P3.5-423

Application and Evaluation of Deep learning Phase Pickers for Local Aftershock Monitoring

K.U. Semin, T.C. Destici, N.M. Ozel, S. Kocak, M.U. Teoman

Boğaziçi University
Kandilli Observatory & Earthquake Research Institute

INTRODUCTION AND MAIN RESULTS

This poster provides an overall performance comparison of 3 different Deep Learning Pickers by calculating the Precision, Recall and F1 score and pick timings compared to catalogs of two earthquake sequences in Türkiye.

SnT 2025
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

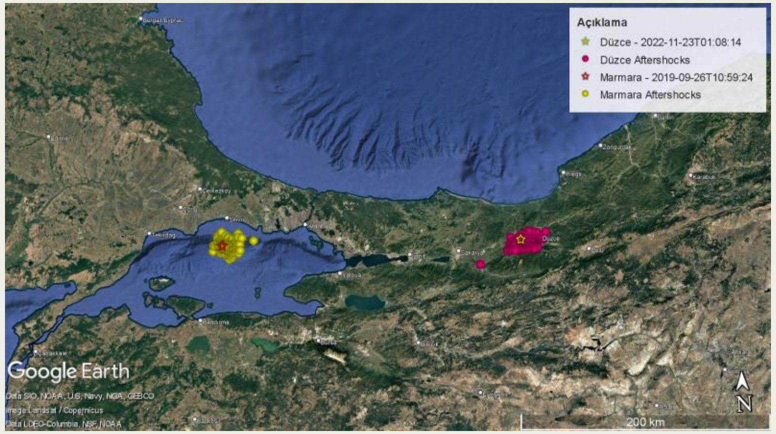
Application and Evaluation of Deep learning Phase Pickers for Local Aftershock Monitoring

K.U. Semin, T.C. Destici, N.M. Ozel, S. Kocak, M.U. Teoman

P3.5-423

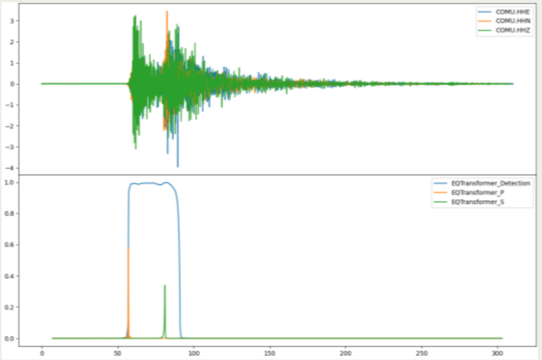
Introduction

The use of Machine Learning (ML) methods in seismology has gained significant attention in recent years, driven by the availability of large, high-quality datasets. While ML is applied to various seismological tasks, it is most commonly used for seismic signal detection, phase picking, and classification. Recent studies show that deep learning models such as EQTransformer, PhaseNet, and Generalized Phase Detection (GPD) perform remarkably well in these areas, often rivaling waveform correlation methods, particularly in low-seismicity regions. In this study, we applied these pickers to aftershocks of two moderate earthquakes in occurred on Northwestern part of Türkiye (***Mw 5.7 Marmara Sea on 26.09.2019 and Mw 6.0 Düzce on 23.11.2022***) and pickers were compared with the Kandilli Observatory catalogue, focusing on the correct identification of P and S phases and arrival time differences between analyst picks and ML outputs.



