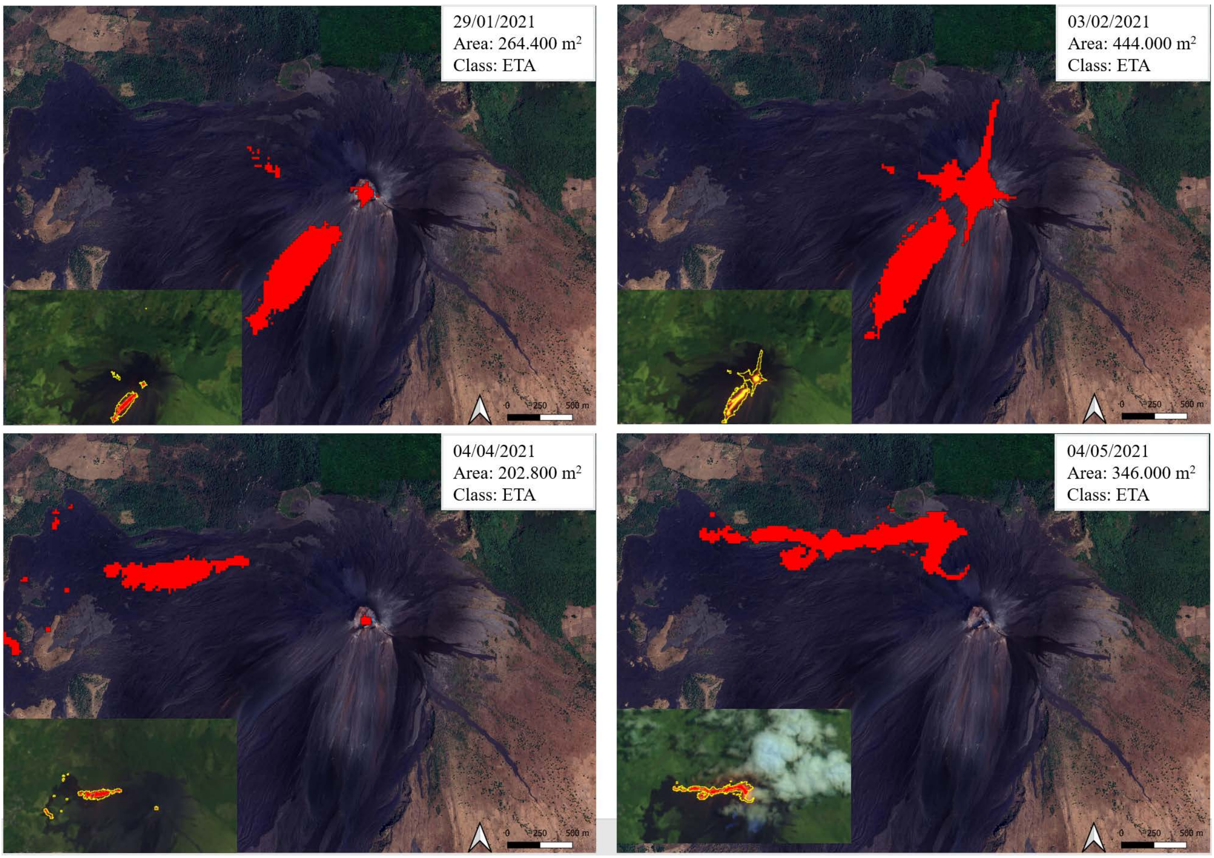

Figure 10. A sequence of images corresponding to the identified points of increase in Fig. 8. This series visually illustrates the key moments of heightened volcanic activity, capturing the dynamics of the eruptions and the development of thermal anomalies over time. Reference system used WGS84.

investigation of potential thermal anomalies even in images classified as cloudy. In fact, even when anomalies are partially obscured by clouds, our model maintains the ability to identify and quantify them. Additionally, leveraging high spatial resolution data from the MSI sensor on Sentinel-2 enhances our capability to distinguish genuine thermal signals from artifacts caused by cloud presence or other atmospheric factors. This advanced approach ensures a more reliable assessment of volcanic activity, particularly in environments where traditional monitoring methods might struggle due to adverse atmospheric conditions. The integration of cloud filtering and anomaly detection within our workflow represents a significant step forward in addressing the operational challenges of monitoring tropical volcanoes like Pacaya.

