Thessaloniki, 30/03/2022

Machine learning in Seismology

Dario Jozinović

Swiss Seismological Service @ ETH Zurich

Funded by the Horizon 2020 Framework Programme of the European Union

COST ACTION CA17137
A NETWORK FOR GRAVITATIONAL WAVES, GEOPHYSICS AND MACHINE LEARNING

Schweizerischer Erdbebendienst
Service Sismologique Suisse
Servizio Sismico Svizzero
Swiss Seismological Service

ETH zürich

G2NET

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Contents

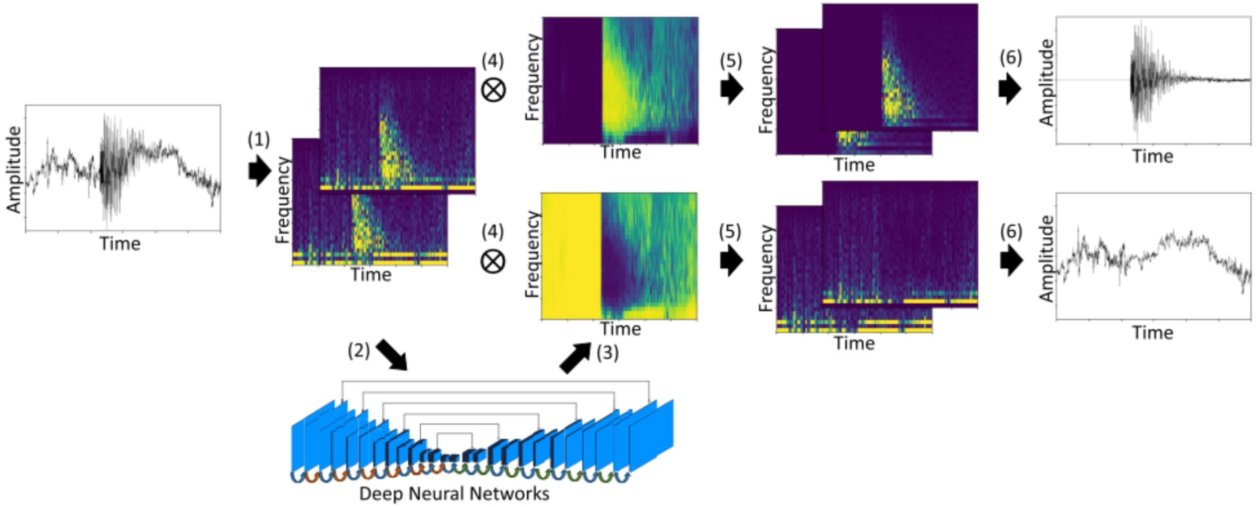
- Earthquake waveforms denoising
- Seismological data for ML
- Most common uses
- Earthquake early warning
- Seismic waveforms as time-series examples

Deep Denoiser

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 9476-9488.

- U-net that takes spectrograms of a noisy seismogram as input, and produces separate seismograms of earthquake and noise
- Dataset: 56k earthquake waveforms and 179k noise waveforms

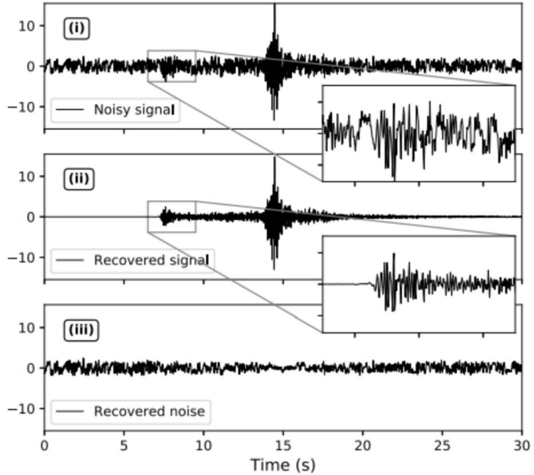
- Earthquake and noise waveforms are randomly combined during training to obtain noisy earthquake waveforms

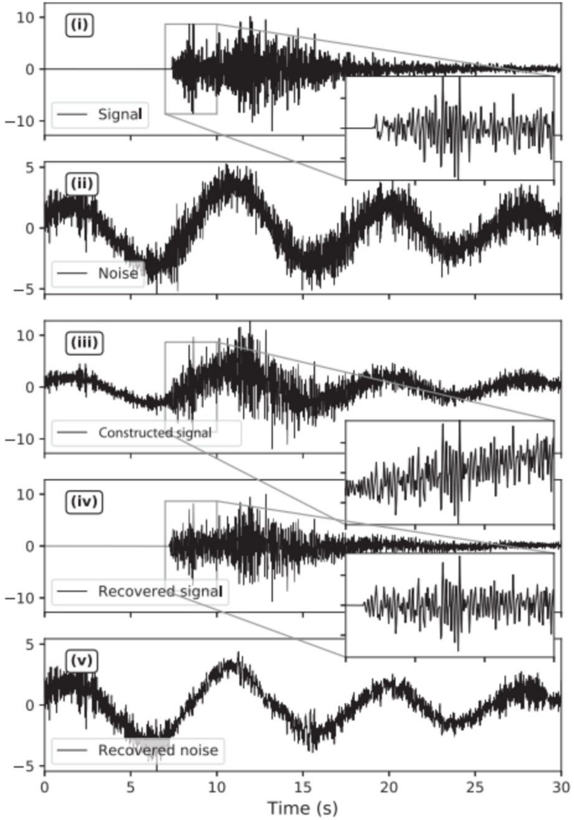


Deep Denoiser

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 9476-9488.