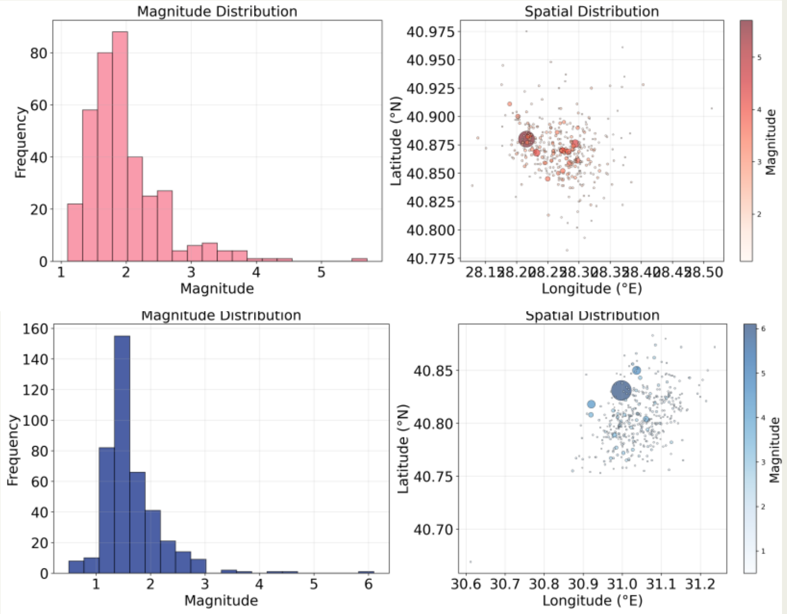
Method

- The **Seisbench** python library was used to perform the comparison of the ML pickers. Waveforms are preprocessed according to original training method of each picker.
- All the waveforms are sampled with 100 Hz, For the EQTransformer, detrend and 1-45 Hz bandpass filter was applied. For the PhaseNet, data were normalized and lastly for GPD detrending and 2 Hz High-pass filter were applied.
- All of the pickers used pre-trained models with original datasets. We haven't done any training using our own data. We considered **abs(T_{cat} – T_{ml}) <= 1.0** between analyst pick and ML picks as True Positive (TP) , else the pick considered as False Positive (FP) and if the analyst pick in the catalog was not detected by pickers it is labelled as False Negative (FN).



A successful detection of the main event 2019-09-26 10:59 Marmara Sea earthquake by EQTransformer at COMU station

Data



Marmara Sea (top) and Düzce (bottom) Earthquake Sequences

Figure shows the magnitude distribution histogram along with spatial and temporal distribution of the aftershocks. For Marmara Sea earthquake we collected 7 days of aftershocks from Sep. 26 to Oct. 2, **369** earthquakes in total. The catalog contains a total of 6968 P-wave phases and 1494 S-wave phases picked by the analyst. On the other hand, we have selected 4 days of aftershocks and **412** events from Nov. 23 to Nov. 27 from Düzce earthquake sequence. The catalog contains a total of 4784 P-wave phases and 1408 S-wave phases manually picked by the analyst.

Application and Evaluation of Deep learning Phase Pickers for Local Aftershock Monitoring

K.U. Semin et. al.

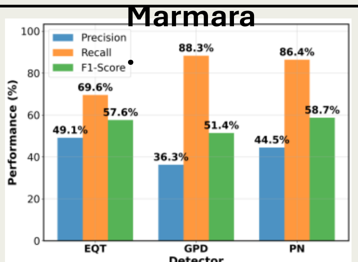
Results / Overall Performance

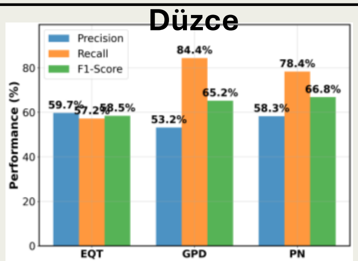
Overall results given in the tables below suggest that PhaseNet consistently achieves the highest F1-scores (66.8% Düzce, 58.7% Marmara), establishing it as the most balanced performer for operational earthquake monitoring. EQTransformer demonstrated better precision rates (59.7% Düzce, 49.1% Marmara) which means it has the lowest False Positives while GPD exhibited the best recall values (84.4% Düzce, 88.3% Marmara) optimal for comprehensive event detection despite elevated false positive rates.

Tables show the statistical results of ML detectors for Marmara (top) and Düzce (bottom) earthquakes.

