Intra-domain and cross-domain transfer learning for time series data—How transferable are the features?

Valencia, 13. April 2022

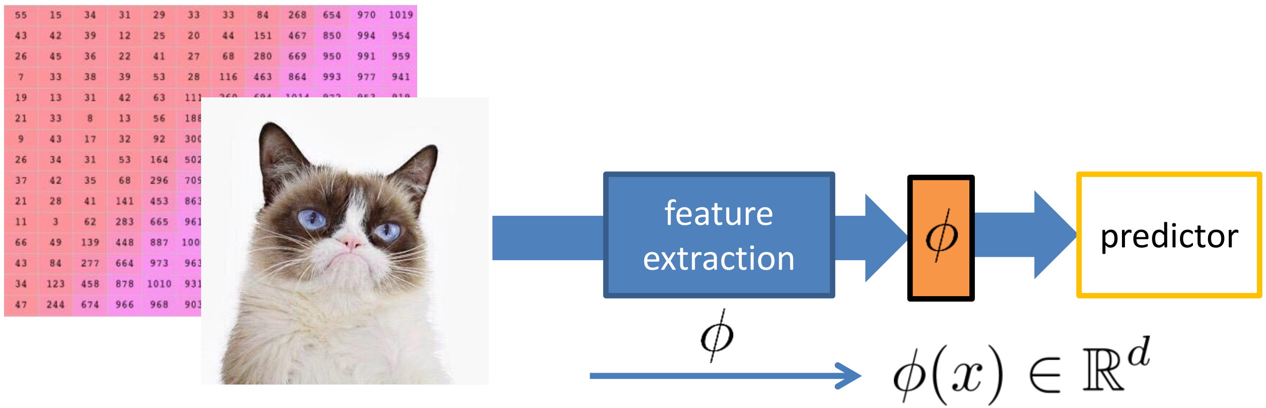
E. Otović, M. Njirjak, D. Jozinović, G. Mauša, A. Michelini, I. Štajduhar

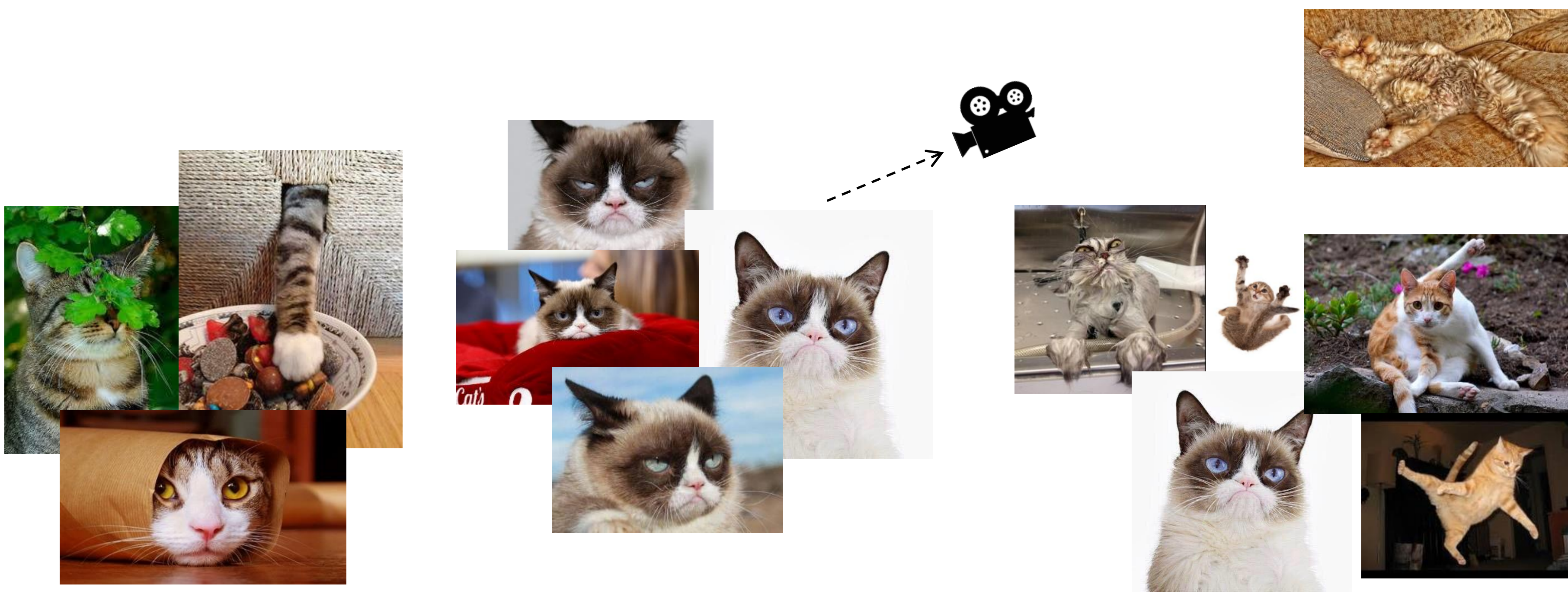
istajduh@riteh.hr
www.riteh.uniri.hr
airi.uniri.hr