6. Conclusions

This study highlights the significant potential of leveraging an accurate, automated thermal anomaly detection model for Near-Real time volcano monitoring. Through the detailed analysis of volcanic activity at Etna, Stromboli, and Pacaya, we have uncovered previously unseen patterns, deepening our understanding of the complex processes that govern these dynamic systems. Thanks to the increased accuracy, the model is able to accurately detect thermal changes otherwise unseen by previous approaches. As a result, thermal trends associated with eruptive dynamics become apparent and can now be observed directly by experts, rather than solely by the model. This approach not only reveals new features but also enhances our ability to identify hidden magma dynamics, thereby improving forecasting capabilities and informing volcanic hazard mitigation strategies.

Detecting subtle growth trends in thermal data presents challenges, particularly in identifying small, incremental abnormalities that may indicate an impending eruption. The combined use of satellite thermal data with advanced AI detection techniques is indispensable for pinpointing these critical changes. The cascading model processes each new Sentinel-2 image within 3 minutes, allowing near real-time application as soon as the image becomes available (approximately 2 hours after acquisition). Since the time scale of the pre-eruptive trends that we show are far larger than the acquisition/processing times (days versus hours), this approach can represent a valuable mean for earlier warnings and more effective volcano monitoring, especially for highly active volcanoes such as Etna, Stromboli, and Pacaya.

The study effectively demonstrates how AI techniques enhance anomaly detection and quantification, highlighting the advantages of model adaptability and the ability to learn complex data patterns. To further support this argument, a quantitative comparison with classical techniques, particularly fixed-threshold methods, was conducted. The analysis was performed using data from the MIROVA system, which uses satellite data to monitor volcanic thermal activity, providing an ideal context for evaluating different methodologies. The reference dataset, obtained from the MIROVA website (www.mirovaweb.it), corresponds to the activity of the Pacaya volcano and is based on the algorithm described by Massimetti et al. (2020), which relies on images acquired by the Sentinel-2 MSI sensor. A comparison of the areal extent of thermal anomalies detected by both algorithms revealed a strong correlation of 96% between the datasets. This correlation was computed using MATLAB's corr function, which calculates the Pearson correlation coefficient to measure the linear association between two datasets. A coefficient of 96% indicates a very strong positive correlation, suggesting that both algorithms yield highly consistent results in detecting thermal anomalies. Additionally, the Mean Absolute Error (MAE) was computed in MATLAB to further quantify the differences between the two datasets. The MAE is calculated as the average of the absolute differences between corresponding values in the datasets, providing a measure of the overall discrepancy in terms of absolute deviations. In this case, the MAE was found to be 13 m^2. Considering that each pixel in Sentinel-2 MSI imagery corresponds to an area of 20 m × 20 m (400 m^2 per pixel), this implies that the average difference between the two datasets is approximately 0.03 pixels, confirming that the discrepancy between the algorithms is minimal. This metric serves as an estimate of the "error" or deviation between the datasets, reinforcing the reliability of the detection algorithms.

However, challenges persist. The quality and completeness of input data are crucial, highlighting the importance of meticulous data collection and management practices. Additionally, while our approach offers valuable insights, the inherent complexity of volcanic phenomena may limit the interpretability and explainability of detected patterns. To address this, a multidisciplinary approach that integrates diverse scientific perspectives and methodologies is essential for enriching our understanding and refining our interpretations. The ability to collect and analyze satellite data in near real-time (NRT) is crucial for monitoring volcanic hazards and determining the appropriate mitigation actions. Leveraging multiple satellite sensors, both geostationary and polar, is undoubtedly the next critical step forward. In (Cariello et al., 2024), an AI-based platform was developed for NRT monitoring of volcanic activity, combining satellite imagery with offline analysis. It enables informed, timely, and advanced management of volcanic activity, playing a vital role in enhancing safety measures and reducing the impact of volcanic events on communities and infrastructure. This holistic perspective is key to advancing volcanic monitoring and improving hazard assessments, ultimately contributing to more effective disaster preparedness and response strategies.

Acknowledgements. This work was developed within the framework of the Laboratory of Technologies for Volcanology (TechnoLab) at the Istituto Nazionale di Geofisica e Vulcanologia (INGV) in Catania (Italy). This research was funded by the ATHOS Research Program (INGV OB.FU. 0867.010), by the 2019 Strategic Project FIRST – Forecasting eRuptive activity at Stromboli volcano: timing, eruptive style, size, intensity, and duration – of the INGV Volcanoes Department (Delibera n. 144/2020), and by Project INGV Pianeta Dinamico VT_ORME 2023-2025 (INGV OB.FU. 1020.010). We are grateful to European Space Agency (ESA) for satellite data. Note: Google CoLaboratory™ is a trademark of Google LLC – ©2018 Google LLC All rights reserved.

