QuakeSet: A Dataset and Low-Resource Models to Monitor Earthquakes through Sentinel-1

Daniele Rege Cambrin*
Politecnico di Torino
daniele.regecambrin@polito.it

Paolo Garza
Politecnico di Torino
paolo.garza@polito.it

ABSTRACT

Earthquake monitoring is necessary to promptly identify the affected areas, the severity of the events, and, finally, to estimate damages and plan the actions needed for the restoration process. The use of seismic stations to monitor the strength and origin of earthquakes is limited when dealing with remote areas (we cannot have global capillary coverage). Identification and analysis of all affected areas is mandatory to support areas not monitored by traditional stations. Using social media images in crisis management has proven effective in various situations. However, they are still limited by the possibility of using communication infrastructures in case of an earthquake and by the presence of people in the area. Moreover, social media images and messages cannot be used to estimate the actual severity of earthquakes and their characteristics effectively. The employment of satellites to monitor changes around the globe grants the possibility of exploiting instrumentation that is not limited by the visible spectrum, the presence of land infrastructures, and people in the affected areas. In this work, we propose a new dataset composed of images taken from Sentinel-1 and a new series of tasks to help monitor earthquakes from a new detailed view. Coupled with the data, we provide a series of traditional machine learning and deep learning models as baselines to assess the effectiveness of ML-based models in earthquake analysis.

Keywords

Remote Sensing, Earthquake Monitoring, Deep Learning, Machine Learning, Regression, Detection, Change Detection

INTRODUCTION

The application of modern machine learning and deep learning solutions to remote sensing is strictly linked to data availability. Thanks to the launch of Landsat-8 (Wulder et al., 2019), MODIS (Justice et al., 2002), and Sentinel missions (Main-Knorn et al., 2017; Torres et al., 2012), general-purpose datasets were released in recent years (Helber et al., 2019; Sumbul et al., 2019; Y. Wang et al., 2022). Emergency management can benefit from these solutions. However, only a few datasets of satellite imagery were publicly released for a limited number of emergency management tasks (e.g., flood detection (Bonafilia et al., 2020) and burned areas delineation (Cambrin et al., 2023)). Ad-hoc image datasets are unavailable for many emergency management tasks. Among the others, public datasets of satellite imagery tailored to training machine learning models for earthquake analysis have not been released to our knowledge.

Earthquake monitoring is mainly based on seismometers that actively record seismic waves to detect, understand the strength, and identify the origin of these events. However, we cannot have global capillary coverage using this instrumentation. Conversely, satellite instruments have a broader view over the globe. Hence, they can help reach remote places worldwide and understand which areas were affected by the earthquakes without sending people to assess the damages and without using seismometers. Tasks such as identifying affected areas and estimating magnitude can be solved offline by collecting satellite images and then applying machine learning models in a data center. Alternatively, the inference can be made directly onboard the satellites using low-resource models, avoiding storing large amounts of data and enabling the possibility of transmitting only a few significant information (e.g.,

*corresponding author

coordinates of the damaged area, the estimated magnitude, etc). This raises another crucial challenge linked to the resource availability of the used systems: employing a large model requires computational resources unavailable on board. Creating an accurate model is not the only requirement. Low resource consumption is mandatory, and memory, CPU, and energy must be used effectively.

In this context, we propose a new dataset of more than a hundred earthquakes around the Earth, containing thousands of tri-temporal samples of the same event using Sentinel-1 data and ISC annotations. Moreover, we assess the effectiveness of shallow and deep learning models to solve tasks related to earthquake monitoring.

Our contributions can be summarized as follows:

- We propose a series of tasks to analyze earthquake-affected areas that can be solved by applying machine learning solutions to Sentinel-1 data.
- We evaluate a series of deep learning and shallow machine learning solutions in a subset of the proposed tasks.
- We pose the tasks as a balance between performance and resource consumption to democratize the proposed solutions and make feasible the application to low-resource devices.
- We publicly release a collection composed of tri-temporal time-series images, allowing the possibility to analyze time-related changes.

For reproducibility, we released the code for the experiments at https://anonymous.4open.science/r/quakeset-4D73/. The dataset is available at https://zenodo.org/records/10423411 to allow the community to experiment with it. Both code and data are released anonymously to comply with the double-blind review process.

RELATED WORKS

Image Processing through Deep Learning

The rise of convolutional neural networks and, more recently, vision transformers significantly improved the capability of processing images. The great data availability and the use of pre-trained neural networks posed deep learning models as state-of-the-art solutions in computer vision over classical machine learning models in benchmark datasets like ImageNet (Deng et al., 2009). The increasing performance of these models is often linked to the increased computational costs at training time and in inference (Xie et al., 2021). Dealing with real-time constraints or low-resource devices poses the challenge of balancing accuracy with resource consumption. Developing specific deep learning models like MobileNets (Howard et al., 2017; Sandler et al., 2018) or MobileViTs Mehta and Rastegari, 2021, 2022 were strictly linked to these constraints. Other solutions were designed in different fashions, including the possibility of employing less accurate real-time versions or the heavier and best-performing ones (Z. Liu et al., 2022; W. Wang et al., 2021; Woo et al., 2023; Xie et al., 2021). Finally, the effectiveness of the deep learning models in computer vision is highly impacted by data availability. Without ad hoc collections of images, the results decrease rapidly when considering domains for which tailored datasets are unavailable or limited.