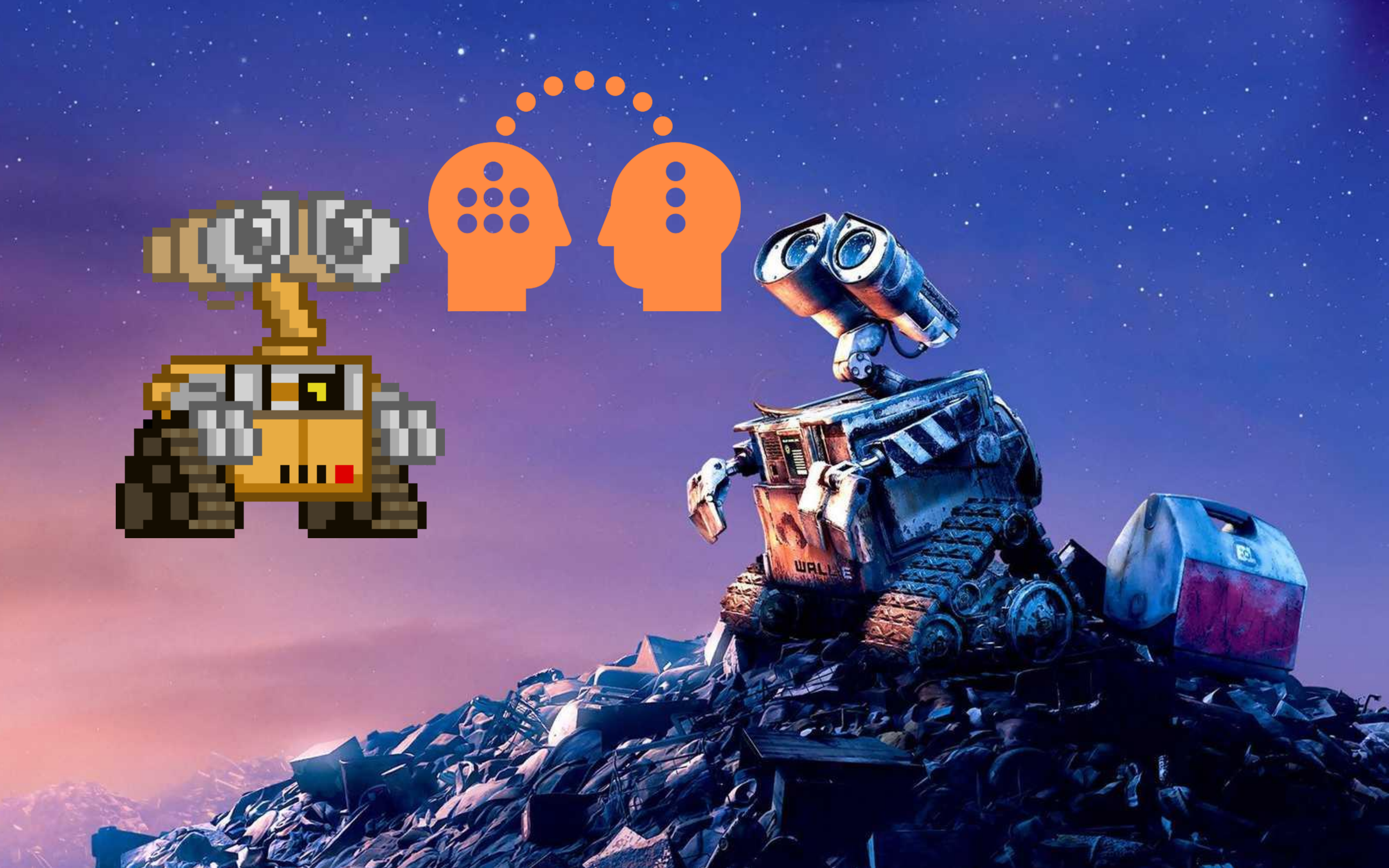
AIRI
Center for Artificial Intelligence and Cybersecurity

Introduction

- Deep learning models require large datasets for successful training.
  - Sometimes it is very hard or not possible to collect enough data.
  - This can be an obstacle to the successful application of deep learning methods in some fields.

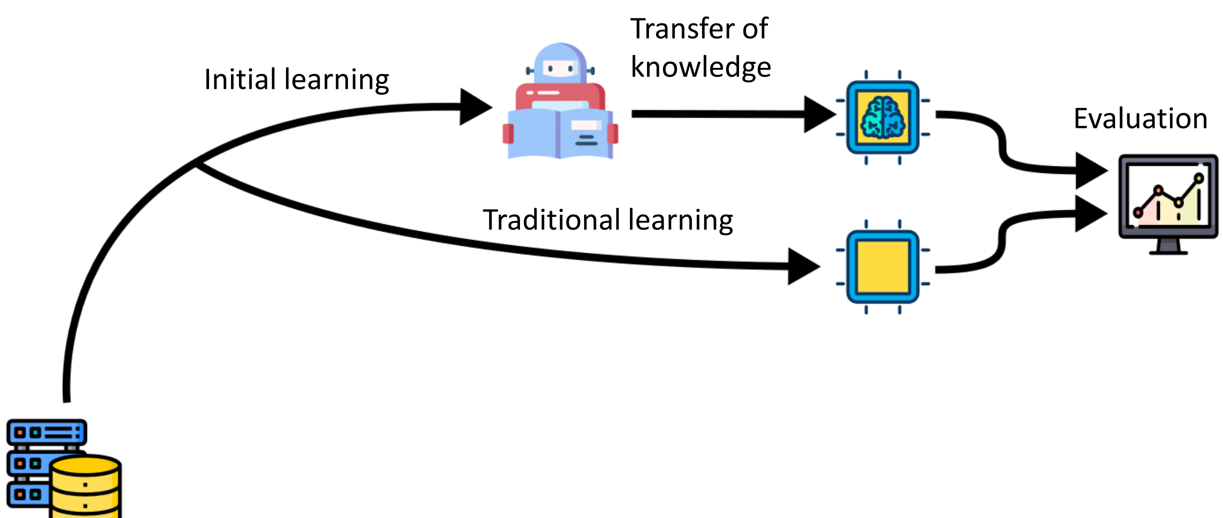

feature extraction
ϕ
predictor
ϕ
$\phi(x) \in \mathbb{R}^d$





Introduction

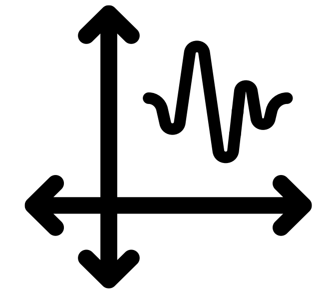
- Transfer learning (TL)
  - One possible way to overcome the problem of small datasets.
  - Transfer knowledge gained on one task to the target task. Therefore, successful training is possible even with a small target dataset.
  - May as well speed up the convergence of the model.