- DeepDenoiser performs extremely well even when the signal and the noise have a similar frequency band
- Robust to presence of non-earthquake signals, introduces low-to-none phase shift
- Generalizes well to non-seen examples with lower Signal-to-Noise ratio
- A test of using it as a pre-processing tool showed improvement in earthquake detection precision from 35% to 95%
- Single-station method

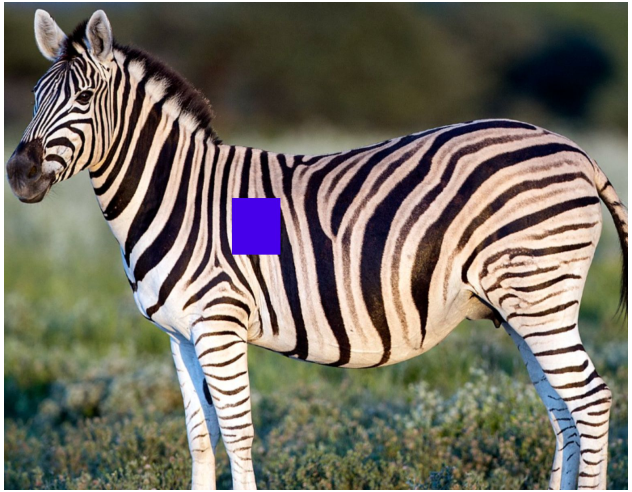


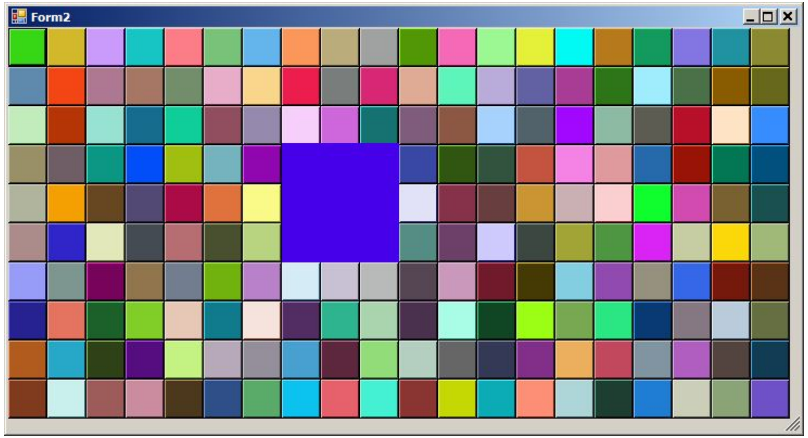


Blind denoising of DAS data

van den Ende, M., Lior, I., Ampuero, J. P., Sladen, A., Ferrari, A., & Richard, C. (2021). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data. IEEE Transactions on Neural Networks and Learning Systems.

- Distributed acoustic sensing (DAS) - technology that turns a fiber-optic cable (including commercial) into a dense array of single-component seismometers
- It allows for densification of existing seismic networks and deployment of sensors in previously inaccessible areas
- Can lead to many more earthquake signals detected - but also records various sources of noise

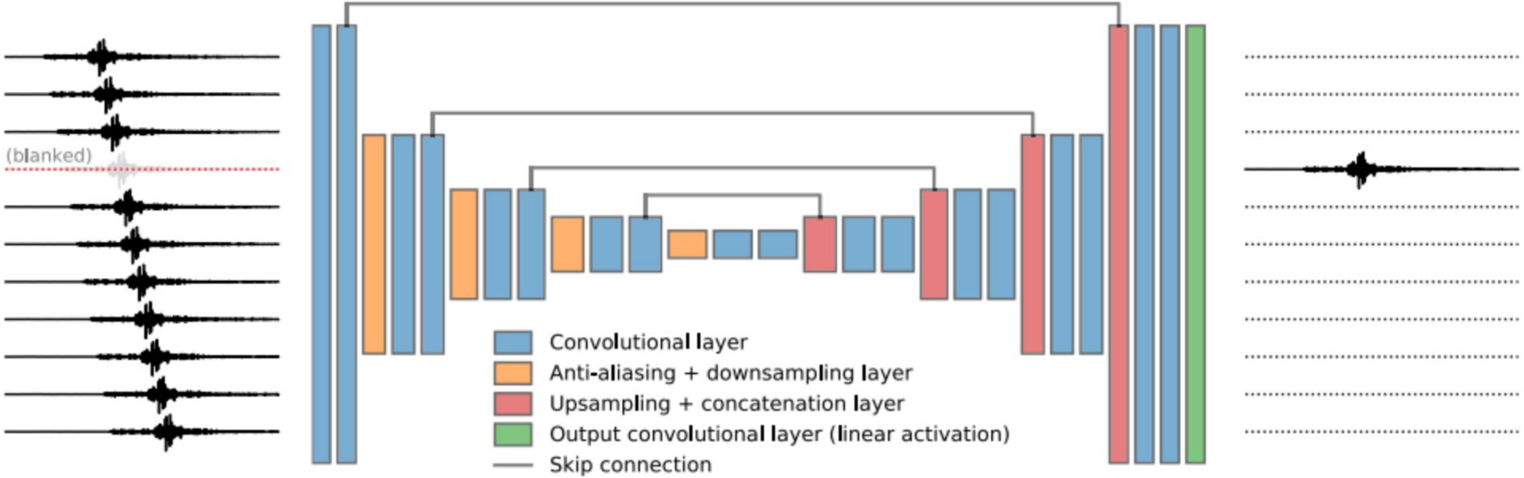




Blind denoising of DAS data

van den Ende, M., Lior, I., Ampuero, J. P., Sladen, A., Ferrari, A., & Richard, C. (2021). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data. IEEE Transactions on Neural Networks and Learning Systems.

- U-NET architecture that takes 11 waveforms as input, of which one is blanked (set to zeros). The output is the waveform that was blanked in the input
- Assumption: signal of interest is spatio-temporally coherent, (most of) the noise is incoherent
- The training is self-supervised: no clean signal and noise data is required