Image Processing and Deep Learning in Crisis Management

Using textual information to analyze the evolution of ongoing emergencies has proven effective with the solutions proposed in the Incident Stream (McCreadie et al., 2019) and Crisis Facts (McCreadie et al., 2023) tracks of TREC. Although, also the analysis of images taken from social media has proven to be effective in the identification of areas affected by fires (Daly & Thom, 2016), earthquakes, hurricanes, and typhoons (Alam et al., 2017; Nguyen et al., 2017). The training of deep convolutional neural networks grants improved results over traditional techniques like Bag-Of-Visual-Words (Nguyen et al., 2017). The release of Incidents1M (Weber et al., 2022) with around a million images of 43 incidents granted the possibility to pre-train vision transformers and finetune existing pre-trained ones (Long et al., 2023). The significant drawback of all existing solutions is that they rely on social media and RGB images. RGB images can limit the comprehension of certain aspects of the events due to the limitations of the visible spectrum. For instance, using the infrared spectrum in Sentinel-2 helps monitor green areas (Drusch et al., 2012), being vegetation more sensitive to these frequencies. Moreover, using social media data also excludes the possibility of reaching sites with limited human presence or low-quality or damaged communications since they rely on information people send over social media. For these reasons, we propose to use satellite data, specifically Sentinel-1 images.

Machine Learning in Remote Sensing

The advancement of neural networks greatly benefits applications in the remote sensing field (Yuan et al., 2020; X. X. Zhu et al., 2017). The networks designed for RGB images have proven adaptable to other spectral bands like infrared and ultra blue. An extensive series of classical computer vision tasks, such as image classification and segmentation, were adapted to the remote sensing domain. The BigEarthNet (Sumbul et al., 2019) and EuroSAT (Helber et al., 2019) datasets pose the problem of landcover detection as an image classification one. The Seaships dataset (Shao et al., 2018) analyzes an object detection problem, while CaBuAr (Cambrin et al., 2023) analyzes a semantic segmentation problem in the burned area identification context. In addition to the extra spectral bands per image, in many cases, these datasets contain spatiotemporal information granting the possibility of solving tasks like change detection, which aims to find in supervised or unsupervised way the difference between two samples of the same spatial area taken at two different timestamps (Asokan & Anitha, 2019). The WHU dataset (Ji et al., 2018) deals with this challenge for urban planning.

Machine Learning in Seismology

There is an increasing interest in applying machine learning and deep learning in the seismology field (Chandrakumar et al., 2023). The current machine-learning solutions were mainly applied to earthquake seismic waves (P-waves and S-waves) to detect an earthquake in real-time. Machine learning was also applied to distinguish between earthquakes and microtremors (H. Liu et al., 2022) and to phase picking (Mousavi et al., 2020; W. Zhu et al., 2022). Current solutions involve convolution neural networks (Khan & Kwon, 2022; W. Zhu et al., 2022), generative adversarial networks (H. Liu et al., 2022), and transformers (Mousavi et al., 2020) applied to seismic waves. Sentinel-1 imagery has already been tested in earthquake analysis without involving automatic systems (Funning & Garcia, 2019). Specifically, the reported manual analysis highlights the feasibility of earthquake detection using Sentinel-1 data. The authors of Funning and Garcia, 2019 managed to identify approximately 70% of considered earthquakes. Compared to the manual analysis, we do not need to apply preprocessing steps for water bodies or atmospheric correction, and, more importantly, we propose completely automated ML-based approaches. Nevertheless, manual analysis in Funning and Garcia, 2019 involves correlation, which fails to understand non-linear relationships in data, while in our study, we employed many non-linear estimators. Moreover, we collected a time series of images to overcome the possible limitation of single image analysis, opening the possibility of leveraging changes instead of simply analyzing the content of single post-event images. Another difference is that we propose automatic solutions to analyze earthquakes from Sentinel-1 images instead of employing waves or manual solutions. Finally, we address several tasks, including detection and magnitude estimation.

SOURCES DESCRIPTION

In this section, we present the sources of our dataset, which includes Sentinel-1 GRD products and ground truth annotations from the International Seismological Centre (ISC).

Sentinel-1 mission

Sentinel-1 (Torres et al., 2012), with its Synthetic Aperture Radar (SAR), proved its effectiveness in a wide range of applications, and it is well suited for monitoring land changes due to its high revisit time and the capacity to see through clouds. It uses a single C-band operating at a center frequency of 5.405 GHz. SAR supports different acquisition modes in single and dual polarization: Stripmap (SM), Interferometric Wide swath (IW), Extra-Wide swath (EW), and Wave (WV). IW mode is mainly employed in land areas, WV in open oceans, EW in coastal areas, and SM for small islands. This work focused on land earthquakes, so we employed the IW mode, which provides VV and VH polarizations. Sentinel-1 products are available in four different formats: Level-0 Raw, Level-1 Ground Range Detected (GRD), Level-1 Single Look Complex (SLC), and Level-2 Ocean (OCN). We employed the Level-1 GRD products at two resolutions: 10×10 (high) and 40×40 (medium). To get all possible information, we kept only high-resolution samples.