Research questions

- How transferable is the knowledge between domains?
  - Intra-domain: between different seismology datasets.
  - Cross-domain: between seismology, audio, medicine and finance datasets.
- How does TL affect the convergence of the model?
- How does target dataset size affect the performance of TL?
  - We use target datasets with 1,500 and 9,000 training instances. These are obtained by reducing the source datasets.
- How does model choice impact the performance of TL?
  - We use two seismology-specific models and two general purpose time-series models.

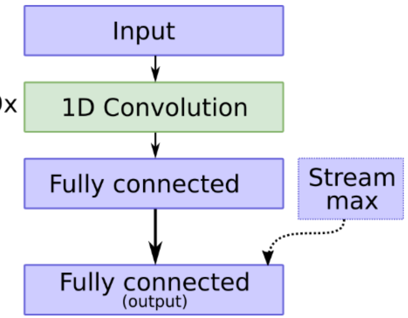
Experiment

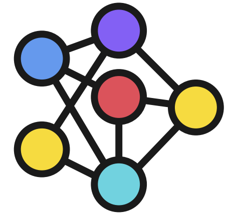


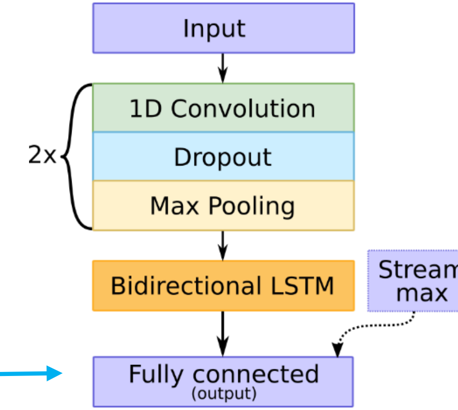
- The input to the model is a raw waveform.
  - Convolutional layers learn to extract features from raw waveform data.
  - Already trained convolutional layers can be transferred to some other task.
- Examine transfer of knowledge between all pairs of domains to find compatible pairs.
- Train referent model only on the target dataset and compare it to the TL model that was pre-trained on the source dataset, and fine-tuned on the target dataset.

Experiment

- Perform grid search to find optimal hyperparameters for TL.
- Rerun experiment multiple times so statistical tests can be applied.
- Summary:
  - 4 models
  - 6 source datasets
  - 6 target datasets having 1,500 training instances
  - 6 target datasets having 9,000 training instances
  - Totally 60 pairs of domains to examine (cases in which target dataset is a subset of source dataset are ignored)

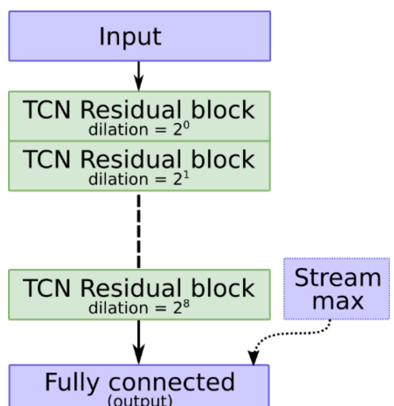


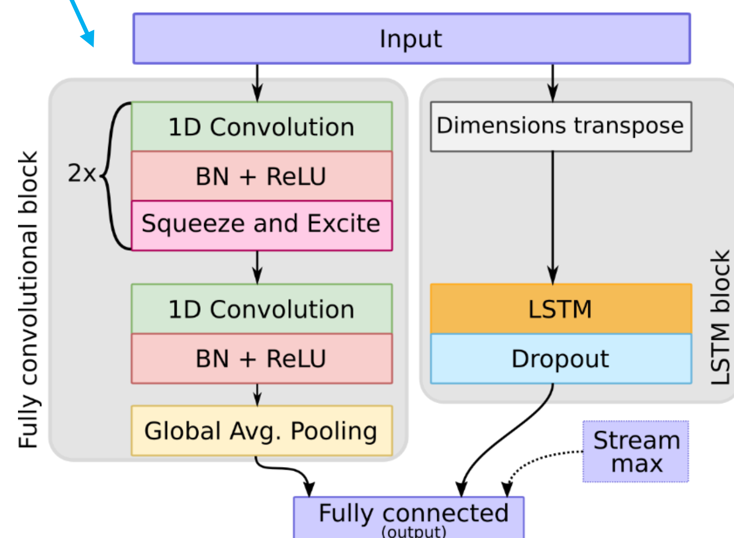




- ConvNetQuake INGV
- MagNet
- MLSTM FCN
- TCN

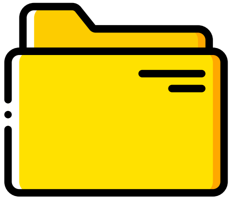
- LOMAX
- LEN-DB
- STEAD
- SPEECH
- EMG
- S&P 500