ML for Seismology

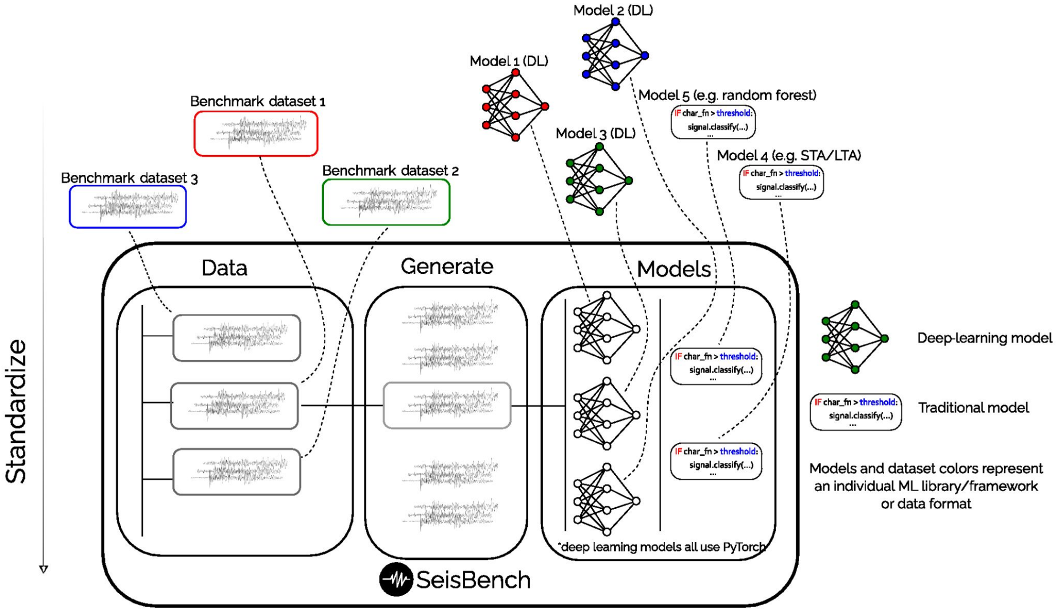
- Seismology - a data intensive field with a vast number of seismic waveforms recorded every day, with most of them publicly available through existing webservices
- Waveforms accompanied by metadata, very often extracted through manual analysis
- A need for automatization of processes that would free up human resources and use them for more science-oriented tasks
- ML benefits from large amount of labeled data

SeisBench
A toolbox for machine learning in seismology

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. SeisBench—A toolbox for machine learning in seismology. Seismological Society of America, 93(3), pp.1695-1709.

- Large amounts of seismic data available and an explosion in ML applications in the recent years - a need for standardization and benchmarking
- SeisBench - a standardization agreement for ML-ready dataset construction

- Result - a python library that allows easy access to a set of benchmark datasets, all having the same metadata
- Library allows for an easy application of available models to any available dataset

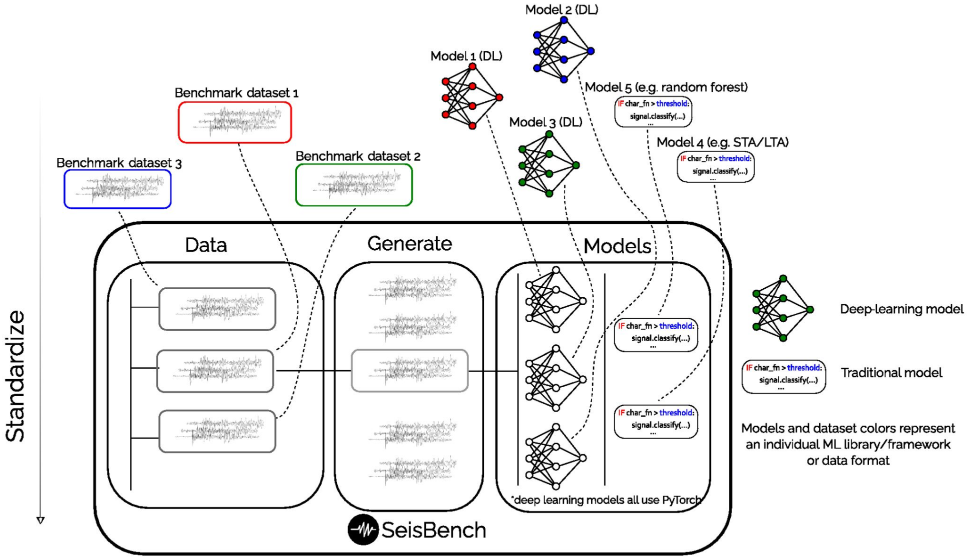


SeisBench

A toolbox for machine learning in seismology

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. SeisBench—A toolbox for machine learning in seismology. Seismological Society of America, 93(3), pp.1695-1709.

- A dataset: waveform hdf5 + .csv file with accompanying metadata, that both follow standard structure and naming scheme
- Metadata contains earthquake (e.g. magnitude), station (e.g. location), path (e.g. phase travel time), waveform (e.g. SNR), and other (e.g. train/val/test split) information
- Metadata easily extensible by adding new columns
- 6 datasets currently available, with a large amount of waveforms (e.g. 8.1 million in SCEDC) and accompanying metadata (e.g. 125 in INSTANCE) available
- Noise data often available



SeisBench

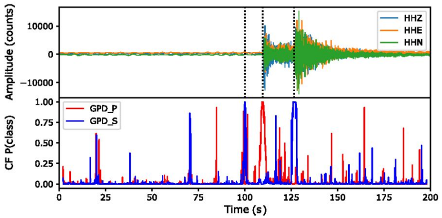
A toolbox for machine learning in seismology

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. SeisBench—A toolbox for machine learning in seismology. Seismological Society of America, 93(3), pp.1695-1709.

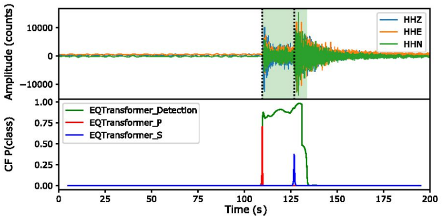
- SeisBench provides a wide range of tools to work with the datasets: filtering them according to the desired metadata characteristics, resampling the data, dealing with missing components, filtering the signals, normalize the signals, etc…
- It allows for a simple generation of a benchmark dataset in the SeisBench format from any data source
- Range of ML models available (e.g. for picking, DeepDenoiser) which can be easily applied to the desired data simply through SeisBench
- Re-training and transfer-learning of the models available on desired data

[1]:
```python
from obspy.clients.fdsn import Client
from obspy import UTCDateTime
import matplotlib.pyplot as plt


# Get a seismic stream
client = Client("GFZ")

t = UTCDateTime("2007/01/02 05:48:50")
st = client.get_waveforms(network="CX", station="PB01", location="*",
 channel="HH?", starttime=t-100, endtime=t+100)
```