ISC Bulletin

The International Seismological Centre (ISC) provides access to a database of all known and suspected earthquakes from more than 130 seismological agencies around the world (Willemann & Storchak, 2001). The ISC Bulletin ("International Seismological Centre (2023), On-line Bulletin", 2023) contains data from 1900 to the present (2023-11-19). The Reviewed ISC Bulletin, manually checked by ISC analysts and relocated (when there is sufficient data), is typically 24 months behind real-time and is currently up to 2021-11-01. The review procedure is applied to

all earthquakes with unknown magnitude or greater than 3.5. The procedure ensures the hypocenters are coherent with the origin region and that any regional and magnitude data are not missing. According to ISC, around 20% of the entire database is reviewed. The bulletin provides various information regarding each earthquake, such as hypocenter (Bondár & Storchak, 2011), the timestamp of the event, and magnitude (Di Giacomo & Storchak, 2016). Using this information, we can select an area of interest, know when to sample data related or not related to earthquakes, and try to understand different aspects of the events. The magnitude scale employed in this analysis and the ones directly provided by ISC is the mb scale, which can be used for medium entity earthquakes without encountering saturated values (Howell Jr, 1981). The following section describes how we use them to create tri-temporal time-series and ground truth annotations.

DATASET

Table 1. Dataset Summary

Earthquakes	155
Temporal window	2018-2021
Image channels	2 (VV and VH)
Area size	$20km \times 20km$
Temporal difference	1-13 days
Time series length	3
Patch size	512×512
Magnitudes	$>$ 4 mb

We collected the epicenters of all earthquakes from 2018 until 2021 that were reviewed by ISC. We kept the ones defined as *known* events. We retained only the earthquakes that could create some visible effects, filtering the ones with mb magnitude greater than 4. We queried the Copernicus system of the European Space Agency to collect samples from Sentinel-1. For each epicenter longitude-latitude coordinates (E_{lg}, E_{lt}), we collected the area from $(E_{lg} - 0.1, E_{lt} + 0.1)$ to $(E_{lg} + 0.1, E_{lt} - 0.1)$. This creates an area around the epicenter of $\approx 20km \times 20km$. We have also constrained the image selection to IW mode samples. We collected two samples for each earthquake, one before (we call it "pre-event") and one after (we call it "post-event") the event. Since the revisit time of Sentinel-1 is 12 days, we created a temporal window of 13 days before and 13 days after the earthquake. We then selected the two images temporally nearest the event, excluding the day of the incident. We have excluded epicenters for which we do not have all the two samples. Furthermore, we collected an extra sample of the same area in a temporal window from 25 days before to 13 days before the earthquake, selecting the furthest day from the "pre-event" sample to avoid excessive similarities (we call it "neutral"). The availability of the third image for each earthquake helps balance the data distribution when considering the tasks based on the bitemporal time series in the following section.

Figure 1 depicts the four temporal windows considered for each earthquake. To summarize, three Sentinel-1 images are collected for each earthquake: one in the post-event, one in the pre-event, and one in the neutral temporal window.

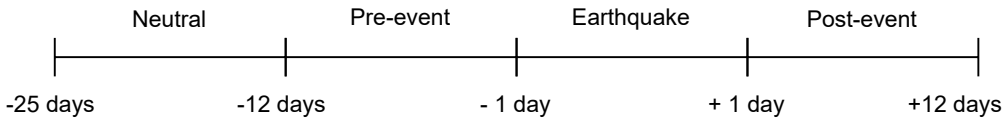


Figure 1. Temporal windows of collected samples

Each sample was preprocessed using SAR backscatter with the sigma0 ellipsoid coefficient. In this way, we collected 155 different events. For each earthquake, we provide three images (neutral, pre-event, post-event), the magnitude expressed in mb, the epicenter coordinates, and the depth (used to identify the hypocenter), when available.

In Figure 2, it is possible to see geographical distribution around the globe. Americas, East Asia, and Oceania are best represented, while Africa has the fewest samples. The identified epicenters cover coastal areas, hinterlands, and islands.

In Figure 3a, you can see the distribution of the magnitudes. We collected sufficient samples to approximate the distribution using a normal with $\mu = 5.64, \sigma = 0.44$. The collected values can be considered non-saturated according to Howell Jr, 1981, which poses the upper limit of the scale to $mb_{sat} \approx 8.1$.