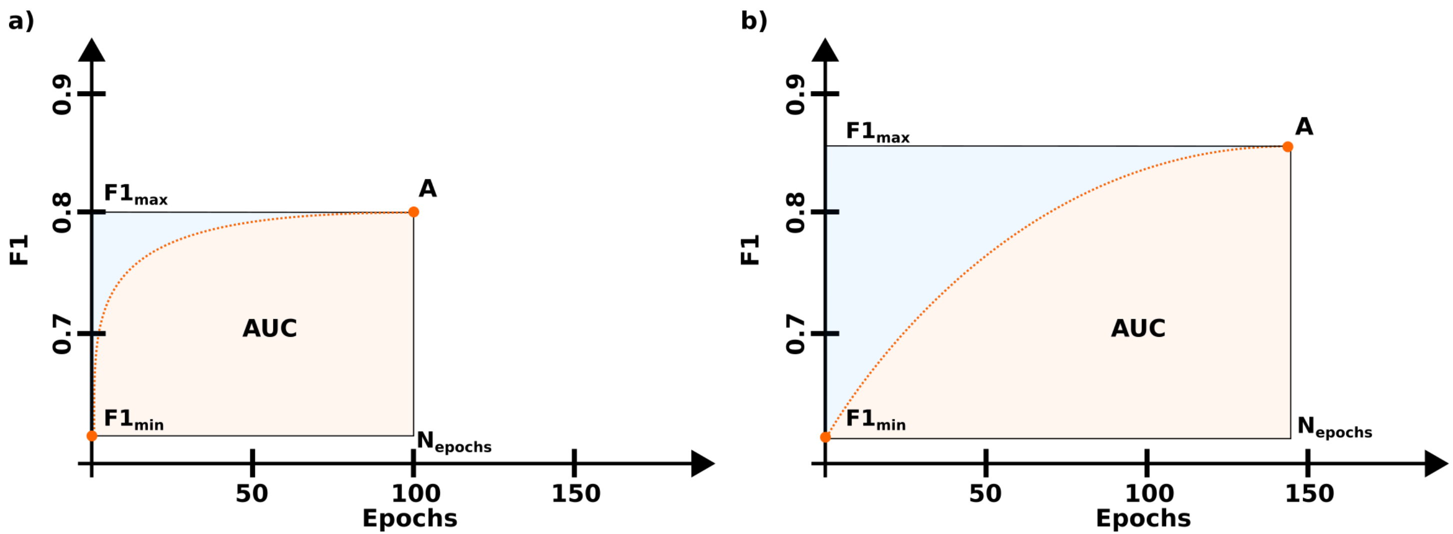
Experiment

Dataset	Task type	Domain	Size	Channels	Sampling frequency	Sampling points per channel	Waveform duration
LOMAX [9]	Regression	Seismology	19,426	3	20 Hz	1,001	50 s
LEN-DB [11]	Regression	Seismology	629,096	3	20 Hz	540	27 s
STEAD [13]	Regression	Seismology	1,031,908	3	100 Hz	6,000	60 s
SPEECH [14]	Classification	Audio	105,829	1	16 kHz (resampled to 8 kHz)	16,000 (8,000 after resampling)	1 s
EMG [15]	Classification	Medicine	32,438	3	200 Hz	80	0.4 s
S&P 500 [19,20]	Regression	Finances	22,681	1	Once every working day	50	50 days