[2]:
```python
from seisbench.models import GPD

# Load pre-trained weights into model
gpd_model = GPD.from_pretrained("original")

# Apply GPD to stream
gpd_annotations = gpd_model.annotate(st, stride=5)
gpd_picks = gpd_model.classify(st, P_threshold=0.95, S_threshold=0.95)
```



[2]:
```python
from seisbench.models import EQTransformer

# Load pre-trained weights into model
eqt_model = EQTransformer.from_pretrained("original")

# Apply EQT to stream
eqt_annotations = eqt_model.annotate(st)
eqt_picks = eqt_model.classify(st)
```



Earthquake monitoring

- Monitoring includes earthquake detection, phase picking, event association, event location and magnitude determination
- Currently, most of the steps are semi-automated and require trained analysts to manually review the automatic solutions, especially phase picking
- The number of seismic stations available is increasing rapidly, and requires more-and-more human resources to analyze
- The existing automated procedures require a great deal of parameter tuning
- ML approaches have potential to assist the human analysts and to increase the number of earthquakes detected
- Large number of labeled data available

Which Picker Fits My Data?

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth, 127(1), p.e2021JB023499.

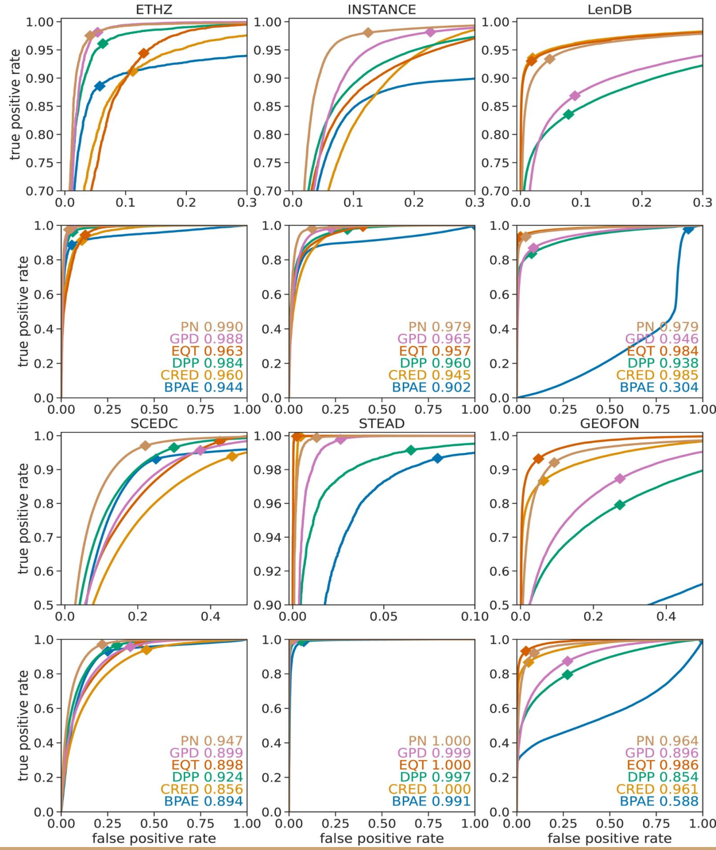
- A comprehensive benchmark of ML methods for earthquake detection, phase picking and phase identification
- 7 models (6 ML, 1 classical) are compared on 8 different datasets
- Datasets are of different sizes, come from different regions and contain local, regional, and teleseismic earthquake signals
- ML models differ in their architecture, their input type, their size etc.
- All ML models are trained on all the available datasets and tested on the same (in-domain) and the other datasets (cross-domain)

SeisBench
A toolbox for machine learning in seismology

EQ detection benchmarking

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth, 127(1), p.e2021JB023499.

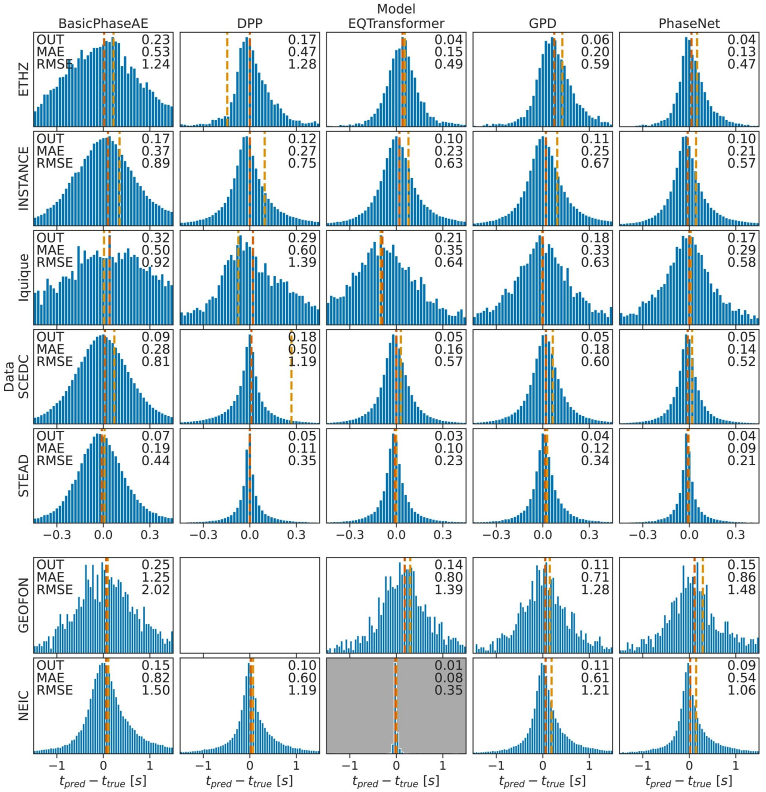
- Most of the models show good overall performance for in-domain earthquake detection, with the AUC ranging from 0.78-0.98
- Deeper models show more robustness and achieve good performance for all the datasets
- Cross-dataset applications more challenging - performance degrades when the datasets differ in their typical epicentral distances



Phase picking and identification

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth, 127(1), p.e2021JB023499.