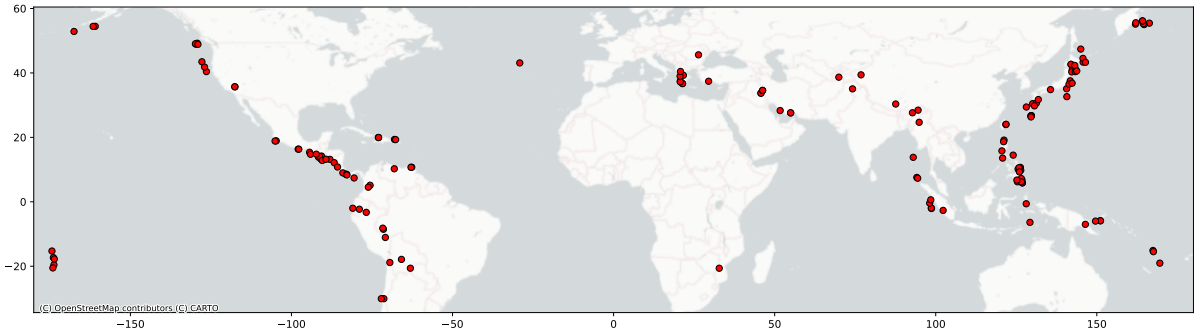
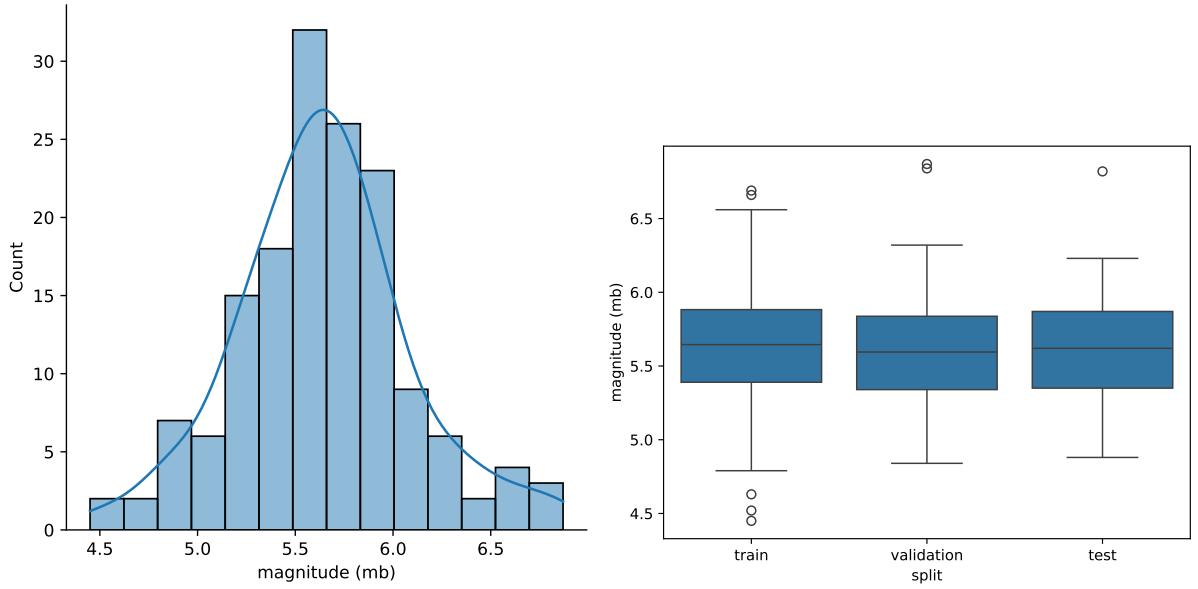


Figure 2. Earthquakes epicenters around the globe

We patched the samples from the center to obtain images of size $512 \times 512 \times 2$, which are more tractable by machine learning models. In this way, we created 1906 patches. We reported a dataset summary in Table 1, highlighting the main features.



(a) Magnitudes distribution

(b) Magnitude distribution over splits

Figure 3. Magnitude distributions

Train, Validation and Test Splits

To create representative splits, we divided the dataset into 70%-15%-15% sets for train, validation, and testing, respectively. We decided to create the splits according to the distribution of the magnitudes. All the images (neutral, pre-event, and post-event) associated with the same earthquake event are in the same set/split. We sample from the original dataset with a uniform distribution. We evaluated the magnitude distribution using the two-sample Kolmogorov-Smirnov test to ensure similar distributions between train-validation and train-test. We obtain a p-value of 0.985 between train-validation and 0.995 between train-test sets. The obtained distributions in Figure 3b highlight the similarities between the three sets in terms of quantiles.

Tasks

Several tasks related to earthquake monitoring can be solved using the data we collected, and we are sharing. Among the others, we suggest using our data to handle the following emergency earthquake-related tasks:

1. Earthquake detection with bitemporal image time series
2. Magnitude regression with bitemporal image time series

3. Hypocenter or epicenter regression with bitemporal image time series
4. Change detection with bitemporal image time series
5. Earthquake detection with a single image
6. Magnitude regression with a single image
7. Hypocenter or epicenter regression with a single image

In the following, we formalize each type of task regardless of the input image time series size (the input can be either a single image or a time series composed of two consecutive images).

Earthquake Detection

The earthquake detection tasks (Tasks 1 and 5) can be formulated as follows. We have a set of time-series $\mathcal{T}$, where each time-series $T_s \in \mathcal{T}$ is composed of N images, with $N \leq 2$, of size $W \times H \times C$ related to the same spatial area S at timesteps $\{T_1, \ldots, T_N\}$ and a ground truth value $G_t \in \{0, 1\}$ (where 1 indicate S was affected by an earthquake, 0 otherwise). Given a training dataset D_{tr}, composed of a set of pairs (T_s, G_t), we train a machine learning model M. Given a test dataset D_{ts}, composed of a set of T_s, we can predict G_t for each sample T_s using M. This can be framed as a supervised classification task.