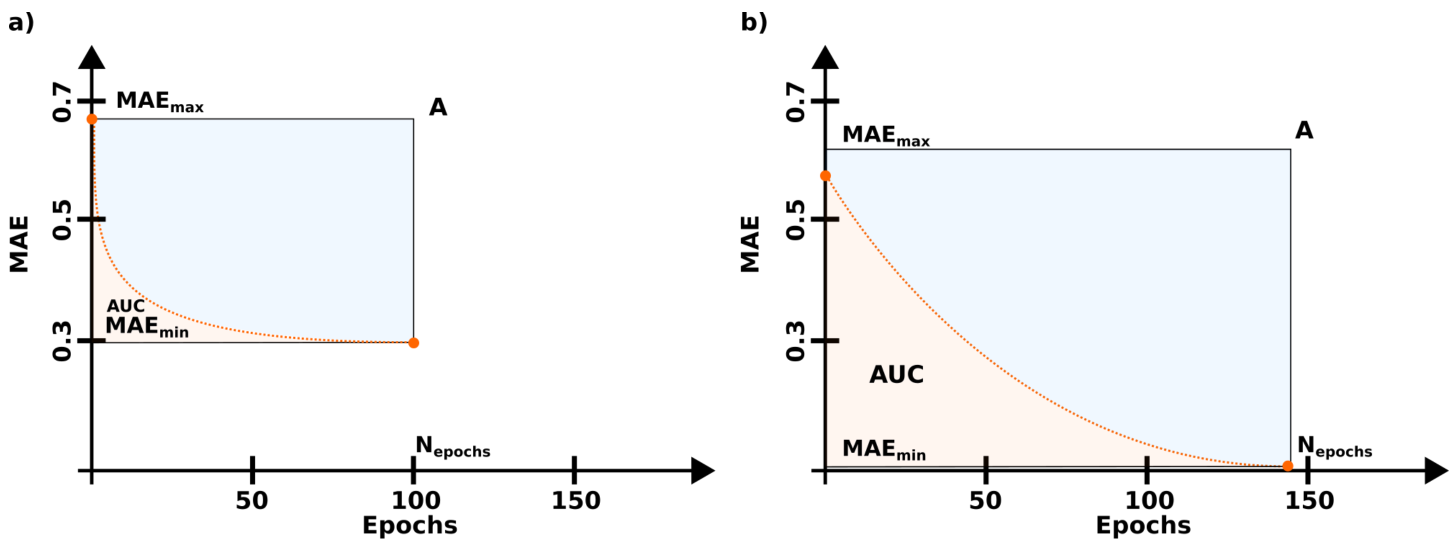
Evaluation metrics - classification

- Accuracy: weighted F1
- Convergence rate

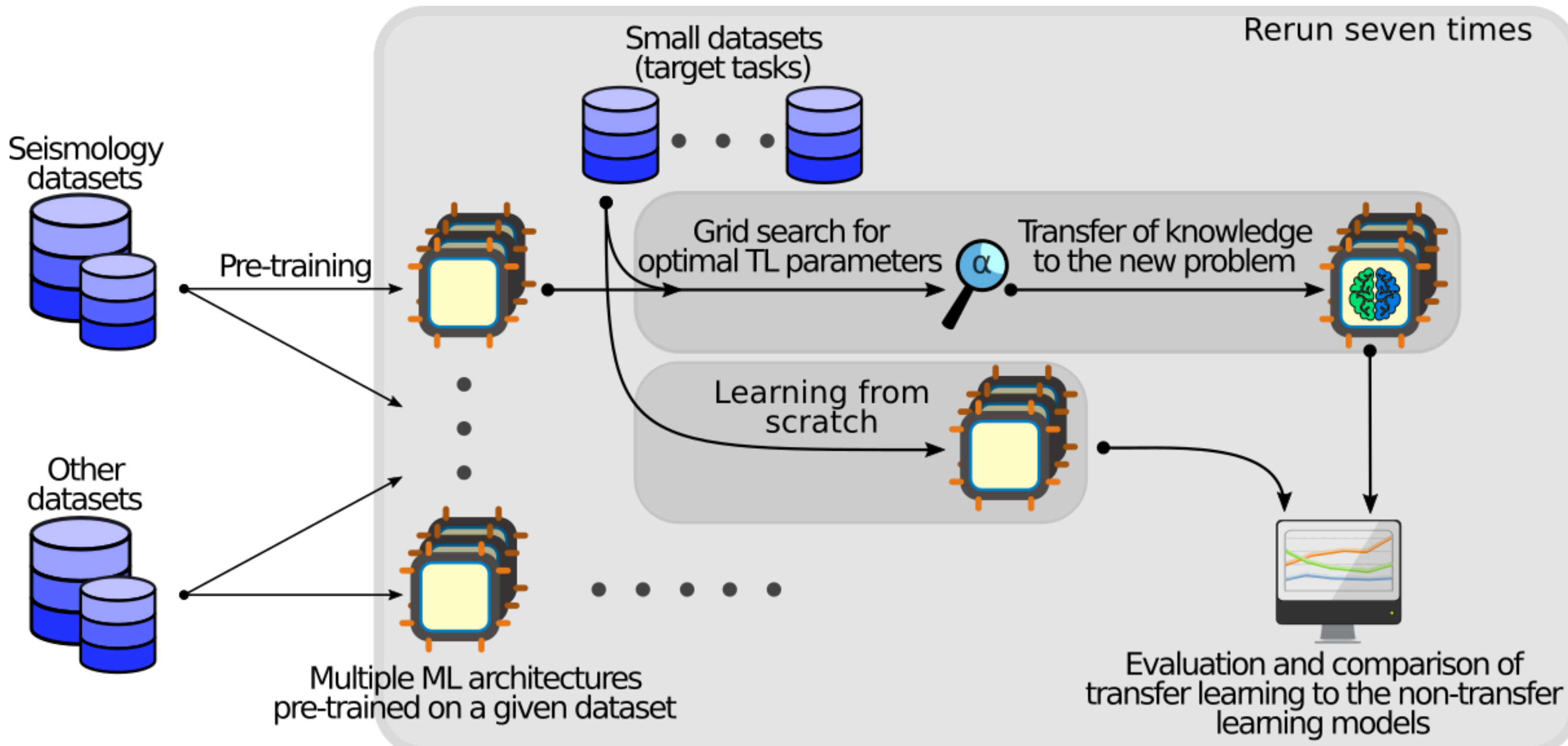


Evaluation metrics - regression

- Accuracy: mean absolute error (MAE)
- Convergence rate

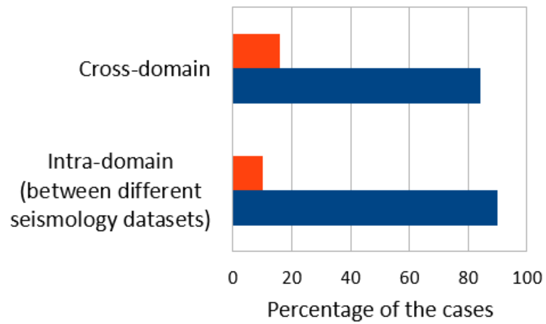


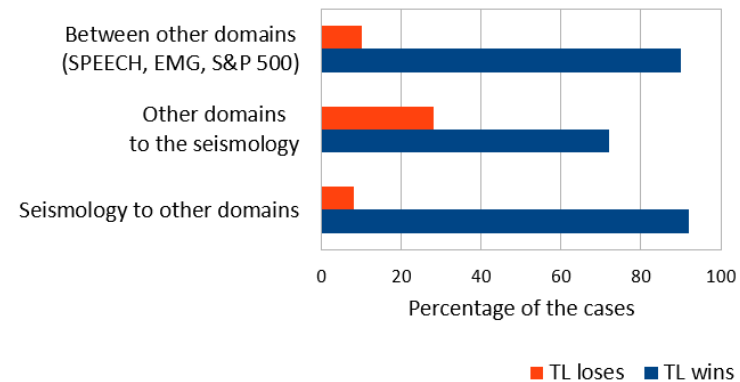
Experiment workflow



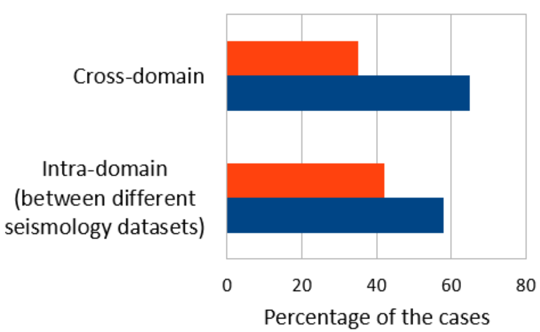
Results

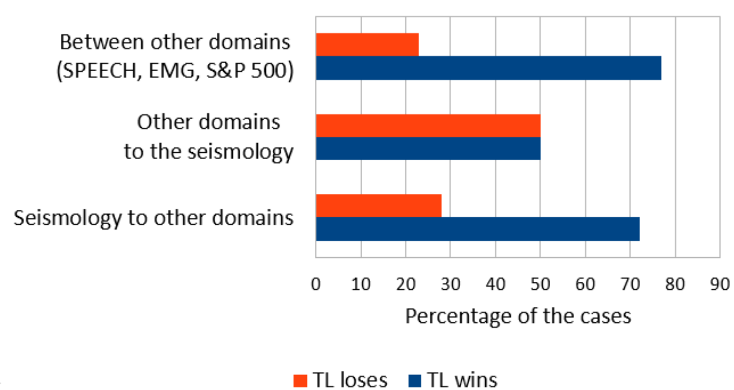
Accuracy:



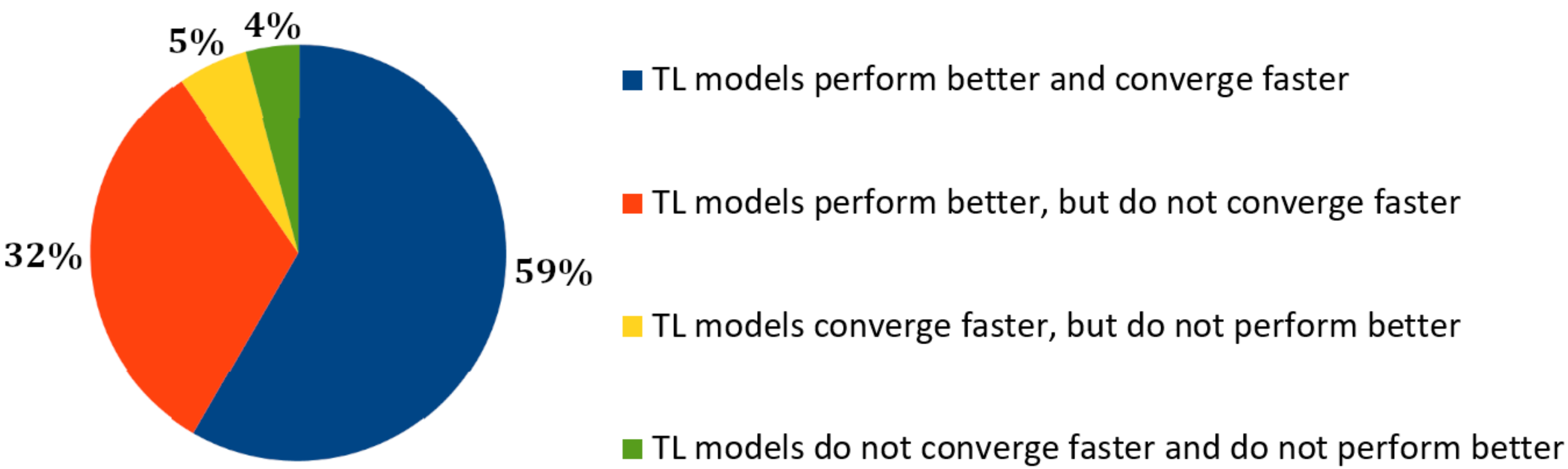


Convergence rate:





Results


4%
5%
32%
59%
TL models perform better and converge faster
TL models perform better, but do not converge faster
TL models converge faster, but do not perform better
TL models do not converge faster and do not perform better

Results – accuracy

	Target datasets					
	LOMAX 1k5	LOMAX 9k	LEN-DB 1k5	LEN-DB 9k	STEAD 1k5	STEAD 9k
LOMAX	–	–	**8.21%**	**3.17%**	**6.63%**	**3.96%**
LEN-DB	**8.25%**	**4.01%**	–	–	**2.49%**	**2.52%**
STEAD	1.16%	−1.87%	0.56%	−0.52%	–	–
SPEECH	**9.29%**	**3.58%**	**4.47%**	**3.19%**	**−2.31%**	0.07%
EMG	**9.12%**	**5.34%**	−0.06%	0.41%	1.15%	**2.04%**
S&P 500	0.12%	0.12%	−0.92%	1.14%	**−7.23%**	**−2.29%**

