Report of Transnational Access Projects

(Note: the information here will be publicly disclosed in the Geo-INQUIRE website, do not include sensitive information)

Project ID: *C1_TA3-84-1_2*

Principal investigator: *Maya Keren, Tel Aviv University (Israel)*

Project team (if applicable):

Project title: AI-based back-azimuth for EEWS

Project acronym: *AI-BAZ*

Hosting installation: *TA3-84-1, Università di Napoli Federico II, Physics Department*

Hosting team: *Gaetano Festa[1], Antonio Scala[1], Francesco Scotto di Uccio[1], Claudio Strumia[1], Sonia Sorrentino[1]*

[1] Università di Napoli Federico II – Department of Physics

Period of access: *From 16/05/2025 to 26/06/2025*

Report of activities:

Robust back-azimuth assessment using single-station seismic recordings is useful for real-time earthquake location and earthquake early warning systems. Nevertheless, existing approaches often lack in-depth error analysis and suffer from high prediction errors. They also provide little uncertainty quantification and offer limited evaluation of how well models generalize to unseen geographic regions. Moreover, the broader research landscape has paid relatively little direct attention to back-azimuth prediction as a standalone task.

Before the starting of the TA activity, we developed machine learning models trained on a large-scale global seismogram dataset to estimate back-azimuth at individual three-component seismic stations. Our approach combines PCA-derived features with additional time-series and seismological descriptors, optimizes time windows for feature extraction, and relies solely on explainable features. A new method called Adjustable Time Window-PCA (ATW-PCA), chooses the best time window according to the highest east-north absolute correlation, and perform PCA on it. The dataset used in this study is the Stanford Earthquake Dataset (STEAD), a large global collection of single-station seismograms with earthquake metadata.

We train two machine learning models, XGBoost (XGB) and a Bayesian Neural Network (BNN). Their performance in estimating back-azimuth is benchmarked against the classical PCA-based method. On the STEAD dataset, both XGB and BNN substantially outperform PCA: mean absolute angular error decreases

Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call

from 39.4◦ to 29.8◦ (XGB) and 29.2◦ (BNN), median absolute errors from 18.8◦ to 15.2◦ and 14.6◦, and standard deviation from 49.4◦ to 38.9◦ and 38.6◦.

The TA activity has been dedicated to evaluating the overall capacity of globally trained models to unexplored regions. In this frame, we performed an out-of-sample test using events from the Iripinia Near Fault Observatory (INFO) dataset (spanning longitudes 14.7 ◦–16.1 ◦E and latitudes 40.1 ◦–41.2 ◦N), which were not included in the STEAD training data. Although the STEAD dataset contains a single event occurred in the Irpinia region (in 2010), this event predates the data used for the TA experiment (collected since 2014) and does not overlap temporally with the test set used in this evaluation, which consists of 193 events. Importantly, we used the same models that were trained on the STEAD training set, without any additional retraining or fine-tuning, and applied them directly to the INFO data.

*The out-of-sample evaluation confirms that the machine learning models successfully learned meaningful relationships between the input features and back-azimuth and effectively generalized to data not included in the training set. As shown in Table **1**, both the BNN and XGB models achieved performance comparable to or better than the PCA-based baselines. The XGB model achieved the best results, reducing the mean absolute angular error to 29.5°, compared to 36.0° for PCA —an improvement of approximately 6.5°. The median absolute error also decreased to 18.3°, compared to 19.3° for PCA. The BNN model achieved a mean absolute angular error of 32.0° and a median error of 20.4°.*

Table 1: Mean absolute angular errors for the different techniques tested on the INFO dataset

Model	Mean	Median	STD
ISNet PCA	36.0	19.3	38.9
ISNet ATW-PCA	35.5	19.7	37.6
ISNet XGB	**29.5**	**18.3**	**32.8**
ISNet BNN	32.0	20.4	34.3

These results are particularly notable given that the INFO dataset is entirely independent of STEAD and may differ in instrumentation, noise characteristics, and underlying geophysical conditions. The characteristics of the events differ significantly, as shown in the panels of Figure 1. Furthermore, the performance of the XGB model on INFO dataset closely matches its performance on the STEAD test set: a mean absolute angular error of 29.8°. This consistency highlights the ability of the model to generalize across various seismic regions and suggests strong potential for globally applicable machine learning models in seismological applications.

Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call

Project outcomes:

Integrated data, consisting of the catalogue reporting the back-azimuth angle estimates for the analysed events from the different techniques can be openly accessed at the following DOI: https://doi.org/10.5281/zenodo.19348231 under a CC-BY 4.0 open license.

Note: Data, products, software and publications resulting from TA activities must be publicly accessible under a CC-BY 4.0, GPLv3 or equivalent open license. No embargos beyond June 2026 are allowed. They must cite Geo-INQUIRE as the source of funding. Minimal citation: "*Geo-INQUIRE is funded by the European Union (GA 101058518)*".

Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call