Magnitude Regression

The magnitude regression tasks (Tasks 2 and 6) can be formulated as follows. We have a set of time-series $\mathcal{T}$, where each time-series $T_s \in \mathcal{T}$ is composed of N images, with $N \leq 2$, of size $W \times H \times C$ related to the same spatial area S at different timesteps $\{T_1, \ldots, T_N\}$ and a ground truth value $G_t \in \{0...M_m\}$ (where M_m is the maximum value for the given magnitude scale). Given a training dataset D_{tr}, composed of a set of pairs (T_s, G_t), we train a machine learning model M. Given a test dataset D_{ts}, composed of a set of T_s, we can regress G_t for each sample T_s using M. This regression task is a supervised one.

Epicenter or Hypocenter Regression

The Epicenter or Hypocenter regression tasks (Tasks 3 and 7) can be formulated as follows. We have a set of time-series $\mathcal{T}$, where each time-series $T_s \in \mathcal{T}$ is composed of N images, with $N \leq 2$, of size $W \times H \times C$ related to the same spatial area S at different timesteps $\{T_1, \ldots, T_N\}$ and a ground truth composed by the coordinates (x, y, d) (where x and y are the epicenter coordinates and d is the depth of the hypocenter) and a binary label $G_t \in \{0, 1\}$ (1 if the area S contains the hypocenter of the earthquake, otherwise 0). Given a training dataset D_{tr}, composed of a set of tuples (T_s, G_t, x, y, d), we train a machine learning model M. Given a test dataset D_{ts}, composed of a set of T_s, we can predict (G_t, x, y, d) for each sample T_s using M. If S is affected by an earthquake but does not contain the hypocenter, it should recognize that the tuple (x, y, d) cannot be determined. The task is supervised and combines classification (presence or absence of the epicenter) and regression (coordinates of the hypocenter).

Change Detection

Change detection (Task 4) can be formulated as follows. We have a set of time-series $\mathcal{T}$, where each time-series $T_s \in \mathcal{T}$ is composed of 2 images of size $W \times H \times C$ related to the same spatial area S at different timesteps $\{T_1, T_2\}$. We want to generate a binary mask M_b of size $W \times H$ for each T_s. The value 1 indicates the pixel is affected by a substantial change; otherwise, it is set to 0. The task is an unsupervised semantic segmentation task. Given a training dataset D_{tr}, composed of a set of T_s, we train a machine learning model M. Given a test dataset D_{ts}, composed of a set of T_s, we can generate M_b for each sample T_s using M.

The previous formulations are general, and the mentioned tasks can be solved using single images (degenerate time series) or bi-temporal time series composed of two images.

EXPERIMENTS

In this section, we present the results obtained for a subset of the proposed tasks, which are Tasks 1, 2, 5 and 6. We have solved the most promising tasks to evaluate the quality of the collected data and the feasibility of the more relevant proposed tasks.

We employed four deep-learning models designed for low-resource devices: MobileNetV2 (Sandler et al., 2018) and ConvNextV2-Atto (Woo et al., 2023), which are convolutional neural networks (CNN), and MiT-B0 (Xie

et al., 2021) and MobileViTV2-1.0 (Mehta & Rastegari, 2022), which are vision transformers (ViT). We also considered some classical shallow learning models: Support Vector Machines (SVM) with 3rd-grade polynomial and Radial Basis Function kernels and Random Forest (RF). We have chosen these solutions because SVMs can exploit non-linear relations, while RFs, being an ensemble, are more robust.

Experimental Settings

This section reports all settings for the employed models and the input datasets.

Input data

Single Image. In this setting, the input data is a collection of single images (i.e., a collection of degenerate time series, each one composed of a single image). We use all the post-event images (as representative of areas already affected by an earthquake) and all the pre-event ones (as representative of areas unaffected by an earthquake). This corresponds to a balanced distribution of the two classes under analysis (affected/unaffected by an earthquake). In this setting, we do not consider the neutral images (see Figure 1). Each image has shape $512 \times 512 \times 2$.

Bi-temporal Time Series. This second setting considers as input a collection of bitemporal time series of images. When dealing with bitemporal time series of images, we use samples of shape $512 \times 512 \times 4$ concatenating the two components of the time series of shape $512 \times 512 \times 2$ along the channel axis (this approach allows using machine learning models not designed for time series). Each time series is related to a specific earthquake and comprises a pre-event and a post-event image (example of affected area class object) or a neutral and a pre-event sample (example of unaffected area class object). This ensures a balanced distribution of positive and negative examples for the two classes.

Models

Deep Learning Models. We trained convolutional neural networks for six epochs while vision transformers for ten epochs. This was done due to the slower learning of transformers. The batch size is 16. We used an AdamW optimizer with a cosine-annealing learning rate scheduler with a warmup of 0.1 of the total training steps. The learning rate starts from 0.0001 for CNNs and from 0.001 for ViTs. All models are initialized without any pre-trained weights. The loss is Mean Squared Error (MSE) for regression tasks and Cross-Entropy loss for classification.

Shallow Models. We trained the Support Vector Machines with 3rd-degree polynomial and RBF kernels. Due to the high number of features, we applied PCA before using classical models. In this way, the size of the features of each sample is reduced to ≈ 2000.