Mar.	TP	FP	FN	Preci.	Recall	F1
EQT	5,041	5,217	2,207	49.1%	69.6%	57.6%
GPD	4,576	8,042	605	36.3%	**88.3%**	51.4%
PN	5,491	6,857	865	44.5%	86.4%	**58.7%**

Düzce	TP	FP	FN	Preci.	Recall	F1
EQT	3,232	2,179	2,415	59.7%	57.2%	**58.5%**
GPD	3,526	3,106	651	53.2%	**84.4%**	65.2%
PN	3,851	2,757	1,064	58.3%	78.4%	**66.8%**





Results / Phase Specific

Detector	Mar. Preci.	Mar. Recall	Mar. F1	Düzce Preci.	Düzce Recall	Düzce F1
EQT P	83.7%	67.8%	74.9%	89.4%	53.9%	67.2%
GPD P	**89.2%**	**85.8%**	**87.4%**	**91.3%**	**81.1%**	**85.9%**
PN P	83.5%	84.2%	83.8%	88.7%	75.1%	81.4%
EQT S	**18.9%**	77.5%	**30.4%**	**32.3%**	68.1%	43.8%
GPD S	13.4%	**96.5%**	23.6%	26.8%	**93.4%**	41.7%
PN S	16.8%	95.1%	28.6%	29.9%	88.9%	**44.7%**

Tables show the statistical results of ML detectors for Marmara (top) and Düzce (bottom) sequences by Phase type.

Phase specific results above shows that, there is a drastic difference between P-waves and S-waves. If we look at the P-wave results, GPD achieves best results in each performance metrics for both earthquakes, followed by PhaseNet and EQTransformer.

S-wave pick performance of all the detectors exhibits low performance in Precision and F1 scores. This is due to low number of analyst S-picks in the catalog as we are comparing for catalog completeness, most of the real S-picks by the ML detectors labelled as FP which lowers the precision and F1 scores. ML detectors actually very good in picking S phases evident from the very high Recall values especially for GPD and PhaseNet.

Results and Conclusion

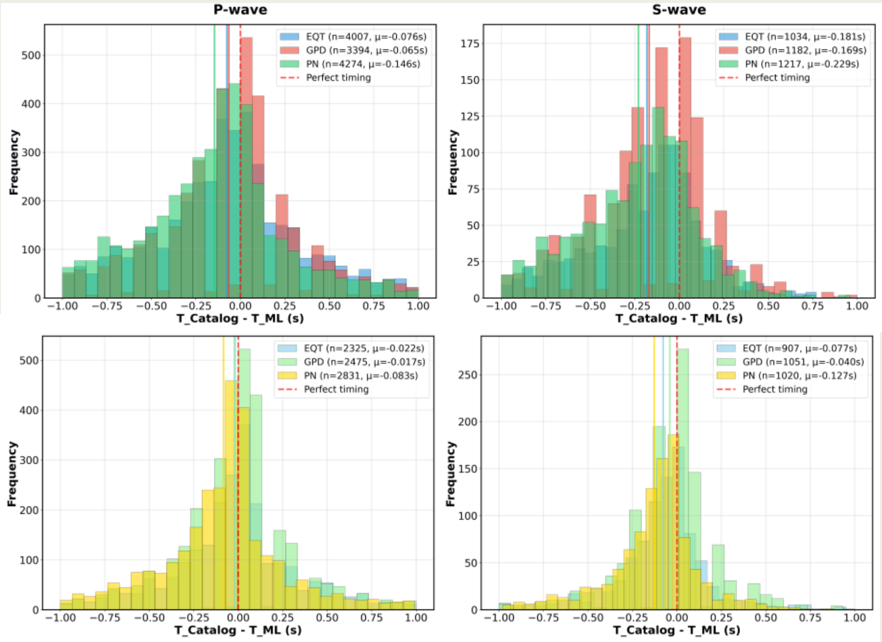


Figure shows the P and S-wave phase pick time differences between the catalog and AI pickers as histogram plots for Marmara (top) and Düzce (bottom) aftershocks.

Arrival time differences between catalog and ML picks indicate that EQT and GPD show the best agreement with catalog picks, while PN has higher mean difference. For the Düzce earthquake, EQT and GPD are nearly aligned with the catalog, indicating better overall timing accuracy.

Our study indicates that picker performances are lower than in their original papers. However, the results are based on pre-trained models and original probability thresholds, therefore, training with local data and/ or adjusting the thresholds could improve the performance.