Towards Constructing Generalized Foundation Models For Time Series Classification

Ali Ismail-Fawaz

Université de Haute Alsacem MSD IRIMAS, Mulhouse France

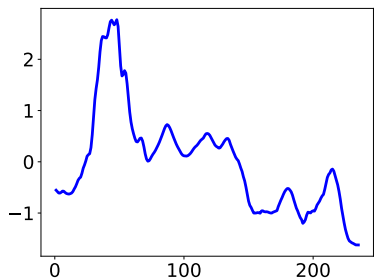
16 May 2023

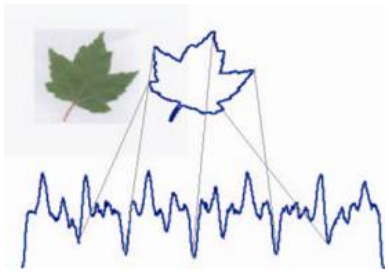
Outline

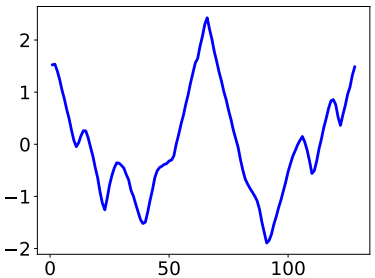
Contents

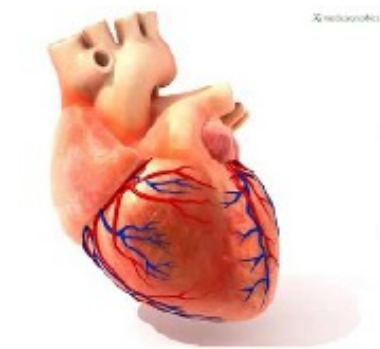
Time series are ubiquitous

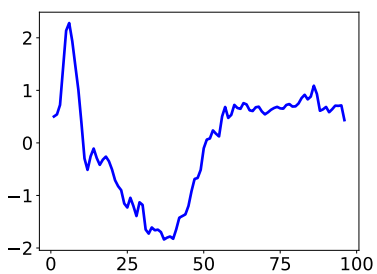
- Type of data present in numerous applicative domains
- Allow to study the evolution in time of a process, a behavior, etc.

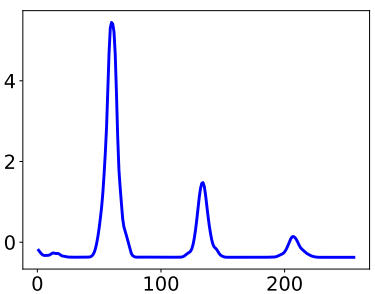








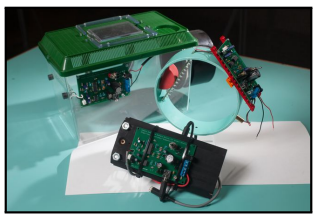


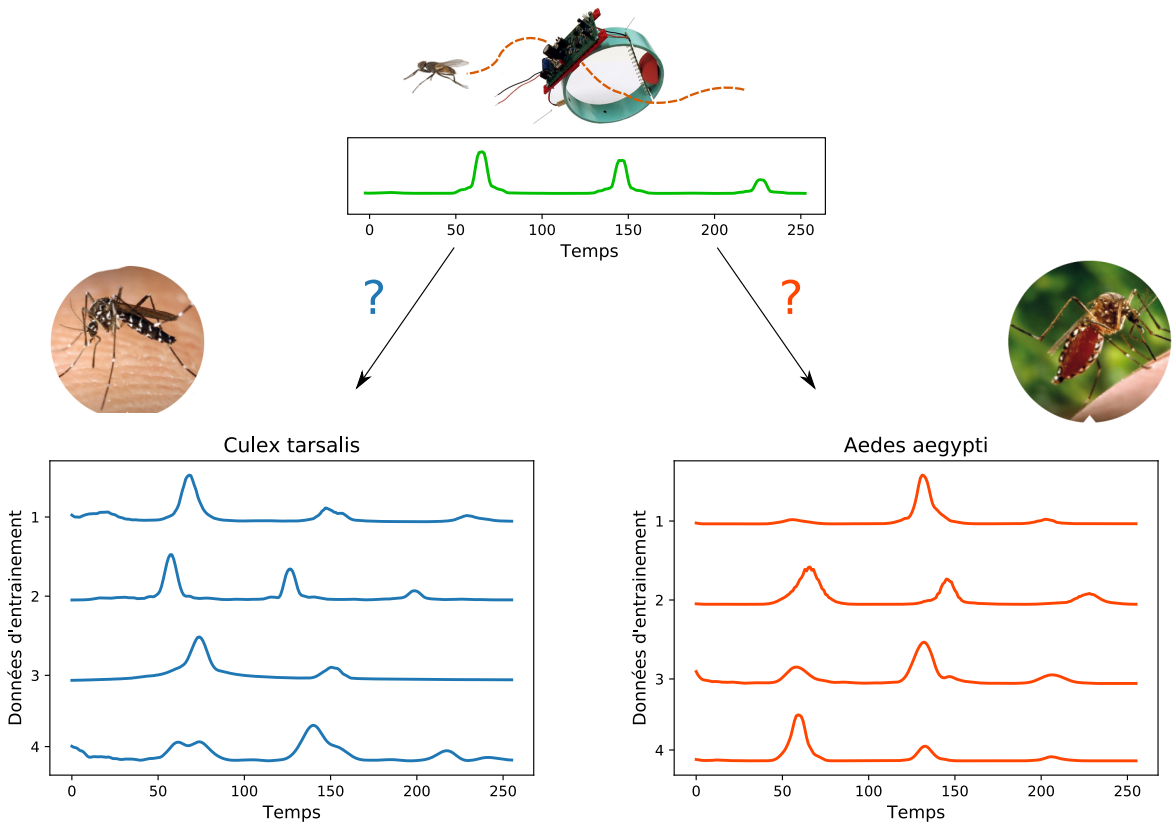


source : `http://timeseriesclassification.com/`

Time series classification

- Time series are regrouped in classes (e.g. Culex / Aedes)
- The goal is to affect a class to new time series

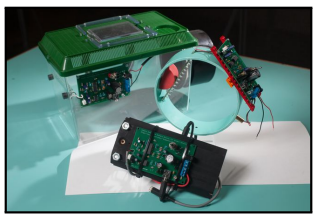




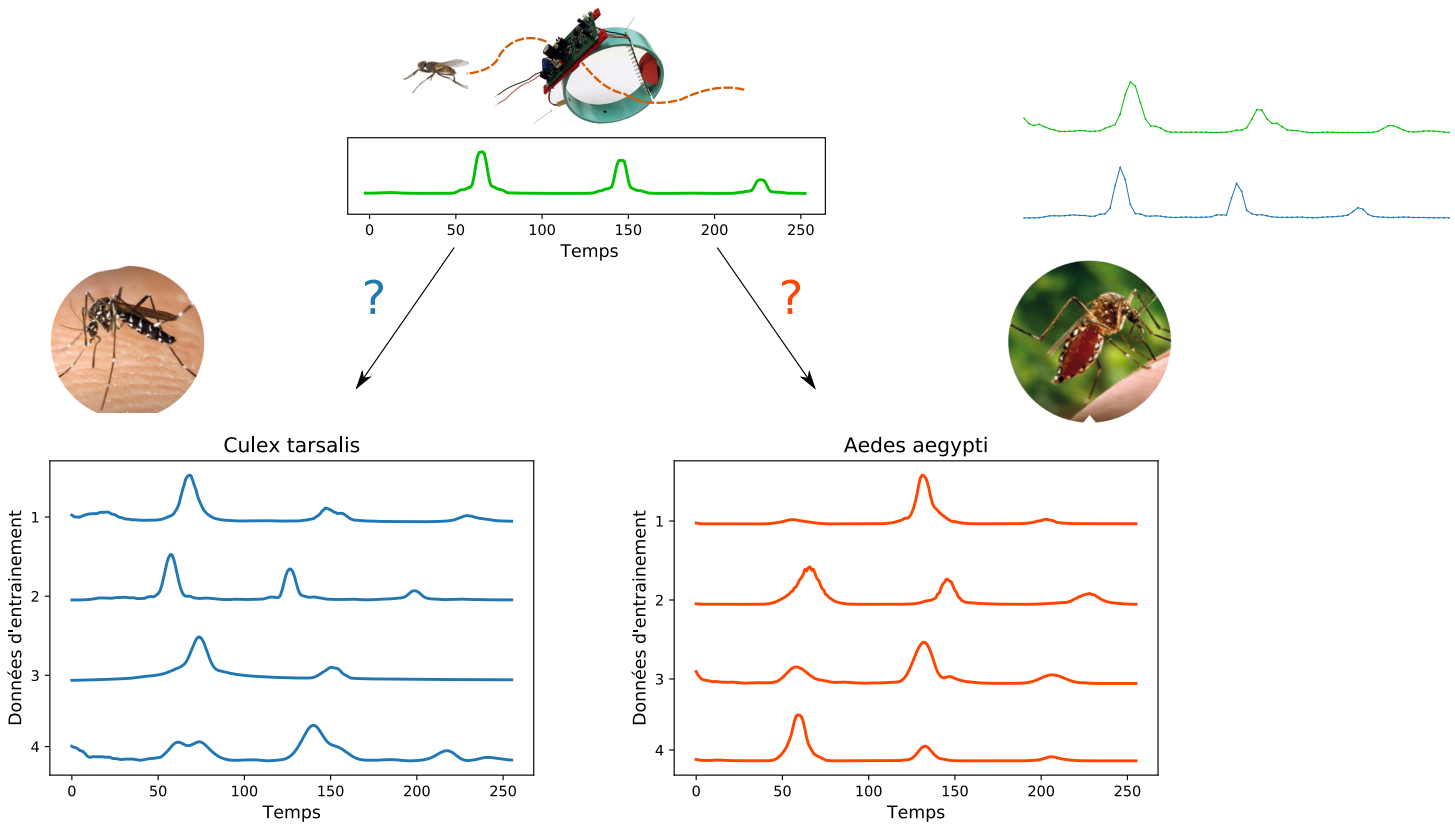
Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. In IEEE International Conference on Data Mining (pp. 470-479)
ICDM 2023 10-year highest-impact paper Award