Metrics

We analyzed the results in terms of accuracy for classification tasks since the class distribution is balanced ($\approx 50\%$ of positive and negative samples). Accuracy is defined as:

$$Accuracy = \frac{CC}{N} \tag{1}$$

where CC is the number of samples predicted correctly, and N is the number of samples we asked for a prediction.

We analyzed the results regarding Mean Absolute Error (MAE) for the regression tasks. Still, we also evaluated the accuracy in identifying the occurrence of an earthquake thresholding the predicted magnitude (if magnitude < 1, then no earthquake occurred; otherwise, an event affected the area). Mean Absolute Error is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i}^{N} |y_i - \hat{y_i}| \tag{2}$$

where N is the number of samples, y_i is the ground truth magnitude, and $\hat{y_i}$ is the predicted magnitude for the ith sample.

We are also interested in the resource consumption of the presented models, so we report the number of parameters for deep learning models, inference time in seconds, and Mega Floating Points Operation Per seconds (MFLOPs) computed on an Intel(R) Core(TM) i9-10980XE CPU.

Earthquake Detection with Bitemporal Time Series

We report in Table 2 the results for Task 1. The Random Forest Classifier is the best shallow model in terms of accuracy, providing reasonably high accuracy with a few FLOPs.

Regarding deep learning models, MobileViTV2 struggles to compete with the others, obtaining −20% accuracy and 2× FLOPs compared to Mit-B0. The best accuracy results are obtained by the CNNs, which are also less resource-hungry than transformers. MobileNet achieves the best result (94.72%) with fewer FLOPs than the other deep learning models. The differences between this model and the RFC classifier regarding FLOPs and accuracy are evident. Hence, deep learning models are needed to achieve high accuracy values at the cost of a more elevated resource consumption (two orders of magnitude more FLOPs than RFC). However, we notice that the "smallest" model (fewer parameters) of each deep learning category performs better than the "bigger" model of the same category. Specifically, MobileNetV2 (CNN) is more accurate than ConvNextV2 (CNN) and MiT-B0 (ViT) is more accurate than MobileVitV2 (ViT). Hence, smaller and less resource-demanding deep learning models are better for solving this task.

Table 2. Performance of models for earthquake detection with bitemporal time-series

Model	Params	Accuracy ↑	Time (s) ↓	MFLOPs ↓
SVC (RBF kernel)	-	0.6341	0.3640	1.0486
SVC (Poly kernel)	-	0.5440	0.3138	1.0486
RFC	-	0.7241	0.3130	1.0508
MobileNetV2 (CNN)	2.2M	0.9472	0.1109	207.5949
MiT-B0 (ViT)	3.4M	0.8865	0.0909	430.7062
ConvNextV2 (CNN)	3.7M	0.9374	0.1003	373.7727
MobileViTV2 (ViT)	4.9M	0.6536	0.1746	964.5956

Magnitude Regression with Bitemporal Time Series

In Table 3, we report the results obtained for Task 2. Classical solutions like Support Vector Regressors and Random Forest Regressors struggle to compete with deep learning models. In this case, the SVR with RBF kernel gets the best results in terms of accuracy among shallow machine learning models, despite RFR being the best in MAE. This means SVR can better distinguish between positive and negative samples, while RFR tends to approximate the magnitude better. Another critical factor is the accuracies of the shallow models, which are all near 0.5, being generally incapable of making distinctions between the classes.

Deep learning models are also more accurate (in terms of accuracy and MAE) for this task. Deep learning models provide higher accuracies and MAEs, distinguishing affected and unaffected areas. MiT-B0, which achieves worse MAE than CNNs (ConvNextV2 and MobileNetV2), gets an accuracy comparable to ConvNextV2 and MobileNetV2. Hence, Mit-B0, ConvNext, and MobileNet have the same ability to distinguish between earthquake-affected/unaffected areas. However, ConvNextV2 and MobileNetV2 estimate the magnitude values better. The performances of MobileViTV2 are not as satisfactory as the other deep-learning models. Similarly to the previous task, MobileNetV2 obtains the best performance under all metrics among the deep learning models despite being the "smallest" deep learning model. Furthermore, MobileNetV2 (CNN) is again more accurate (accuracy) and precise (MAE) than ConvNextV2 (CNN), and MiT-B0 (ViT) is more accurate and precise than MobileVitV2 (ViT). Hence, smaller and less resource-demanding deep learning models are also better for solving this task.

Table 3. Performance of models for magnitude regression with bitemporal time-series

Model	Params	MAE ↓	Accuracy ↑	Time (s) ↓	MFLOPs ↓
SVR (RBF kernel)	-	2.2368	0.5519	0.3640	1.0486
SVR (Poly kernel)	-	2.6015	0.5440	0.3138	1.0486
RFR	-	1.9930	0.5440	0.3130	1.0508
MobileNetV2 (CNN)	2.2M	0.5456	0.9374	0.1109	207.5949
MiT-B0 (ViT)	3.4M	0.8496	0.9276	0.0909	430.7062
ConvNextV2 (CNN)	3.7M	0.6494	0.9315	0.1003	373.7727
MobileViTV2 (ViT)	4.9M	1.7612	0.7378	0.1746	964.5956

Earthquake Detection with a Single Image

Table 4 reports the results for Task 5. It shows the models struggle to discriminate between areas affected by earthquakes and the non-affected ones. The benefits in reducing the computation costs compared to the models trained on bitemporal input data are evident only for shallow models ($\approx 0.5\times$ FLOPs than bitemporal settings), while the deep learning models gain is not worth the decrease in terms of accuracy (compare Table 4 with Table 2).