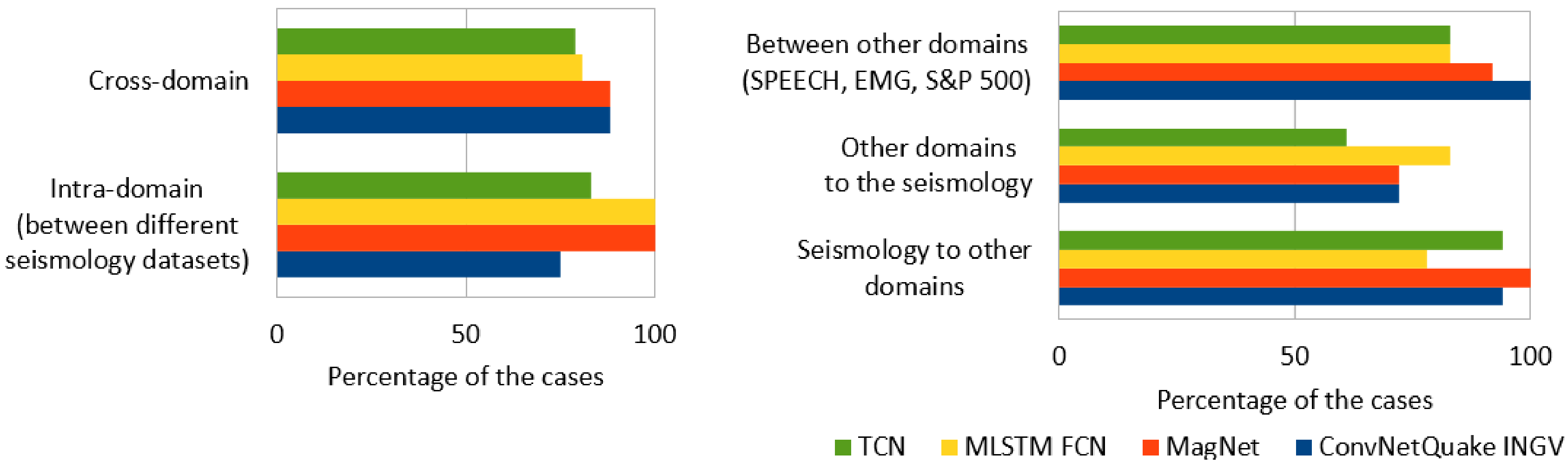
	Target datasets					
	SPEECH 1k5	SPEECH 9k	EMG 1k5	EMG 9k	S&P 500 1k5	S&P 500 9k
LOMAX	**45.7%**	**871%**	**24.2%**	**17.6%**	**45.4%**	**22.8%**
LEN-DB	**309%**	**1210%**	**21.9%**	**12.1%**	**44.4%**	**22.1%**
STEAD	**323%**	**1080%**	**18.7%**	**10%**	**44.3%**	**22.7%**
SPEECH	-	-	**13.5%**	9.87%	**44.1%**	**21.6%**
EMG	**367%**	**1820%**	–	–	**41.8%**	**18.2%**
S&P 500	**122%**	**885%**	**15.2%**	**11.9%**	-	-

Results – convergence rate

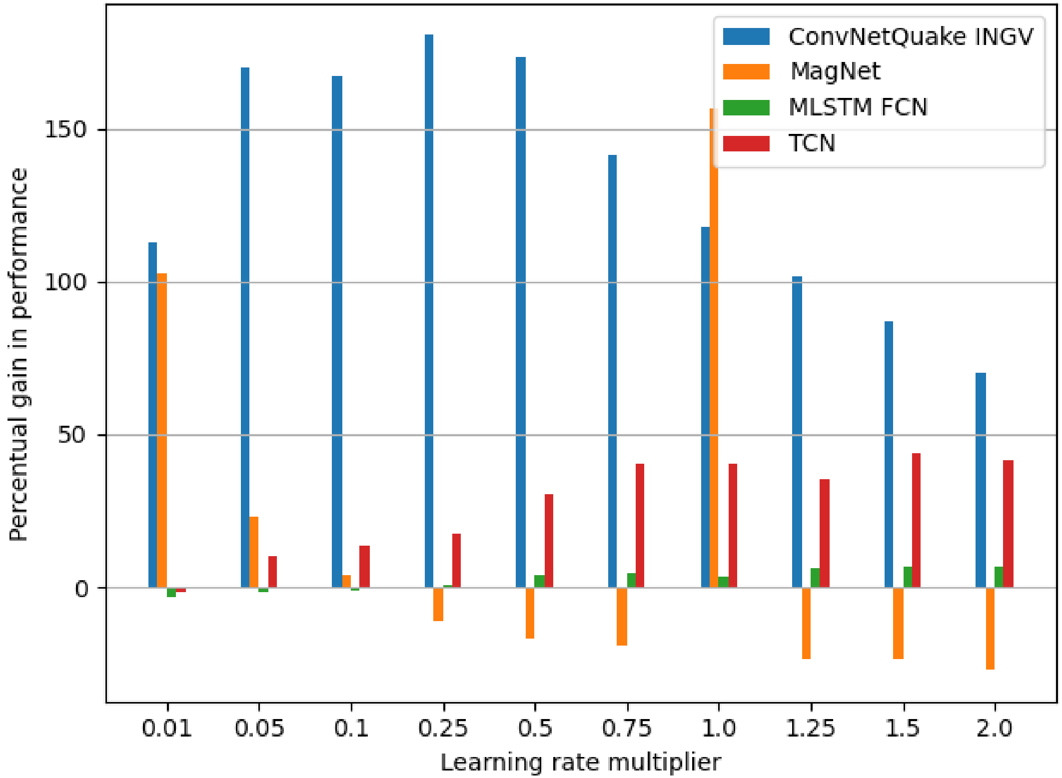
	Target datasets					
	LOMAX 1k5	LOMAX 9k	LEN-DB 1k5	LEN-DB 9k	STEAD 1k5	STEAD 9k
LOMAX	–	–	9.08%	−3.15%	16.6%	11.6%
LEN-DB	**11.6%**	6.72%	–	–	17.5%	11.8%
STEAD	**16.5%**	4.59%	**22.6%**	7.59%	–	–
SPEECH	10.2%	3.19%	**5.45%**	1.35%	6.73%	8.71%
EMG	10.1%	5.76%	14.9%	−0.31%	8.98%	12.8%
S&P 500	4.04%	4.33%	13.8%	−1.39%	**9.76%**	13.9%