Time series classification

- Time series are regrouped in classes (e.g. Culex / Aedes)
- The goal is to affect a class to new time series



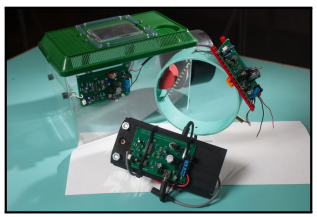
UNIVERSITY OF CALIFORNIA
UCRIVERSIDE



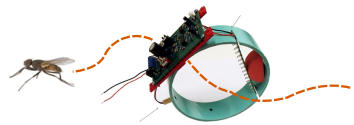
Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. In IEEE International Conference on Data Mining (pp. 470-479)
ICDM 2023 10-year highest-impact paper Award

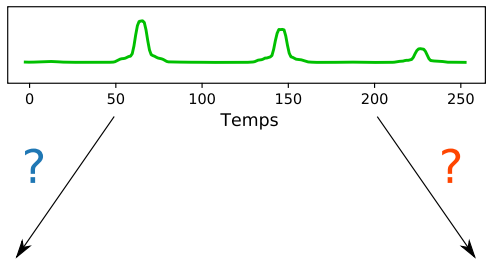
Time series classification

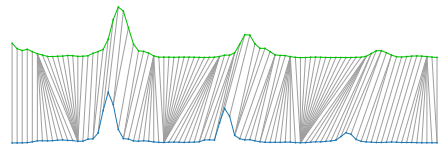
- Time series are regrouped in classes (e.g. Culex / Aedes)
- The goal is to affect a class to new time series



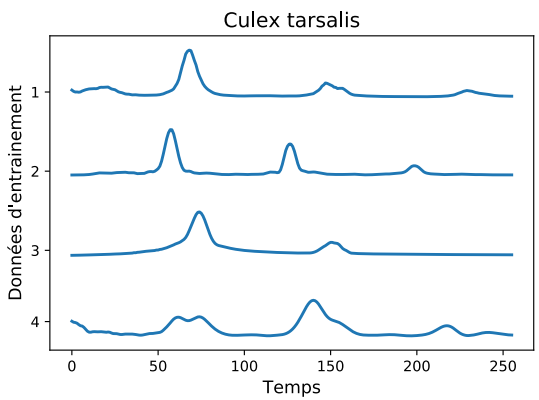
UNIVERSITY OF CALIFORNIA
UCRIVERSIDE

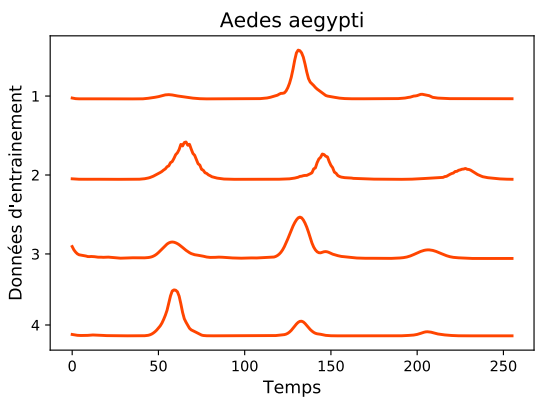






Alignement with
Dynamic Time Warping (DTW)

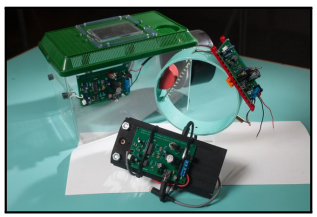




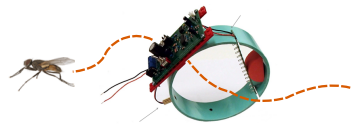
Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. In IEEE International Conference on Data Mining (pp. 470-479)
ICDM 2023 10-year highest-impact paper Award

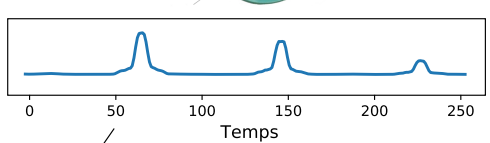
Time series classification

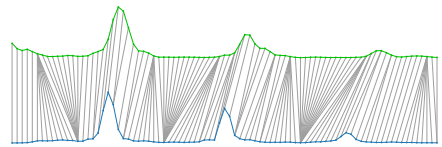
- Time series are regrouped in classes (e.g. Culex / Aedes)
- The goal is to affect a class to new time series



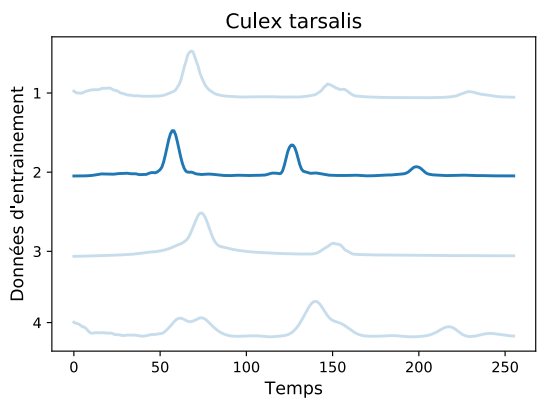
UNIVERSITY OF CALIFORNIA
UCRIVERSIDE

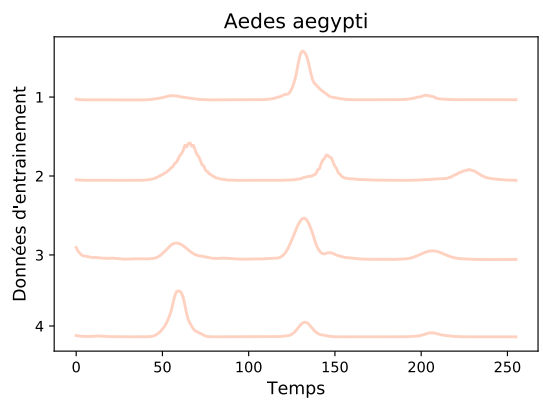






Alignement with
Dynamic Time Warping (DTW)