The best shallow model to solve Task 5 is SVC with RBF kernel, which achieves 0.60 of accuracy in a few FLOPs. Transformers get worse results than this model (-6% in the worst case). MobileNetV2 provides the best results (with $+10\%$ than SVC with RBF kernel), followed by ConvNextV2.

MobileNetV2 with single images achieves -24% in terms of accuracy than bitemporal settings (95%), and the results are comparable to those of Random Forest (RFC) with bitemporal input data (72%).

A consistent performance decrease using only one image was expected, as shown in the manual analysis of interferograms of Funning and Garcia, 2019 in which the authors highlighted that earthquake detection is feasible by looking at the differences between the two phases.

Table 4. Performance of models for earthquake detection with a single image

Model	Params	Accuracy ↑	Time (s) ↓	MFLOPs ↓
SVC (RBF kernel)	-	0.6047	0.1488	0.524288
SVC (Poly kernel)	-	0.5440	0.1472	0.524288
RFC	-	0.5186	0.1655	0.526554
MobileNetV2 (CNN)	2.2M	0.7104	0.0862	203.3172
MiT-B0 (ViT)	3.4M	0.5440	0.0710	424.8801
ConvNextV2 (CNN)	3.7M	0.6419	0.0609	370.8897
MobileViTV2 (ViT)	4.9M	0.5988	0.1171	960.3690

Magnitude Regression with a Single Image

The results in Table 5 for Task 6 show, as expected from the previous experiments, that the models are not well capable of identifying affected areas using one single image. We can conclude from the reported accuracies that this is common for both shallow and deep learning models.

The differences between the used models in terms of MAE and accuracy are less evident than before. Shallow models align with previous performances for Task 2. Conversely, deep learning models decrease their mean performance of -30% in terms of accuracy and +1.3 in MAE compared to the bitemporal setting.

MobileNetV2 gets the best results in terms of MAE but does not perform well in distinguishing between affected and unaffected areas. The decrease in accuracy compared to Task 5 is -16%. ConvNextV2 has the best accuracy, while MiT-B0 is the most balanced solution in both metrics. MobileViTV2 struggles to compete with shallow models in both metrics, too.

Table 5. Performance of models for magnitude regression with a single image

Model	Params	MAE ↓	Accuracy ↑	Time ↓	MFLOPs ↓
SVR (RBF kernel)	-	2.4088	0.5519	0.1488	0.524288
SVR (Poly kernel)	-	2.5960	0.5440	0.1472	0.524288
RFR	-	2.6204	0.5440	0.1655	0.526554
MobileNetV2 (CNN)	2.2M	2.1595	0.5734	0.0862	203.3172
MiT-B0 (ViT)	3.4M	2.1821	0.5832	0.0710	424.8801
ConvNextV2 (CNN)	3.7M	2.2353	0.5851	0.0609	370.8897
MobileViTV2 (ViT)	4.9M	2.4645	0.5440	0.1171	960.3690

Preliminary Integration of a Global Seismic Hazard Map

The integration of the right in-domain information could be beneficial in addressing the proposed tasks. In our preliminary assessment, we evaluated the effectiveness of integrating a global seismic hazard map (Johnson et al.,

2023) as input. Specifically, we performed preliminary experiments by adding an extra channel associated with the seismic hazard information to the bi-temporal images. According to the preliminary results, this additional information does not provide any benefits in earthquake detection and magnitude regression. In the future, we will investigate other in-domain sources as well as other ways of integrating domain-specific information in the training of the models.

LIMITATIONS

The study employed only visual features without any external knowledge. Including other historical features and employing domain knowledge (e.g., geophysics, optics, and earth sciences) could strengthen the results on more complex tasks like hypocenter regression, which could benefit from extra information like geological plaques.

We employed single images and bi-temporal time series, although a longer time series could be investigated to include more historical information about the analyzed areas.

CONCLUSION AND FUTURE DIRECTIONS

Our experiments show the possibility of exploiting machine learning techniques to understand earthquakes using the visual components collected by the SAR instrument with satisfactory results.

We explored the usage of Sentinel-1 imagery, which is easily available to the public and for which there were studies proving effectiveness, although a different source (e.g., from drones and airplanes) or multi-source approach could be employed to adopt less expensive instrumentation.

We employed low-resource solutions (which are generally "small" models in terms of parameters) to show the unnecessary need for complex and bigger models. Results show that smaller deep-learning models are more accurate and precise than the bigger ones, contrasting expectations. Designing ad-hoc low-resource networks is an underexplored direction that should be addressed in remote sensing and hazard management.

In future works, we plan to investigate the remaining proposed tasks, the design of new ad-hoc low-resource models, and the inclusion of in-domain knowledge.

REFERENCES

Alam, F., Imran, M., & Ofli, F. (2017). Image4act: Online social media image processing for disaster response. *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017*, 601–604.