References

Abrams, M., E. Abbott, A. Kahle et al. (1991). Combined use of visible, reflected infrared, and thermal infrared images for mapping Hawaiian lava flows, J. Geophys. Res. Solid Earth, 96, B1, 475-484, doi:10.1029/90JB01392.

Amato, E. (2022). Machine learning and best fit approach to map lava flows from space, Il Nuovo Cimento C, 45, 4, 1-12, doi:10.1393/ncc/i2022-22080-1.

Amato, E., C. Corradino, F. Torrisi, C. Del Negro et al. (2023). A Deep Convolutional Neural Network for Detecting Volcanic Thermal Anomalies from Satellite Images, Remote Sens., 15, 15, 3718, doi:10.3390/rs15153718.

Anantrasirichai, N., J. Biggs, F. Albino, P. Hill et al. (2018). Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data, J. Geophys. Res. Solid Earth, 123, 8, 6592-6606, doi:10.1029/2018JB015911.

Aziz, A., A. Indahwati and B. Sartono (2019). Improving prediction accuracy of classification model using cascading ensemble classifiers, IOP Conf. Ser. Earth Environ. Sci., 299, 1, 012025, doi:10.1088/1755-1315/299/1/012025.

Bonaccorso, G. (2017). Machine Learning Algorithms, Packt Publishing Ltd, Birmingham, UK, 337, ISBN:978-1-78588-962-2.

Calvari, S., F. Di Traglia, G. Ganci, V. Bruno et al. (2022). Multi-parametric study of an eruptive phase comprising unrest, major explosions, crater failure, pyroclastic density currents and lava flows: Stromboli volcano, 1 December 2020-30 June 2021, Front. Earth Sci., 10, doi:10.3389/feart.2022.899635.

Calvari, S. and G. Nunnari (2024). Reawakening of Voragine, the Oldest of Etna's Summit Craters: Insights from a Recurrent Episodic Eruptive Behavior, Remote Sens., 16, 22, 4278, doi:10.3390/rs16224278.

Cariello, S., C. Corradino, F. Torrisi, C. Del Negro et al. (2024). Cascading Machine Learning to Monitor Volcanic Thermal Activity Using Orbital Infrared Data: From Detection to Quantitative Evaluation, Remote Sens., 16, 1, 171, doi:10.3390/rs16010171.

Coppola, D., M. Laiolo, C. Cigolini, F. Massimetti et al. (2020). Thermal Remote Sensing for Global Volcano Monitoring: Experiences From the MIROVA System, Front. Earth Sci., 7, doi:10.3389/feart.2019.00362.

Corradino, C., E. Amato, F. Torrisi, C. Del Negro et al. (2022). Data-Driven Random Forest Models for Detecting Volcanic Hot Spots in Sentinel-2 MSI Images, Remote Sens., 14, 17, 4370, doi:10.3390/rs14174370.

Corradino, C., G. Ganci, A. Cappello, G. Bilotta et al. (2020). Recognizing Eruptions of Mount Etna through Machine Learning Using Multiperspective Infrared Images, Remote Sens., 12, 6, 970, doi:10.3390/rs12060970.

Crafford, A. E. and E. Venzke (2017). Global Volcanism Program. Report on Pacaya (Guatemala), Bull. Glob. Volcanism. Netw., 42, 5, Smithsonian Institution, doi:10.5479/si.GVP.BGVN201705-342110.

Eesa, A. S. and W. K. Arabo (2017). A Normalization Methods for Backpropagation: A Comparative Study, Sci. J. Univ. Zakho, 5, 4, 319-323, doi:10.25271/2017.5.4.381.

Fabio, L., D. Piga, U. Michelucci and E. G. Safouane (2024). BenchCloudVision: A Benchmark Analysis of Deep Learning Approaches for Cloud Detection and Segmentation in Remote Sensing Imagery, arXiv, doi:10.48550/arXiv.2402.13918.

Gama, J. and P. Brazdil (2000). Cascade Generalization, Mach. Learn., 41, 3, 315-343, doi:10.1023/A:1007652114878.