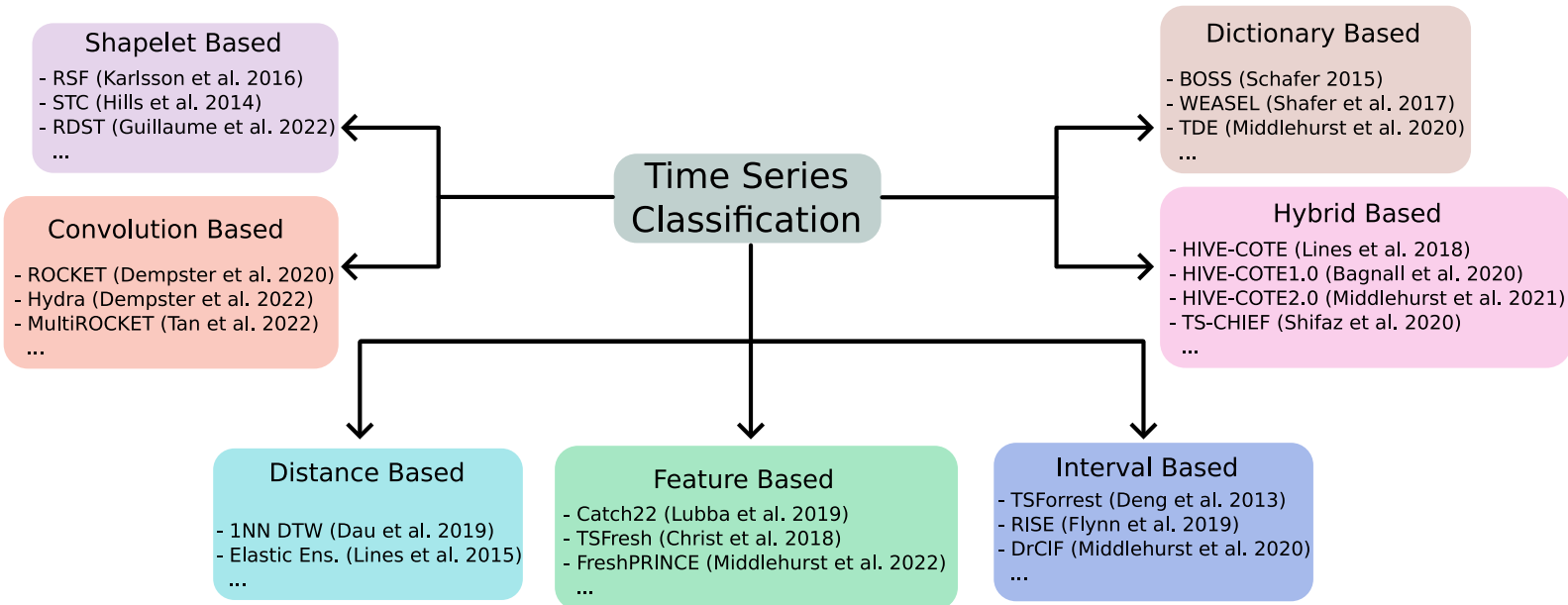




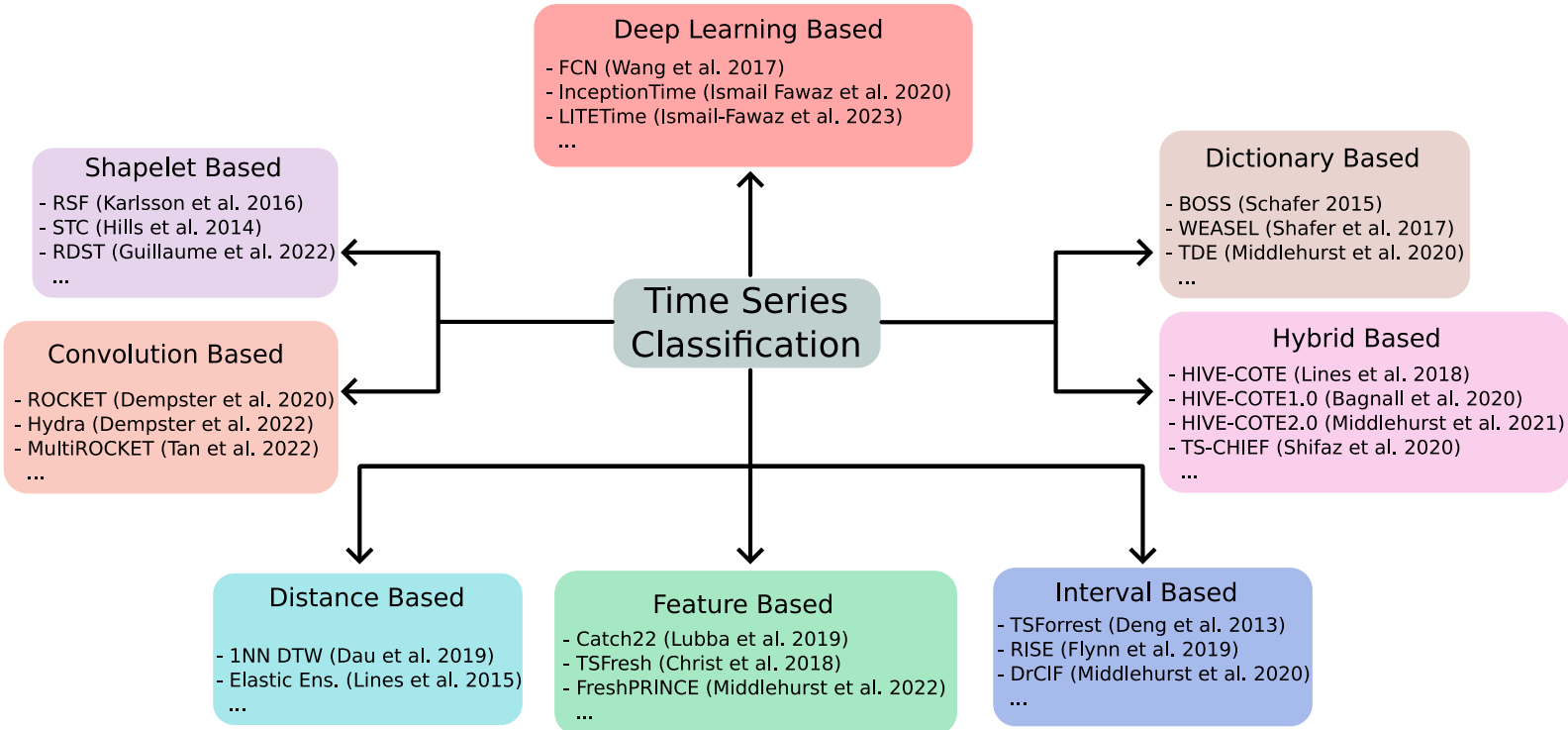
Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. In IEEE International Conference on Data Mining (pp. 470-479)
ICDM 2023 10-year highest-impact paper Award

Contents

Different Tasks



Different Tasks



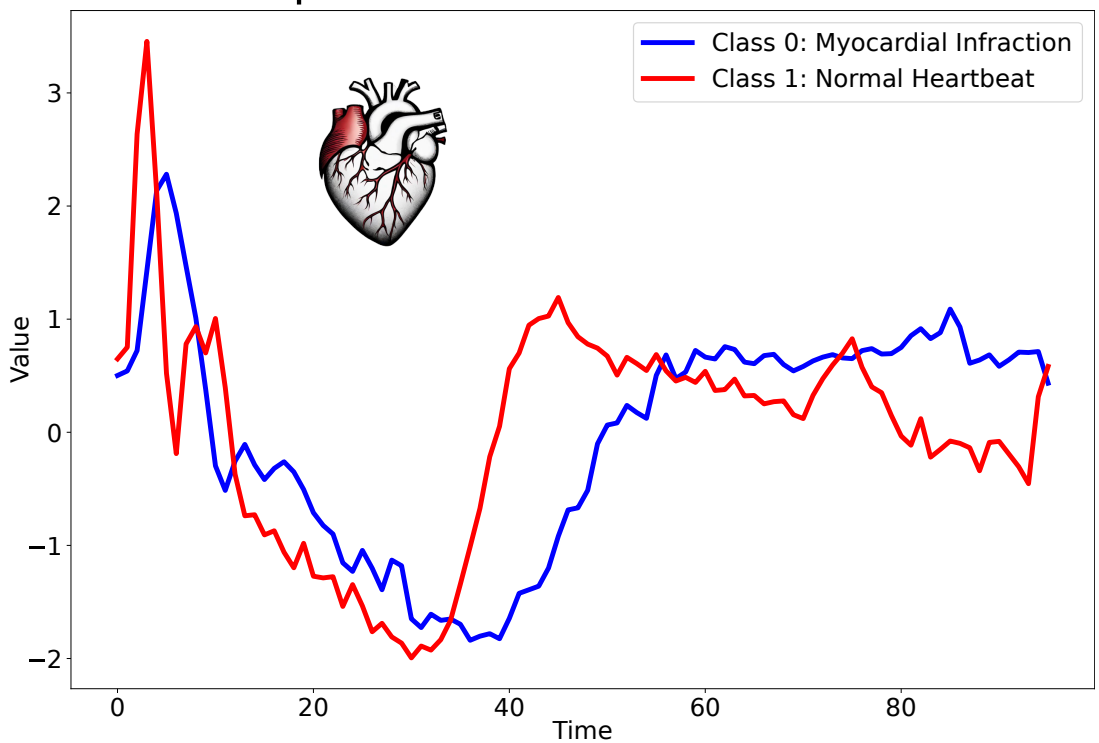
- Middlehurst, Matthew, Patrick Schäfer, and Anthony Bagnall. "Bake off redux: a review and experimental evaluation of recent time series classification algorithms." Data Mining and Knowledge Discovery (2024): 1-74.

Time Series Classification Datasets

The UCR archive, the largest archive for TSC datasets.

- The archive is made of 128 datasets of different tasks: ECG signals, Motion data, Sensors, etc. Already split into train and test.

Example on the ECG200 dataset:

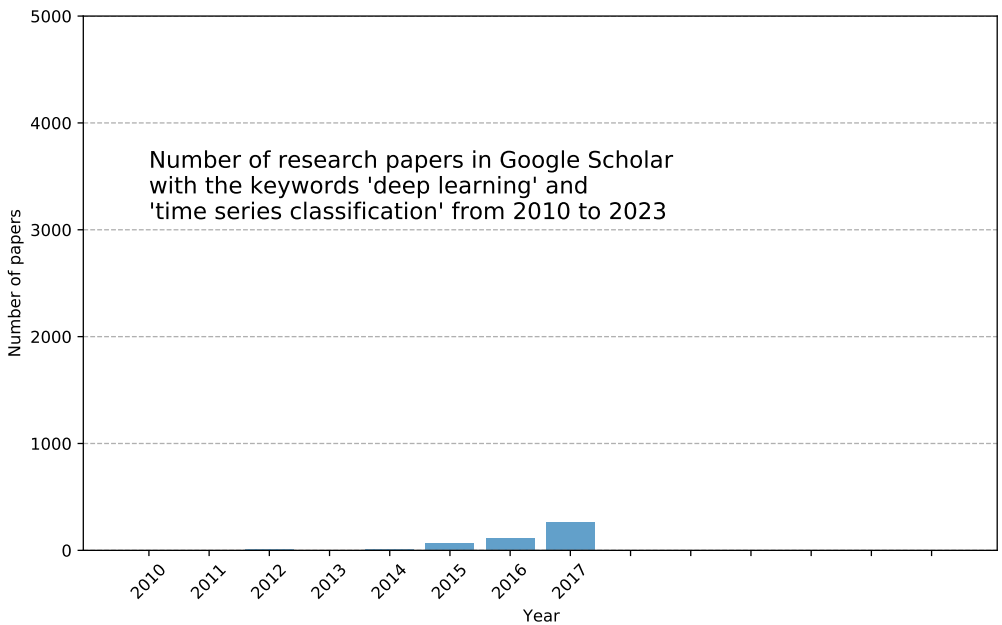


Dau, Hoang Anh, et al. "The UCR time series archive." IEEE/CAA Journal of Automatica Sinica 6.6 (2019): 1293-1305.