Asokan, A., & Anitha, J. (2019). Change detection techniques for remote sensing applications: A survey. *Earth Science Informatics*, *12*, 143–160.

Bonafilia, D., Tellman, B., Anderson, T., & Issenberg, E. (2020). Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Bondár, I., & Storchak, D. (2011). Improved location procedures at the international seismological centre. *Geophysical Journal International*, *186*(3), 1220–1244.

Cambrin, D. R., Colomba, L., & Garza, P. (2023). Cabuar: California burned areas dataset for delineation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, *11*(3), 106–113. https://doi.org/10.1109/MGRS.2023.3292467

Chandrakumar, C., Prasanna, R., Stephens, M., Tan, M. L., Holden, C., Punchihewa, A., Becker, J. S., Jeong, S., & Ravishan, D. (2023). Algorithms for detecting p-waves and earthquake magnitude estimation: Initial literature review findings.

Daly, S., & Thom, J. A. (2016). Mining and classifying image posts on social media to analyse fires. *ISCRAM*, 1–14.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, 248–255.

Di Giacomo, D., & Storchak, D. A. (2016). A scheme to set preferred magnitudes in the isc bulletin. *Journal of Seismology*, *20*, 555–567.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al. (2012). Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, *120*, 25–36.

Funning, G. J., & Garcia, A. (2019). A systematic study of earthquake detectability using sentinel-1 interferometric wide-swath data. *Geophysical Journal International*, *216*(1), 332–349.

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(7), 2217–2226.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Howell Jr, B. (1981). On the saturation of earthquake magnitudes. *Bulletin of the Seismological Society of America*, *71*(5), 1401–1422.

International seismological centre (2023), on-line bulletin. (2023). https://doi.org/https://doi.org/10.31905/D808B830

Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on geoscience and remote sensing*, *57*(1), 574–586.

Johnson, K., Villani, M., Bayliss, K., Brooks, C., Chandrasekhar, S., Chartier, T., Chen, Y.-S., Garcia-Pelaez, J., Gee, R., Styron, R., Rood, A., Simionato, M., & Pagani, M. (2023). Global seismic hazard map. https://doi.org/10.5281/ZENODO.8409647

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., & Morisette, J. (2002). An overview of modis land data processing and product status. *Remote sensing of Environment*, *83*(1-2), 3–15.

Khan, I., & Kwon, Y.-W. (2022). P-detector: Real-time p-wave detection in a seismic waveform recorded on a low-cost mems accelerometer using deep learning. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5.

Liu, H., Li, S., & Song, J. (2022). Discrimination between earthquake p waves and microtremors via a generative adversarial network. *Bulletin of the Seismological Society of America*, *112*(2), 669–679.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11976–11986.

Long, Z., McCreadie, R., & Imran, M. (2023). Crisisvit: A robust vision transformer for crisis image classification. https://doi.org/http://dx.doi.org/10.59297/SDSM9194

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2cor for sentinel-2. *Image and Signal Processing for Remote Sensing XXIII*, *10427*, 37–48.

McCreadie, R., Buntain, C., & Soboroff, I. (2019). Trec incident streams: Finding actionable information on social media.

McCreadie, R., Buntain, C., Soboroff, I., & Ellis, A. (2023). Crisisfacts: Buidling and evaluating crisis timelines. *Proceedings of the 20th International Conference on Information Systems for Crisis Response and Management*. https://doi.org/http://dx.doi.org/10.59297/JVQZ9405

Mehta, S., & Rastegari, M. (2021). Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.

Mehta, S., & Rastegari, M. (2022). Separable self-attention for mobile vision transformers. *arXiv preprint arXiv:2206.02680*.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, *11*(1), 3952.

Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage assessment from social media imagery data during disasters. *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017*, 569–576.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.

Shao, Z., Wu, W., Wang, Z., Du, W., & Li, C. (2018). Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE transactions on multimedia*, *20*(10), 2593–2604.

Sumbul, G., Charfuelan, M., Demir, B., & Markl, V. (2019). Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., et al. (2012). Gmes sentinel-1 mission. *Remote sensing of environment*, *120*, 9–24.

Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., & Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. *Proceedings of the IEEE/CVF international conference on computer vision*, 568–578.

Wang, Y., Braham, N. A. A., Xiong, Z., Liu, C., Albrecht, C. M., & Zhu, X. X. (2022). Ssl4eo-s12: A large-scale multi-modal, multi-temporal dataset for self-supervised learning in earth observation. *arXiv preprint arXiv:2211.07044*.

Weber, E., Papadopoulos, D. P., Lapedriza, A., Ofli, F., Imran, M., & Torralba, A. (2022). Incidents1m: A large-scale dataset of images with natural disasters, damage, and incidents. *IEEE transactions on pattern analysis and machine intelligence*, *45*(4), 4768–4781.

Willemann, R. J., & Storchak, D. A. (2001). Data collection at the international seismological centre. *Seismological Research Letters*, *72*(4), 440–453.

Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16133–16142.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., et al. (2019). Current status of landsat program, science, and applications. *Remote sensing of environment*, *225*, 127–147.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, *34*, 12077–12090.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716.

Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2022). An end-to-end earthquake detection method for joint phase picking and association using deep learning. *Journal of Geophysical Research: Solid Earth*, *127*(3), e2021JB023283.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307