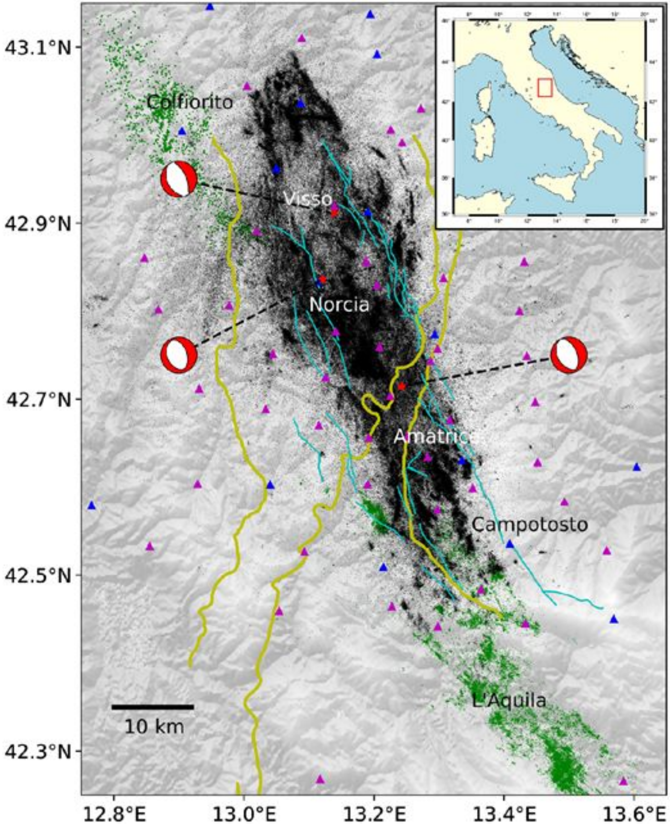
- Model again show good in-domain performance for phase identification, with the MCC (Matthew's correlation coefficient) ranging from 0.71-0.95, with higher variability between the models than for detection
- Cross-domain performance the best when the training dataset contains a variety of tectonic settings
- Phase picking results show similar trends with low picking timing residuals



ML Earthquake Catalog

Tan, Y.J., Waldhauser, F., Ellsworth, W.L., Zhang, M., Zhu, W., Michele, M., Chiaraluce, L., Beroza, G.C. and Segou, M., 2021. Machine‐learning‐based high‐resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central Italy sequence. The Seismic Record, 1(1), pp.11-19.

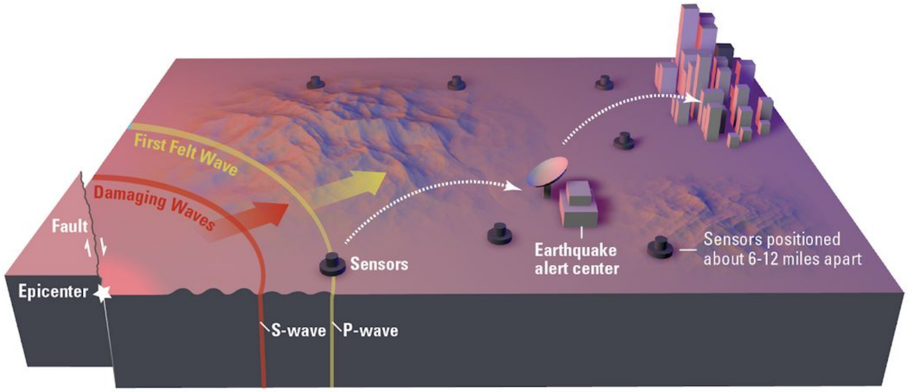
- A ML picker was used to analyze one year of data (2016-2017) on 139 stations
- Usage of the ML derived phase picks led to 900k earthquakes detected during that time-span, an order of magnitude improvement over the INGV catalog
- Detection improvement comes from detection of smaller earthquakes than those present in INGV catalog
- Detection rate shows diurnal variation
- The new catalog allows for a better understanding of the complex fault structures in the area



Earthquake Early Warning

- Seismic waves are far slower than the electromagnetic waves used in modern communication systems to transfer information
- The data from the stations close to the epicenter is used to provide warning to locations further away

- Classic EEW algorithms estimate the earthquake location and size using the first few seconds of recorded signal
- More seconds used means better estimate, but loss in time
- The estimates of the earthquake properties are used to predict ground motion at the locations of interest

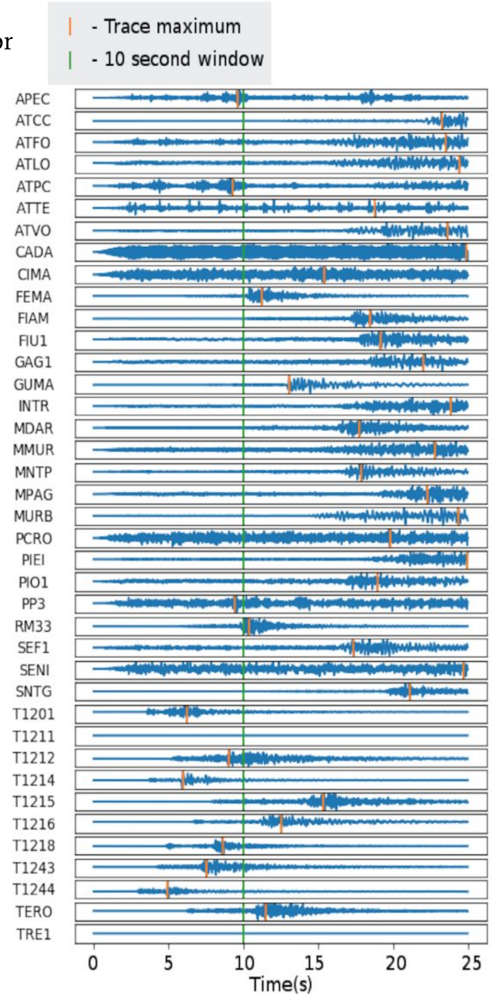


Picture from earthquake.ca.gov

ML for EEW

Jozinović, D., Lomax, A., Štajduhar, I. and Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural networ Geophysical Journal International, 222(2), pp.1379-1389.