Contents

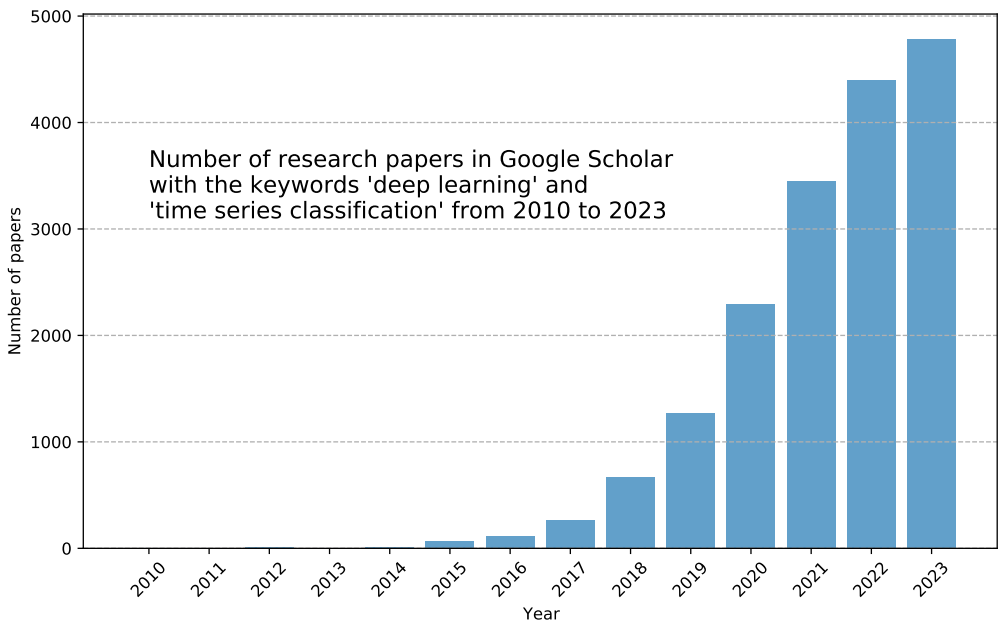
Why Deep Learning for TSC ?

- Deep learning had great success on other type of data (computer vision, NLP, etc.), so why not on time series ?
- Deep Learning models learn to extract features and perform classification with one set of parameters (end-to-end)



Why Deep Learning for TSC ?

- Deep learning had great success on other type of data (computer vision, NLP, etc.), so why not on time series ?
- Deep Learning models learn to extract features and perform classification with one set of parameters (end-to-end)

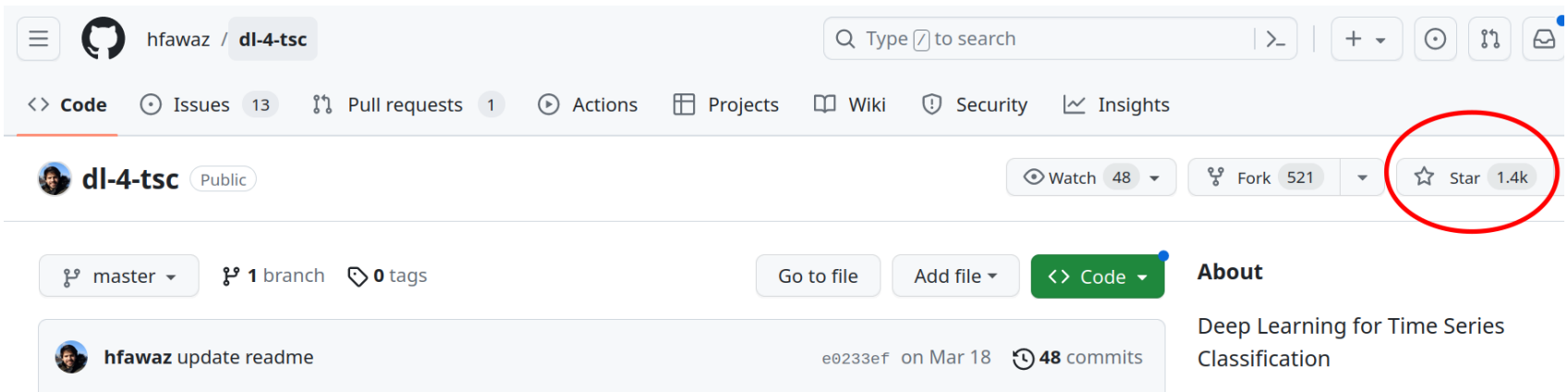


Contents

DL4TSC - 2019

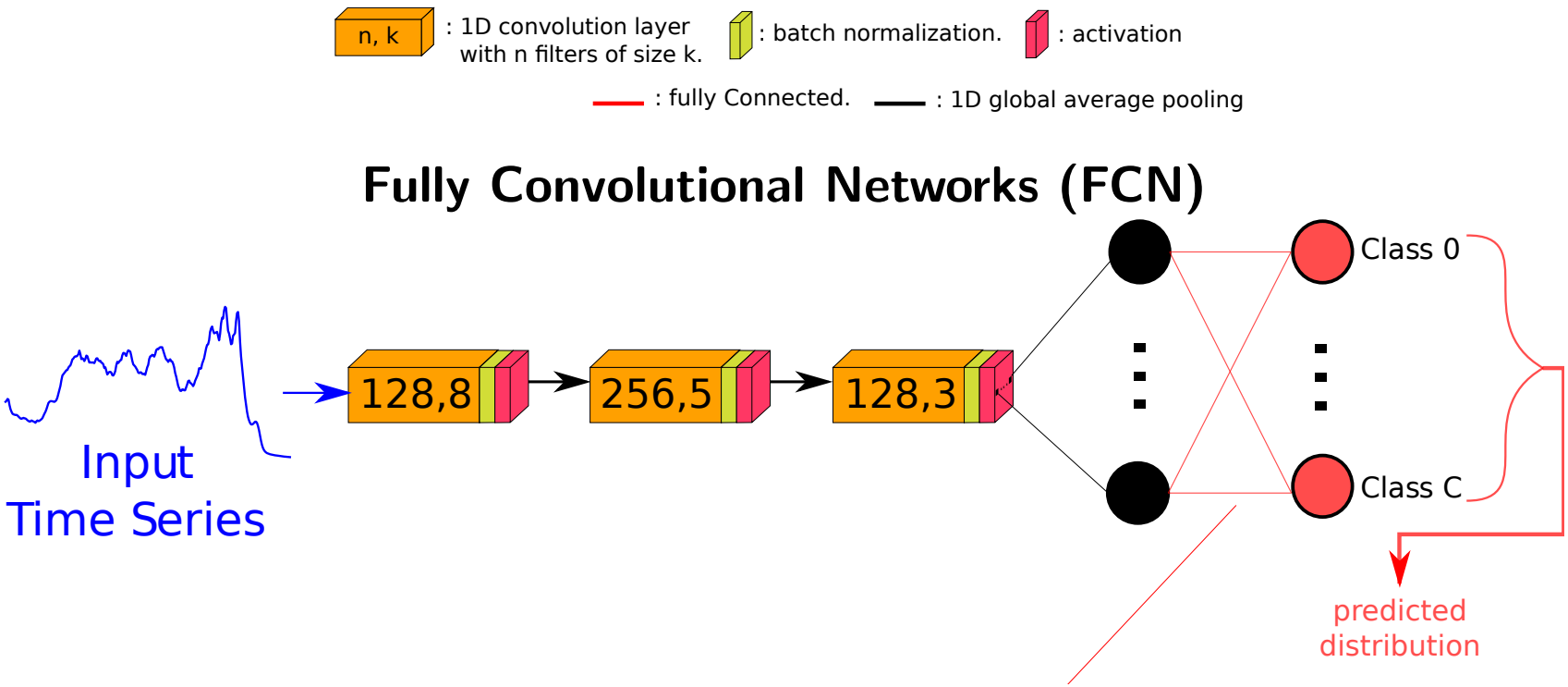
In **2019**, our team presented a study of Deep Learning for Time Series Classification (cited more than 2.5K times (GS)).

- We selected papers with enough details (or code) to reproduce the model's architecture
- We benchmarked all the models on the UCR archive
- We published the code on Github for reproducibility and got very positive feedback



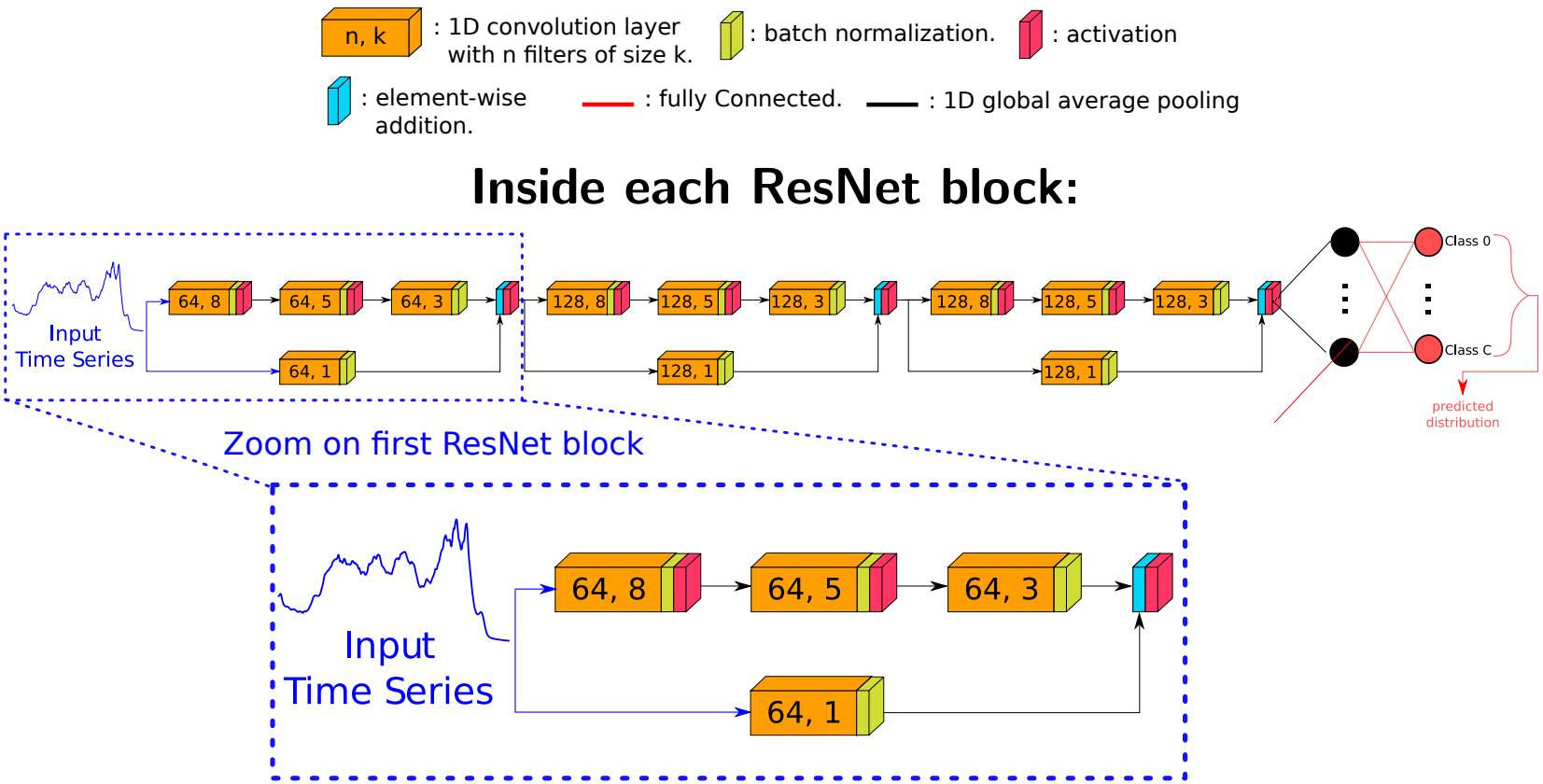
Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. Data mining and knowledge discovery, 33(4), 917-963.