Gargiulo, M., D. A. G. Dell'Aglio, A. Iodice, D. Riccio et al. (2019). A CNN-Based Super-Resolution Technique for Active Fire Detection on Sentinel-2 Data, arXiv, doi:10.48550/arXiv.1906.10413.

Giudicepietro, F., S. Calvari, L. D'Auria, F. D. Traglia et al. (2022). Variations of Stromboli activity related to the 2019 paroxysmal phase revealed by SOM clustering of seismo-acoustic data and its comparison with video recordings and GBInSAR measurements, EGU22, Fasc. EGU22-10482, doi:10.5194/egusphere-egu22-10482.

Gonzalez-Santana, J., C. Wauthier and G. Waite (2024). The 2021 eruption of Pacaya Volcano, Guatemala – Geophysical analysis through satellite geodesy and seismic noise correlations, J. Volcanol. Geotherm. Res., 447, 108027, doi:10.1016/j.jvolgeores.2024.108027.

Goodfellow, I., Y. Bengio and A. Courville (2016). Deep Learning, MIT Press, 781.

Higgins, J. and A. Harris (1997). VAST: A program to locate and analyse volcanic thermal anomalies automatically from remotely sensed data, Comput. Geosci., 23, 6, 627-645, doi:10.1016/S0098-3004(97)00039-3.

Lary, D. J., A. H. Alavi, A. H. Gandomi, A. L. Walker et al. (2016). Machine learning in geosciences and remote sensing, Geosci. Frontiers, 7, 1, 3-10, doi:10.1016/j.gsf.2015.07.003.

Lechner, H. N. and M. D. Rouleau (2019). Should we stay or should we go now? Factors affecting evacuation decisions at Pacaya volcano, Guatemala, Int. J. Disaster Risk Reduct., 40, 101160, doi:10.1016/j.ijdrr.2019.101160.

Mahataa, K., A. Sarkarb, R. Dasb and S. Dasb (2019). Land Use Land Cover Classification by applying Ensemble Approaches on Multi Temporal Remote Sensing Images, Int. J. Manag. Sci. Eng. Manag, 9, 3, 655-660.

Massimetti, F., D. Coppola, M. Laiolo, S. Valade et al. (2020). Volcanic Hot-Spot Detection Using SENTINEL-2: A Comparison with MODIS-MIROVA Thermal Data Series, Remote Sens., 12, 5, 820, doi:10.3390/rs12050820.

Patrick, M. R., J. Kauahikaua, T. Orr, A. Davies et al. (2016). Operational thermal remote sensing and lava flow monitoring at the Hawaiian Volcano Observatory, Geol. Soc. London Spec. Publ., 426, 1, 489-503, doi:10.1144/SP426.17.

Ray, S. and A. De Sarkar (2012). Execution analysis of load balancing algorithms in cloud computing environment, Int. J. Cloud Comput. Serv. Archit., 2, 5, 1-13.

Rose, W. I., J. L. Palma, R. Escobar Wolf, R. O. Matías Gomez et al. (2013). A 50 yr eruption of a basaltic composite cone: Pacaya, Guatemala, in W. I. Rose, J. L. Palma, H. D. Granados and N. Varley (Eds.), Understanding Open-Vent Volcanism and Related Hazards, Geological Society of America, doi:10.1130/2013.2498(01).

Sether, A. (2016). Cloud Computing Benefits, SSRN Scholarly Paper Fasc., 2781593, doi:10.2139/ssrn.2781593.

Torrisi, F., C. Corradino, S. Cariello, C. Del Negro et al. (2024). Enhancing detection of volcanic ash clouds from space with convolutional neural networks, J. Volcanol. Geotherm. Res., 448, 108046, doi:10.1016/j.jvolgeores.2024.108046.

Tramutoli, V., C. Filizzola, N. Genzano, M. Lisi et al. (2018). Robust Satellite Techniques for Detecting Preseismic Thermal Anomalies, in Pre-Earthquake Processes, 241-258, American Geophysical Union, AGU, doi:10.1002/9781119156949.ch14.

Wang, G., S. Han, E. Ding, D. Huang et al. (2021). Student-Teacher Feature Pyramid Matching for Anomaly Detection, arXiv, doi:10.48550/arXiv.2103.04257.

***CORRESPONDING AUTHOR: Simona CARIELLO**,
University of Catania, Department of Electrical, Electronic and Computer Science Engineering, Catania, Italy
Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo, Catania, Italy
e-mail: simona.cariello@ingv.it