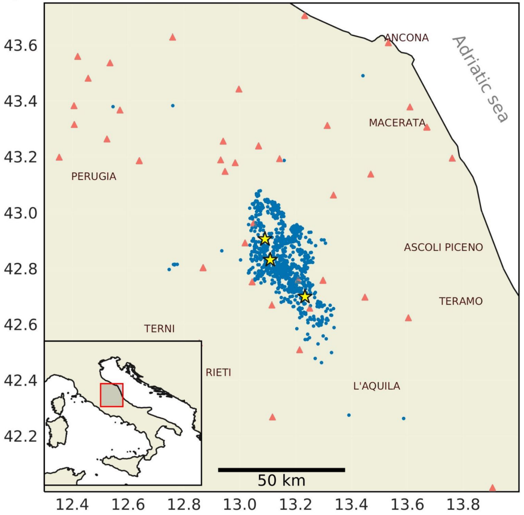
- Use the first 10 seconds of waveforms from 39 stations to predict the peak ground motion (PGM) expected at those 39 stations
- PGM could already be reached at the stations closest to the epicenter, but the algorithm will provide predictions (warnings) to the stations further away
- A direct measure of estimating PGMs avoiding the uncertainties of estimating the earthquake source parameters which are then used to predict PGMs (more uncertainty)
- The ML model: a CNN that looks at the spatio-temporal pattern of ground motions across the network
- Results show better performance than the classical algorithms
- Longer time windows lead to better accuracy

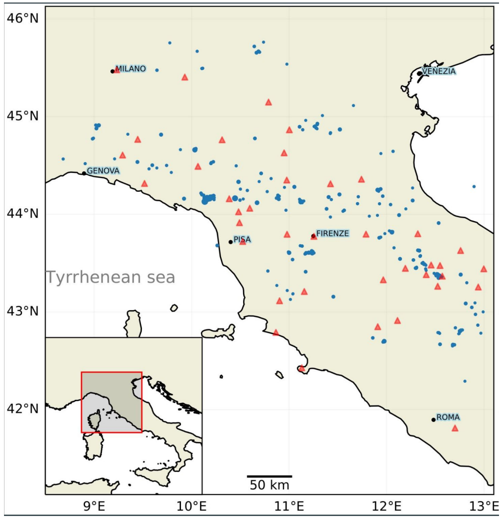


Jozinović, D., Lomax, A., Štajduhar, I. and Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. Geophysical Journal International, 222(2), pp.1379-1389.

Applying it to a new area

- The initial algorithm trained in an area with high density of seismic stations and large earthquake catalog
- How to apply it in a new area that is more challenging in terms of seismic network geometry and data availability?
- The new area - central-western Italy centered in Pisa with smaller number of earthquakes available

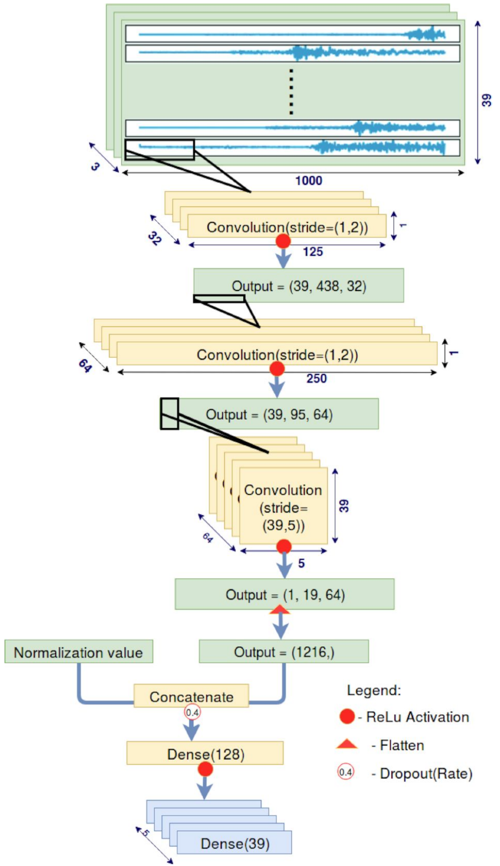




Transfer learning (TL)

Jozinović, D., Lomax, A., Štajduhar, I. and Michelini, A., 2022. Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophysical Journal International, 229(1), pp.704-718.

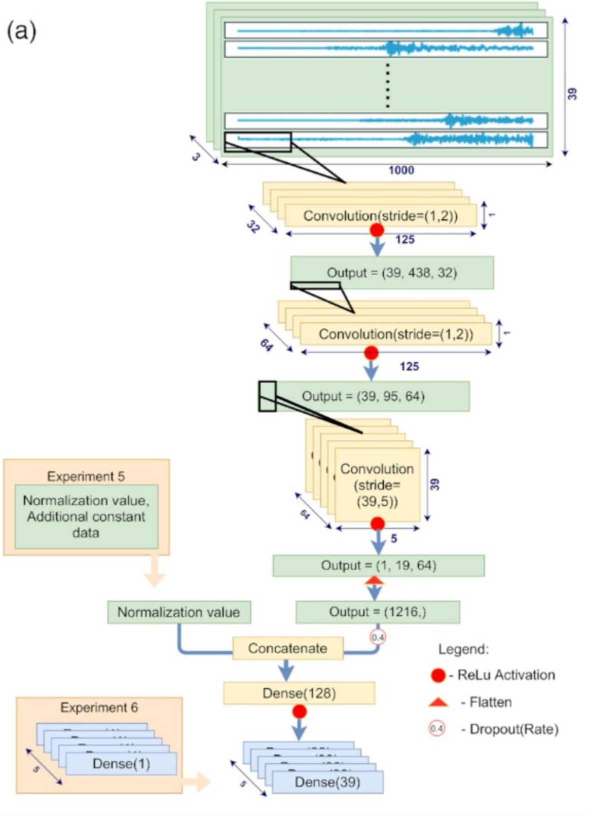
- Re-training of the CNN on the new smaller dataset leads to worse results
- The first two Convolutional layers are single-station layers (kernel height 1), only the third layer is looking at the cross-station pattern
- Usage of transfer learning: the first two layers were pre-trained on the bigger dataset, and only the third and later layers were re-trained (the learning rate of the first two layers set to zero)
- Significant improvement in the results as a consequence of TL
- The results show that the method has significant warning potential for the location of seismic station IV.PII (10 km from VIRGO)



Inclusion of station information

Jozinović, D., Lomax, A., Štajduhar, I. and Michelini, A., 2022. Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophysical Journal International, 229(1), pp.704-718.