DL4TSC - Some Architectures



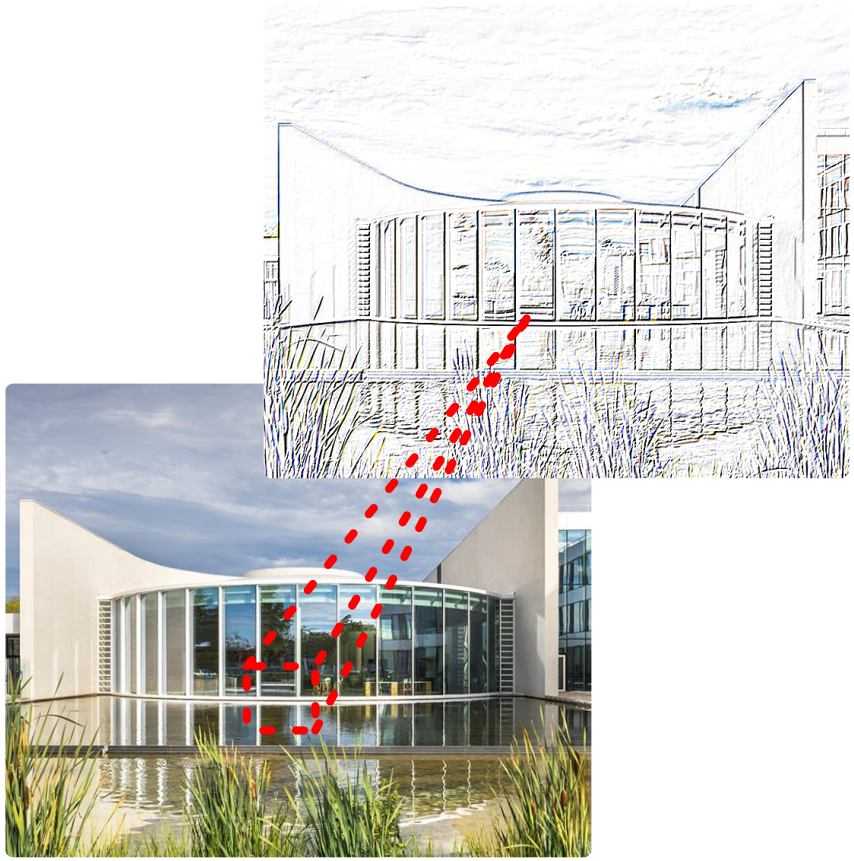
Wang, Z., Yan, W., & Oates, T. (2017, May). Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks (IJCNN) (pp. 1578-1585). IEEE.

DL4TSC - Some Architectures

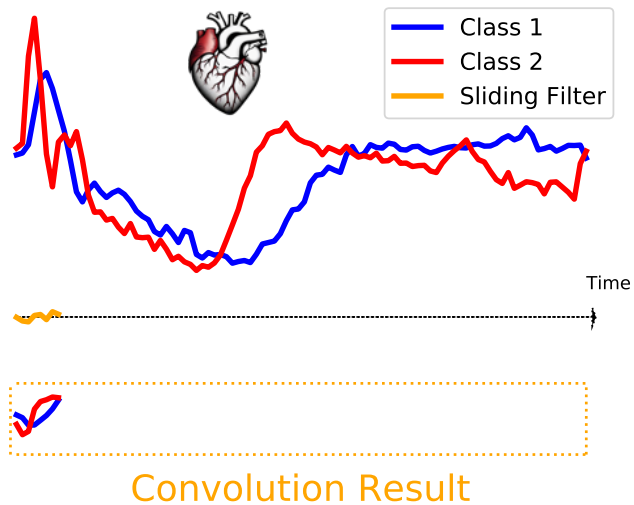


- Wang, Z., Yan, W., & Oates, T. (2017, May). Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks (IJCNN) (pp. 1578-1585). IEEE.

Convolutions on Images vs Time Series

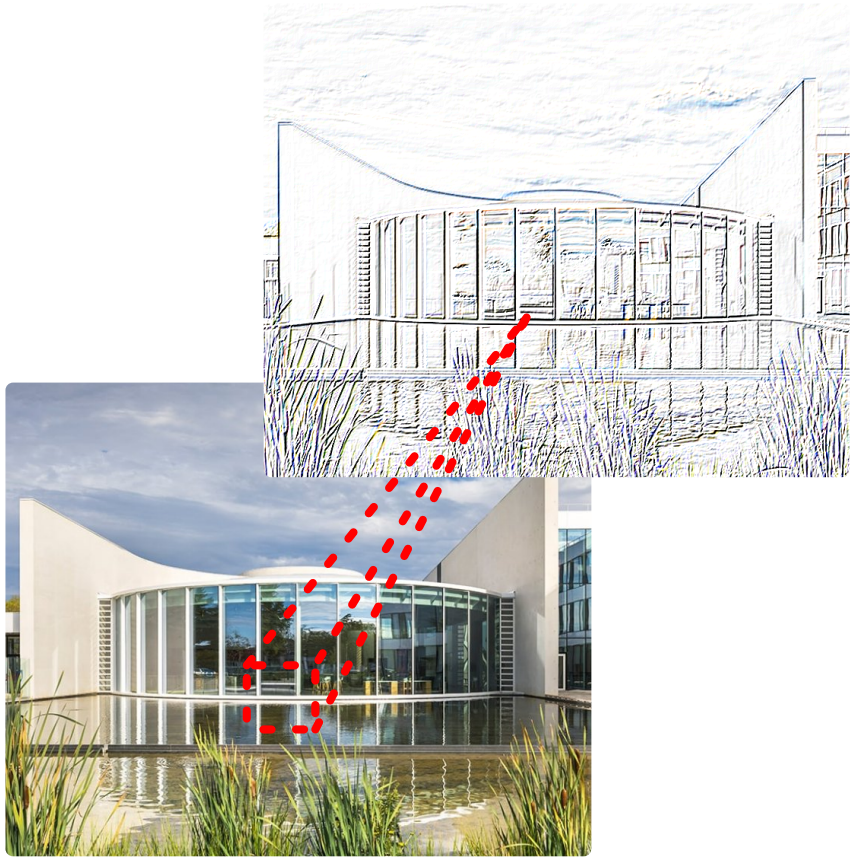


The result of a applying an edge detection convolution on an image

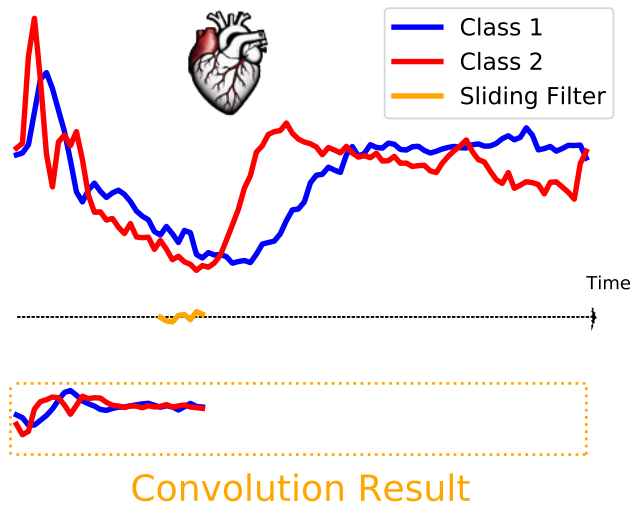


The result of applying a learned discriminative convolution on the ECG200 dataset