	Target datasets					
	SPEECH 1k5	SPEECH 9k	EMG 1k5	EMG 9k	S&P 500 1k5	S&P 500 9k
LOMAX	0.12%	4.55%	13.3%	−1.89%	6.37%	5.39%
LEN-DB	1.13%	7.87%	17.0%	−3.67%	6.43%	3.66%
STEAD	2.35%	3.4%	17.3%	-7.19%	6.16%	9.7%
SPEECH	-	-	13.9%	−5.11%	6.13%	3.13%
EMG	5.76%	9.35%	–	–	6.2%	8.25%
S&P 500	4.8%	5.75%	8.12%	-6.09%	-	-

Results – performance gain



Results – learning rate multiplier



Conclusions

- We found out that TL is very likely to get a better performance score, or at least as good as the model trained from scratch. It is very unlikely to perform worse than models trained from scratch.
- Even seemingly unrelated domains can be mutually compatible enough to yield positive effects.
- TL enabled models to converge on small and difficult datasets, while traditionally trained models could not converge.
- All models had approximately the same probability of achieving better results with TL.

Main message

- Transfer learning is very likely to either result in positive or nonnegative effects.
- The hyperparameters for optimal transfer learning depend a lot on the chosen model.
- A model pre-trained on unrelated task can be better than a randomly initialized model.

- Should work on GW data too.

More info

- https://www.sciencedirect.com/science/article/abs/pii/S0950705121010984
- https://arxiv.org/abs/2201.04449
- https://github.com/ecokeco/tstl

Knowledge-Based Systems 239 (2022) 107976

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Intra-domain and cross-domain transfer learning for time series data—How transferable are the features?

Erik Otović [a], Marko Njirjak [a], Dario Jozinović [b,c], Goran Mauša [a,d], Alberto Michelini [b], Ivan Štajduhar [a,d,*]

[a] Department of Computer Engineering, Faculty of Engineering, University of Rijeka, Rijeka, Croatia
[b] Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy
[c] Department of Science, Roma Tre University, Rome, Italy
[d] Center for Artificial Intelligence and Cybersecurity, University of Rijeka, Rijeka, Croatia

ARTICLE INFO

Article history:
Received 22 August 2021
Received in revised form 7 November 2021
Accepted 13 December 2021
Available online 24 December 2021

Keywords:
Machine learning
Transfer learning
Time series
Fine-tuning
Convolutional neural networks

ABSTRACT

In practice, it is very challenging and sometimes impossible to collect datasets of labelled data large enough to successfully train a machine learning model, and one possible solution to this problem is using transfer learning. In this study, we investigate how transferable are features between different domains of time series data and under what conditions. The effects of transfer learning are observed in terms of the predictive performance of the models and their convergence rate during training. In our experiment, we used reduced datasets of 1500 and 9000 data instances to mimic real-world conditions. We trained two sets of models (four different architectures) on the reduced datasets: those trained with transfer learning and those trained from scratch. Knowledge transfer was performed both within the same application domain (seismology) and between different application domains (seismology, speech, medicine, finance). We observed the prediction performance of the models and their training convergence rate. We repeated the experiments seven times and applied statistical tests to confirm the validity of the results. The overall conclusion of our study is that transfer learning is highly likely to either increase or not negatively affect the model's predictive performance or its training convergence rate. We discuss which source and target domains are compatible for knowledge transfer. We also discuss the effect of the target dataset size and the choice of the model and its hyperparameters on transfer learning.

1. Introduction

In recent years, deep learning techniques have become increasingly popular and have brought new and exciting challenges. One of the major challenges and obstacles in training deep neural networks is the need for datasets that contain sufficient amounts of training instances. Creating such datasets is generally time-consuming, which can slow down the application of deep learning in some domains. For example, this may be the case when it is difficult to collect additional data instances because the observed phenomenon is very rare, or labelling instances for supervised learning is time-consuming because it must be done manually. Transfer learning (TL) is one of the possible approaches to combat these problems.

TL allows a machine learning (ML) model trained to solve one problem to be adapted or fine-tuned to solve another problem. In this way, some of the knowledge contained within the model from the first task is used to solve the second task. Knowledge transfer reduces the number of training instances required to solve another task compared to training with randomly initialised models, reduces training time, and leads to better accuracy. One of the domains where this approach has proven useful is image classification. There are several state-of-the-art models (such as VGG or Inception) pretrained on large image datasets that can be fine-tuned to solve other problems with a much smaller dataset and in much less time (see Review [1]). In this context, TL has enabled the application of these architectures to problems where they could not otherwise be (successfully) applied due to the small amount of training data or due to computationally intensive or lengthy computational operations involved in training models using large datasets.

1.1. Related work

In recent years, some research papers have been published reporting the application of TL for time series (TS) classification

* Corresponding author at: Department of Computer Engineering, Faculty of Engineering, University of Rijeka, Rijeka, Croatia.
E-mail address: istajduh@riteh.hr (I. Štajduhar).

https://doi.org/10.1016/j.knosys.2021.107976

Intra-domain and cross-domain transfer learning for time series data—How transferable are the features?

Valencia, 13. April 2022

E. Otović, M. Njirjak, D. Jozinović, G. Mauša, A. Michelini, I. Štajduhar

istajduh@riteh.hr
www.riteh.uniri.hr
airi.uniri.hr

AIRI
Center for Artificial Intelligence and Cybersecurity