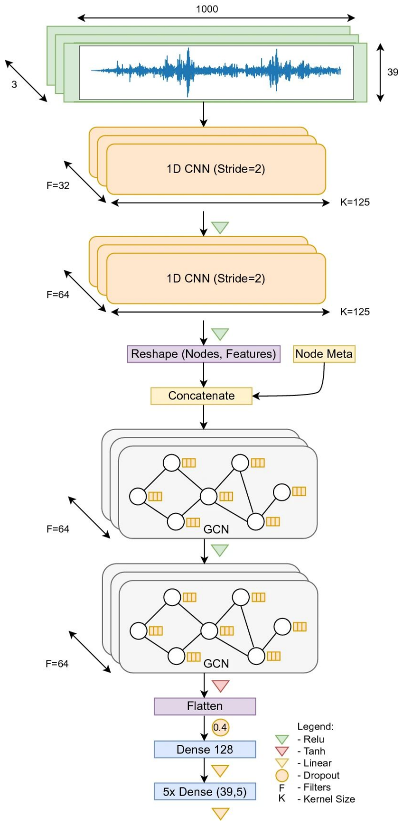
- The CNN is learning the spatial information about the stations implicitly during training
- To help the CNN we add inter-station distances and azimuths (sine and cosine of Az) as additional input to the fully-connected layer at the end
- The addition leads to significant improvement of the results
- To better model the spatial information we use graph convolutional network (GCN) - the GCN layers replace the third convolutional layer and the spatial information is encoded through the adjacency matrix
- The use of GCN improves results for both datasets (on average 16% MSE reduction)
- It allows to achieve the same results with shorter input time windows, thus providing more warning time



Inclusion of station information

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., & Atzmueller, M. (2022). Graph neural networks for multivariate time series regression with application to seismic data. International Journal of Data Science and Analytics, 1-16.

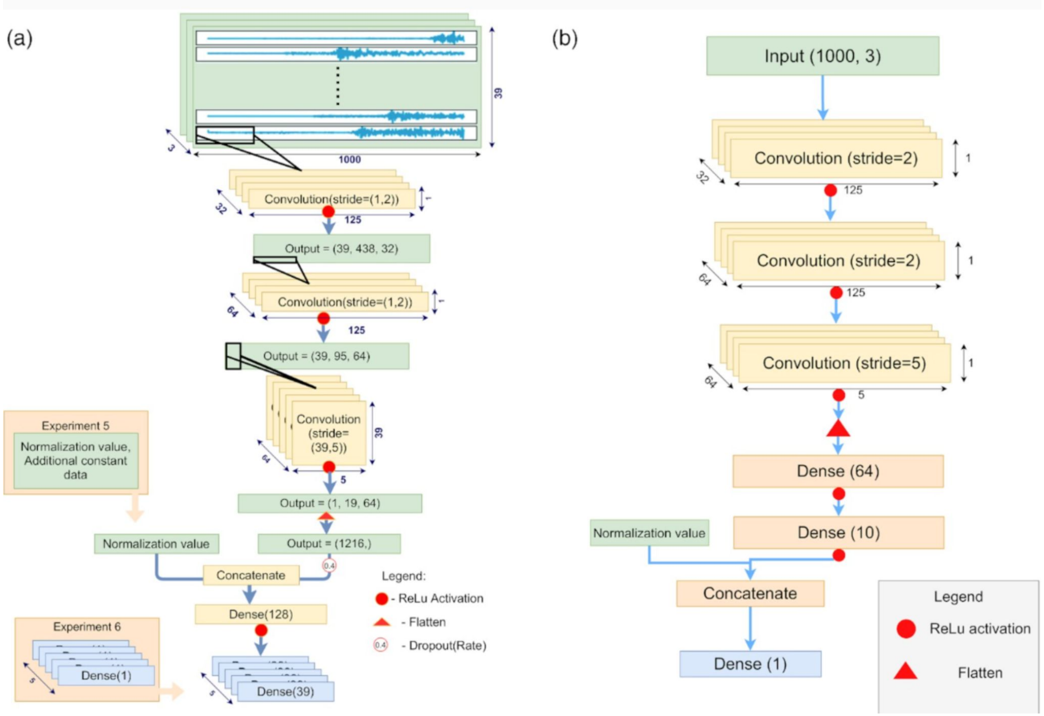
- The CNN is learning the spatial information about the stations implicitly during training
- To help the CNN we add inter-station distances and azimuths (sine and cosine of Az) as additional input to the fully-connected layer at the end
- The addition leads to significant improvement of the results
- To better model the spatial information we use graph convolutional network (GCN) - the GCN layers replace the third convolutional layer and the spatial information is encoded through the adjacency matrix
- The use of GCN improves results for both datasets (on average 16% MSE reduction)
- It allows to achieve the same results with shorter input time windows, thus providing more warning time