Convolutions on Images vs Time Series

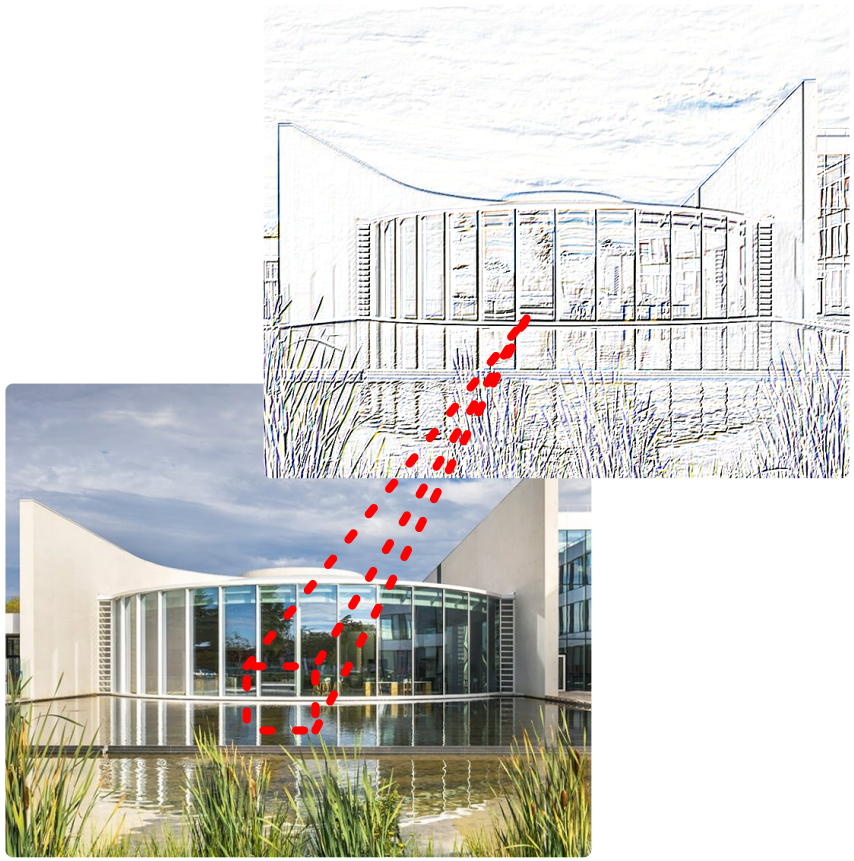


The result of a applying an edge detection convolution on an image

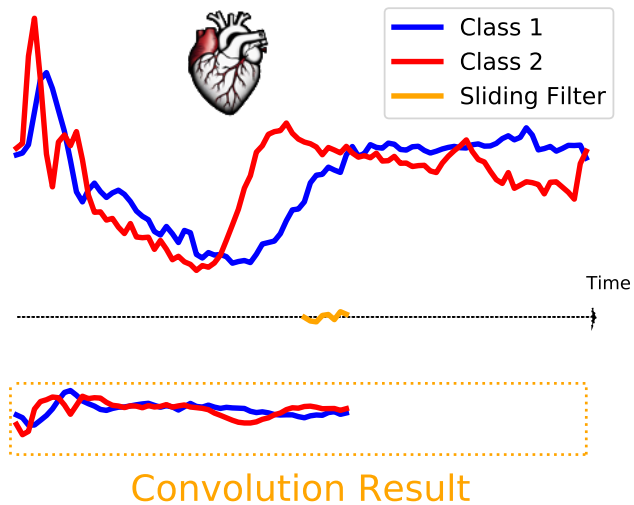


The result of applying a learned discriminative convolution on the ECG200 dataset

Convolutions on Images vs Time Series

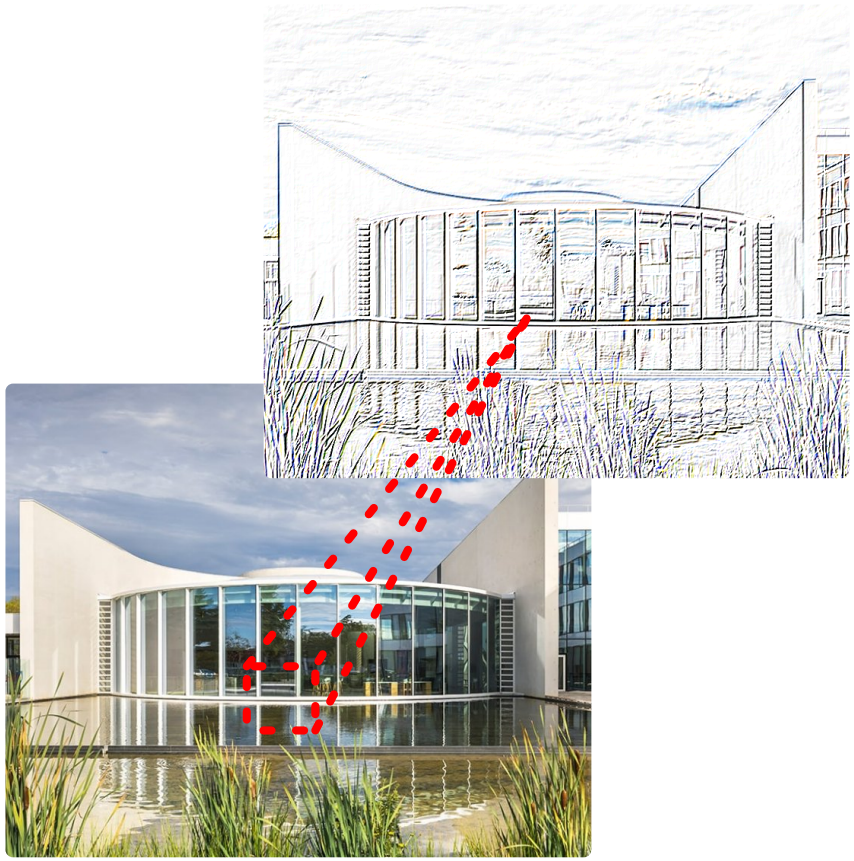


The result of a applying an edge detection convolution on an image

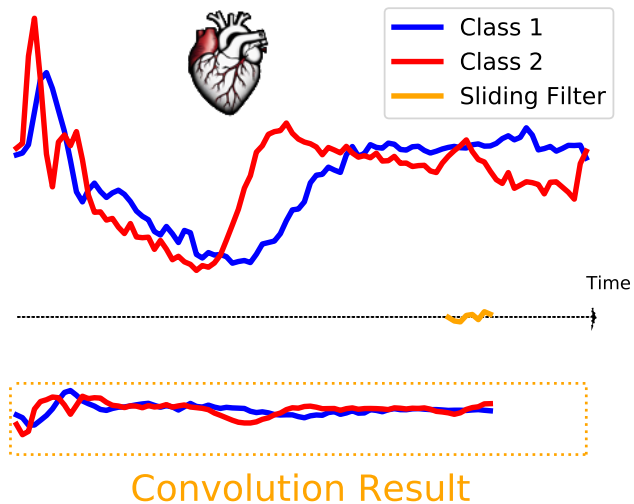


The result of applying a learned discriminative convolution on the ECG200 dataset

Convolutions on Images vs Time Series

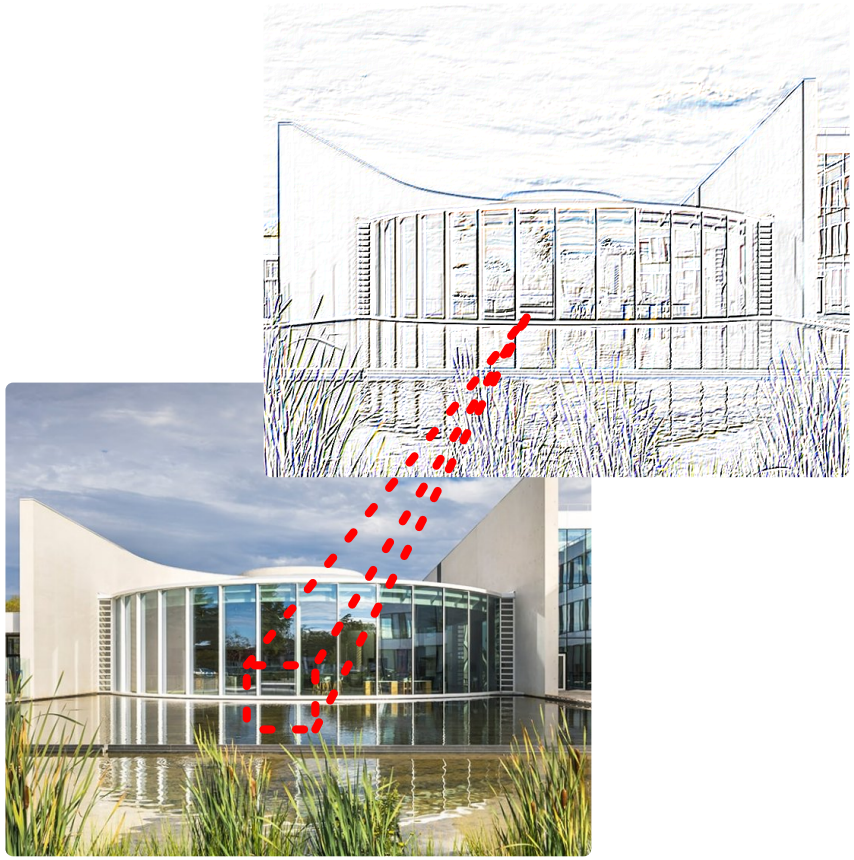


The result of a applying an edge detection convolution on an image

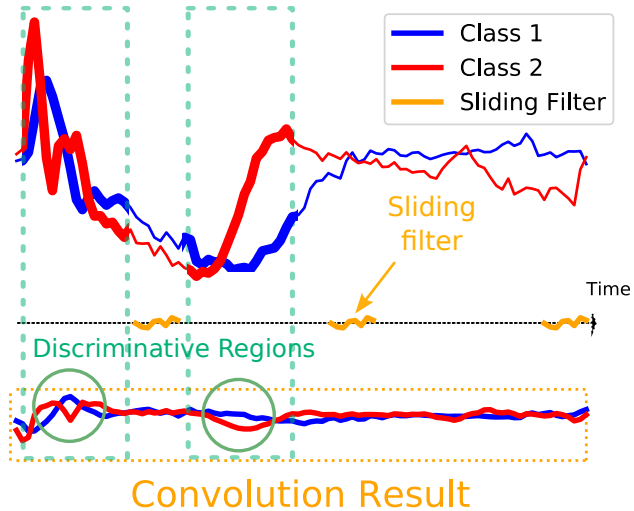


The result of applying a learned discriminative convolution on the ECG200 dataset

Convolutions on Images vs Time Series

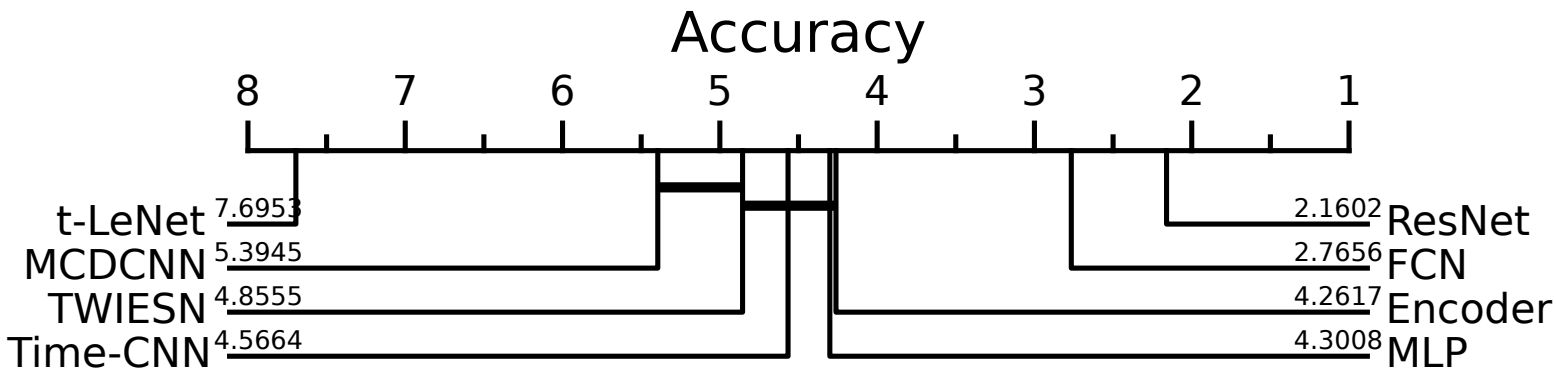


The result of a applying an edge detection convolution on an image



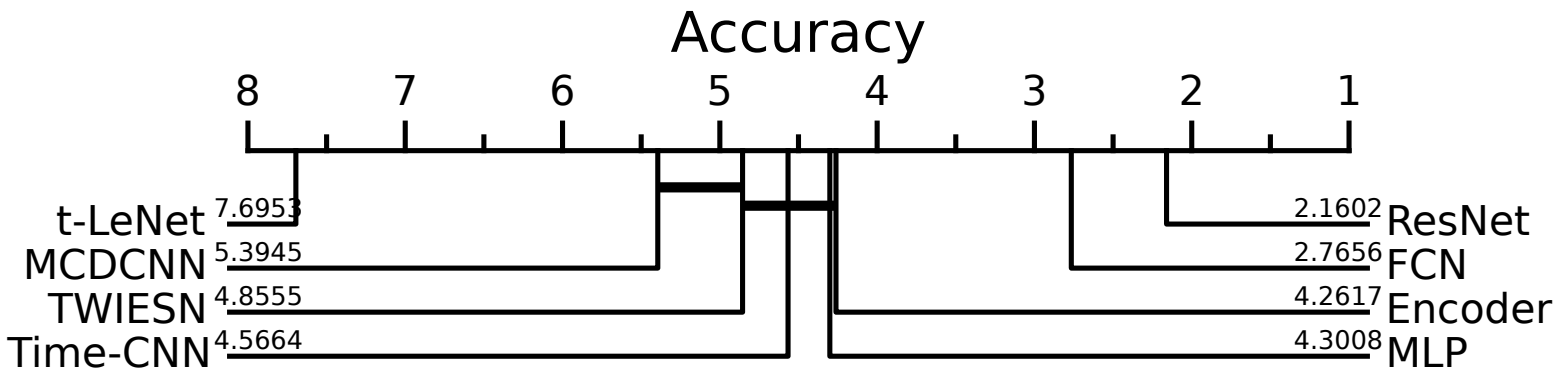
The result of applying a learned discriminative convolution on the ECG200 dataset

DL4TSC - Experiments: Results



Benavoli, Alessio, Giorgio Corani, and Francesca Mangili. "Should we really use post-hoc tests based on mean-ranks?." The Journal of Machine Learning Research 17.1 (2016): 152-161.

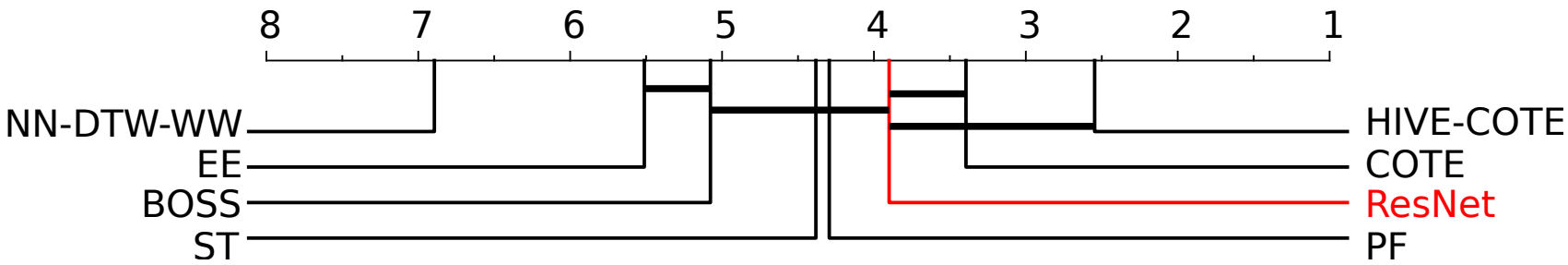
DL4TSC - Experiments: Results



- A Critical Difference Diagram is a visualization technique that displays the performance of multiple classifiers on multiple datasets
- The classifiers in a CDD are ordered following their average rank
- Ranking is made following the accuracy on all test sets of the 128 UCR archive datasets
- Each clique (black line between classifiers) represent a statistically non-significant difference in performance between linked classifiers on the 128 datasets of the UCR archive (using Wilcoxon Signed-Rank test + Holm correction)

Benavoli, Alessio, Giorgio Corani, and Francesca Mangili. "Should we really use post-hoc tests based on mean-ranks?." The Journal of Machine Learning Research 17.1 (2016): 152-161.

DL4TSC - Experiments: Results Compared to 2017 Bake Off



- The best performing deep learning model from our review, ResNet, performs as good as state-of-the-art non-deep models for Time Series Classification on the UCR archive.
- No significance is found in difference in performance between ResNet and Hive-COTE.
- Showcasing the high performance of deep models for TSC

Bagnall, Anthony, et al. "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances." Data mining and knowledge discovery 31 (2017): 606-660.

Contents

InceptionTime

Inception

- Originally proposed by Google for image recognition problems [1]
- Further developed to reach state-of-the-art results on ImageNet [2]

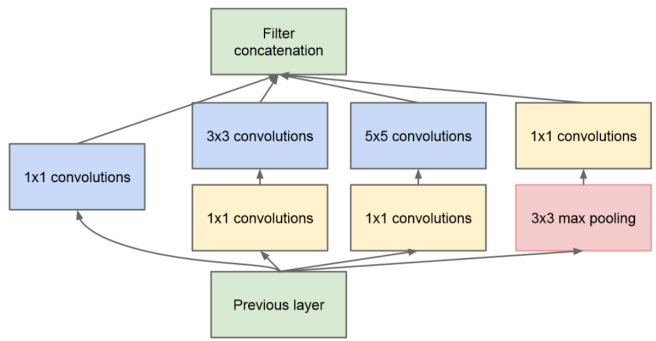


Figure: Inception module for image recognition [1]

- [1] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- [2] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

InceptionTime

Inception

- Originally proposed by Google for image recognition problems [1]
- Further developed to reach state-of-the-art results on ImageNet [2]
- Main ideas:
  - Apply convolutions of different resolutions to capture various patterns
  - Use a bottleneck layer in order to reduce the number of parameters

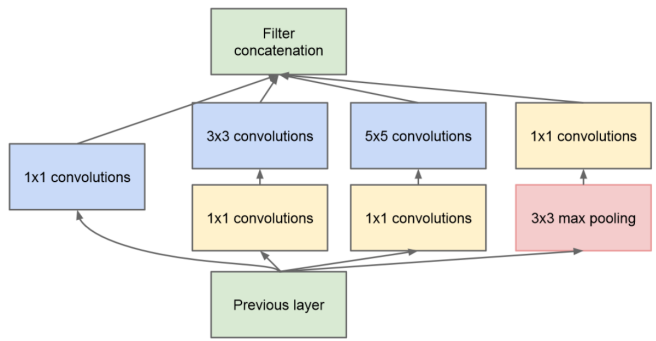
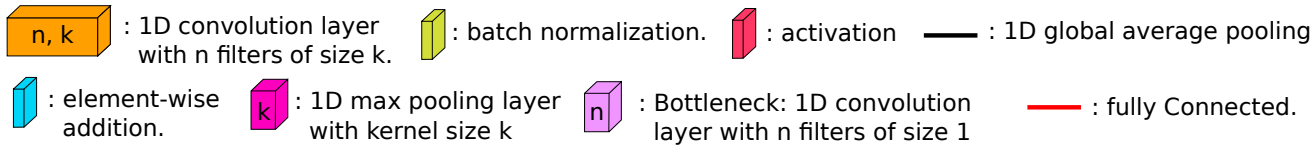