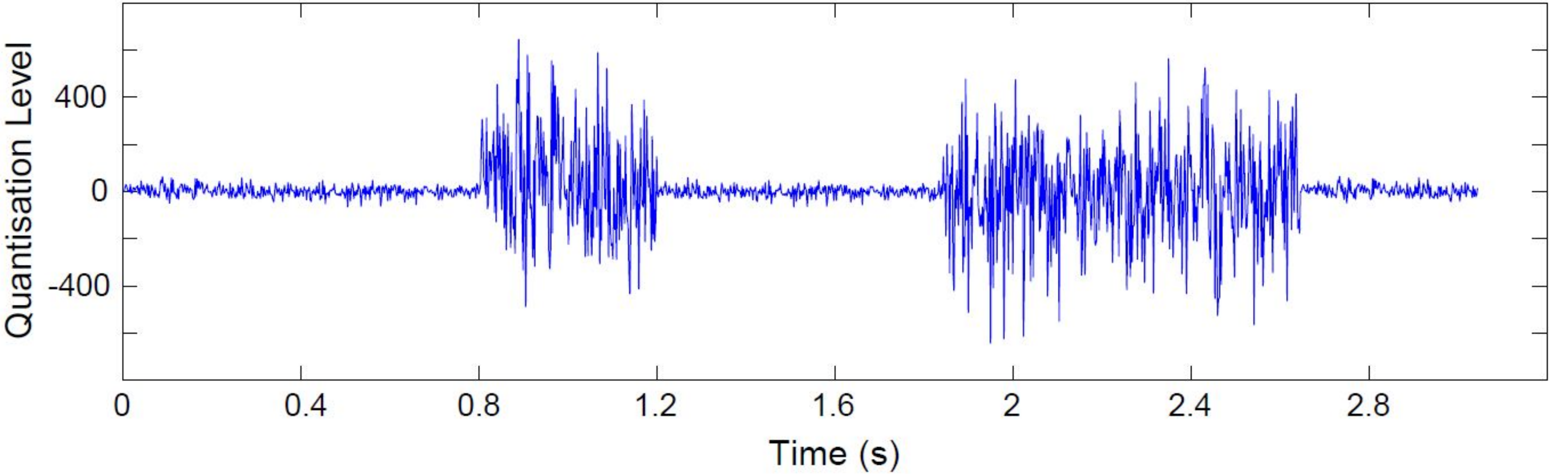


TL from a different problem

Jozinović, D., Lomax, A., Štajduhar, I. and Michelini, A., 2022. Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophysical Journal International, 229(1), pp.704-718.

- The first two layers are extracting features on a low-level, only the later layers combine these features for the task needed
- A new CNN model for a different problem (magnitude determination) is designed as to have identical first two layers as the EEW CNN model
- Doing TL from the magnitude determination problem was successful
- First layer learning rate set to zero, second layer used pre-trained weights but trainable

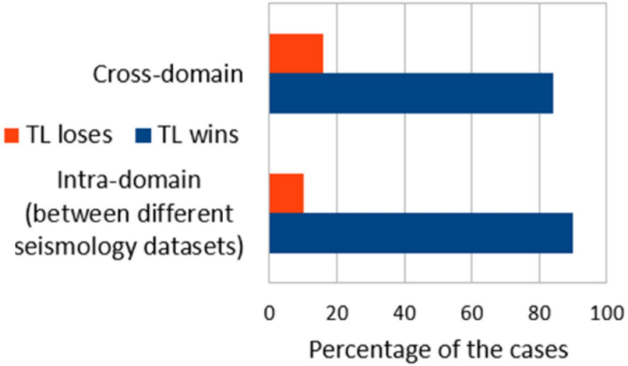




IMG source: https://hackaday.io/project/113338-publys-an-open-source-biosensing-board/log/143756-emg-sensor

Otović, E., Njirjak, M., Jozinović, D., Mauša, G., Michelini, A. and Štajduhar, I., 2022. Intra-domain and cross-domain transfer learning for time series data—How transferable are the features?. Knowledge-Based Systems, 239, p.107976.

TL for time-series



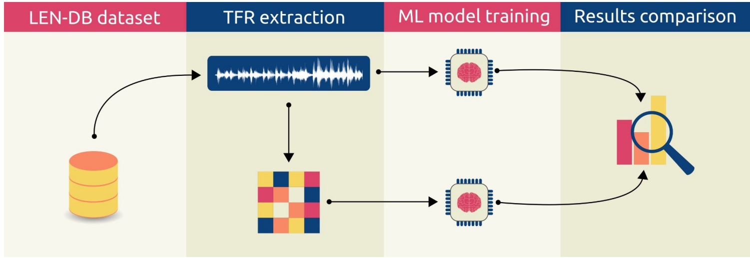
- The paper tested TL between time series data coming from the same (3 seismological datasets) and different domains (3 seismological datasets, 1 speech, 1 EMG, and 1 stock market dataset)
- The seismological datasets differed in their sampling rate, signal epicentral distances, units
- Two training sets were used by randomly selecting 1500 and 9000 waveforms from the training set (test and validation set are always the same)
- Tests were performed using 4 different ML models to reduce the influence of the ML model selection on the conclusions
- Both accuracy (or similar metric) and convergence rate were observed to quantify the benefits of TL
- Conclusion: TL can be successfully used inside and between different time-series domains

	Target datasets			
	SPEECH 1k5	SPEECH 9k	EMG 1k5	EMG 9k
LOMAX	45.7%	871%	24.2%	17.6%
LEN-DB	309%	1210%	21.9%	12.1%
STEAD	323%	1080%	18.7%	10%
SPEECH	-	-	13.5%	9.87%
EMG	367%	1820%	–	–
S&P 500	122%	885%	15.2%	11.9%

Time-frequency representations for ML

Njirjak, M., Otović, E., Jozinović, D., Lerga, J., Mauša, G., Michelini, A. and Štajduhar, I., 2022. The Choice of Time–Frequency Representations of Non-Stationary Signals Affects Machine Learning Model Accuracy: A Case Study on Earthquake Detection from LEN-DB Data. Mathematics, 10(6), p.965.

- Does using different time-frequency representations (TFR) of the same data (non-stationary signals) affect the ML results on those data?
- 150k samples from the LEN-DB seismic (earthquake and noise) dataset, was used for the goal of earthquake detection
- The experiment: use 3 different state-of-the-art CNNs (VGG, AlexNet, ResNet) and train them on 9 different TFRs
- The base model for comparison is a successful CNN for Earthquake detection that takes waveforms in the time-domain as inputs
- Conclusion: TFR of the data can have a statistically significant influence on the ML results



	VGG16	ResNet50	AlexNet
BJ	0	0.01	0.323
BUD	0.042	0	0
CW	0	0	0.007
MH	0	0	0
PWV	0	0	0
RIDB	0	0	0.016
SP	0	0.002	0
SPWV	0	0.014	0
WV	0	0	0

Thank you! Questions?