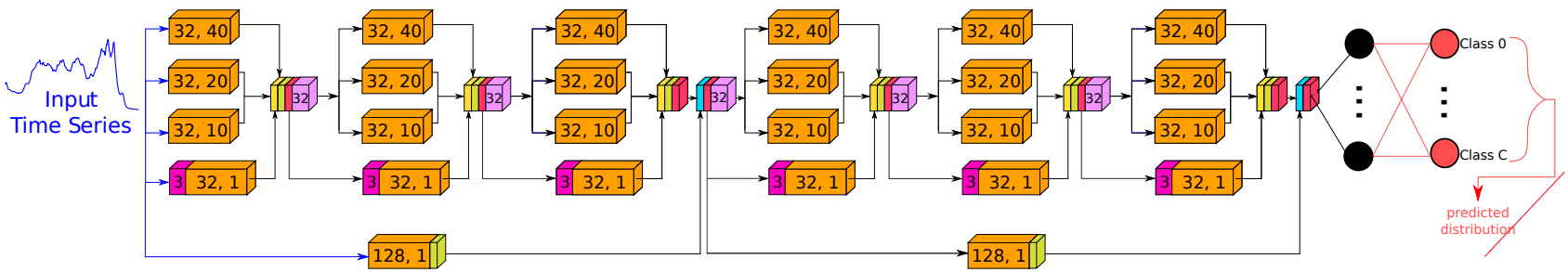
Figure: Inception module for image recognition [1]

- [1] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- [2] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

InceptionTime

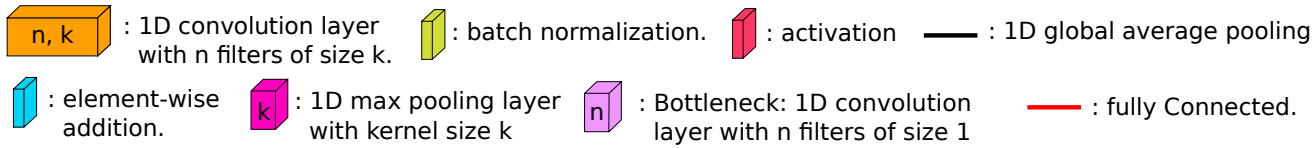


Inception architecture for TSC

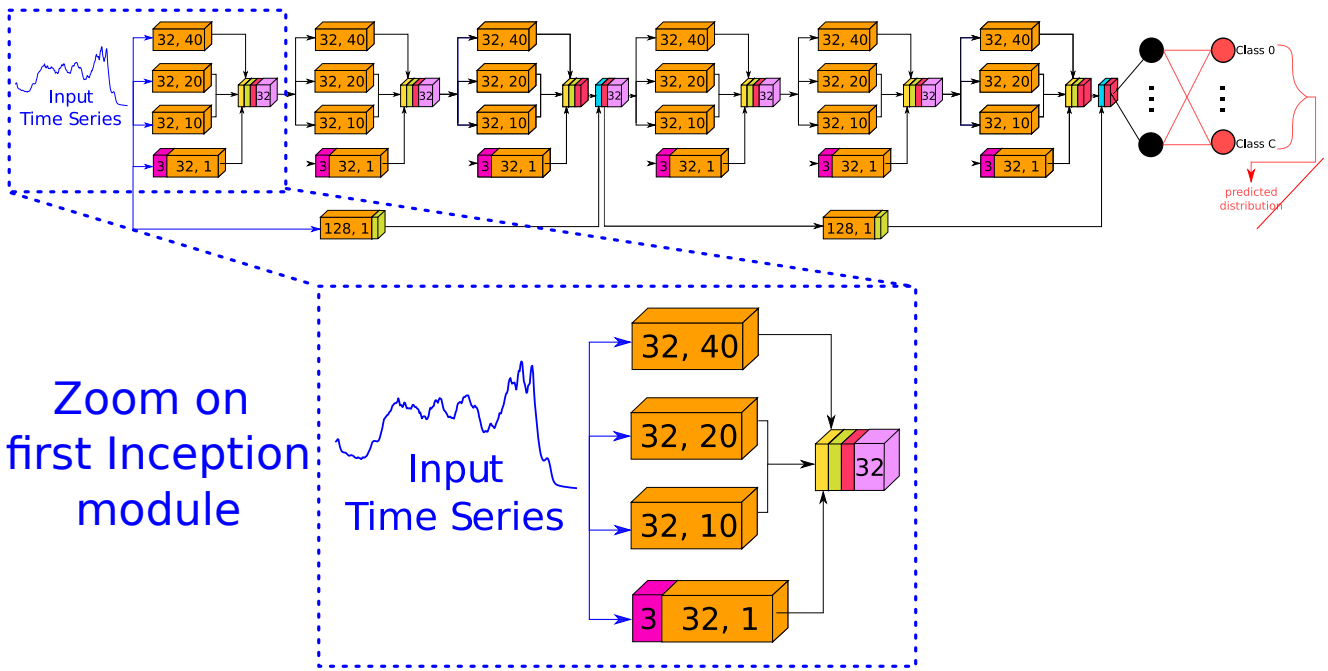


Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." 2020 Data Mining and Knowledge Discovery.

InceptionTime

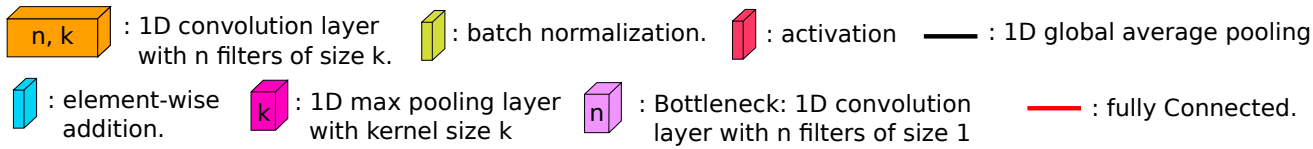


Inception architecture for TSC

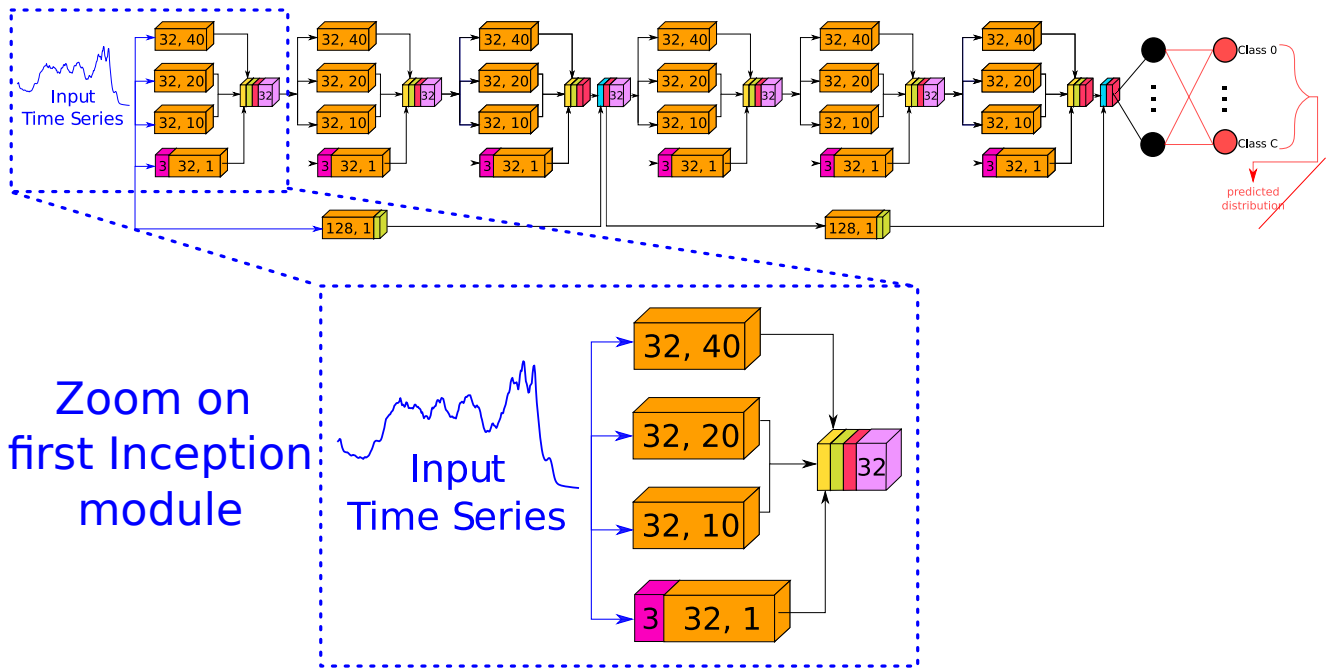


Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." 2020 Data Mining and Knowledge Discovery.

InceptionTime



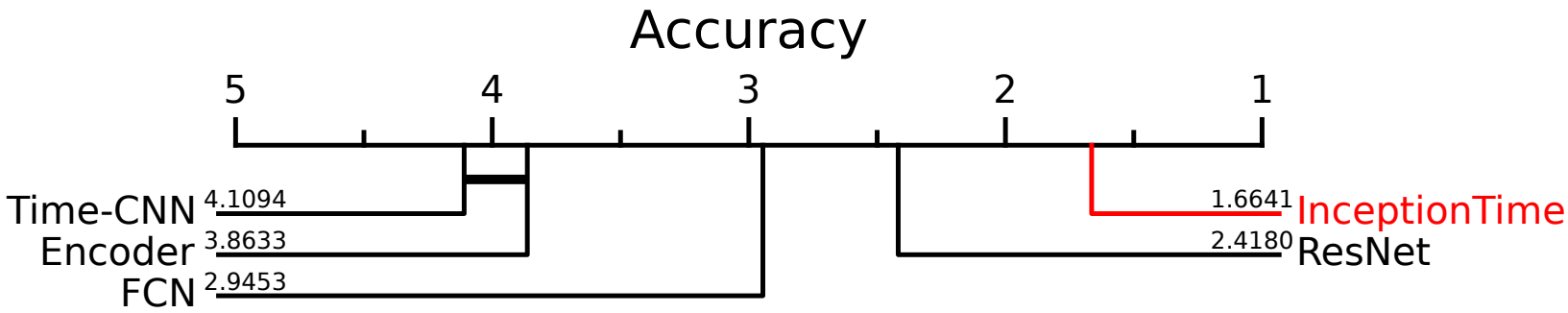
Inception architecture for TSC



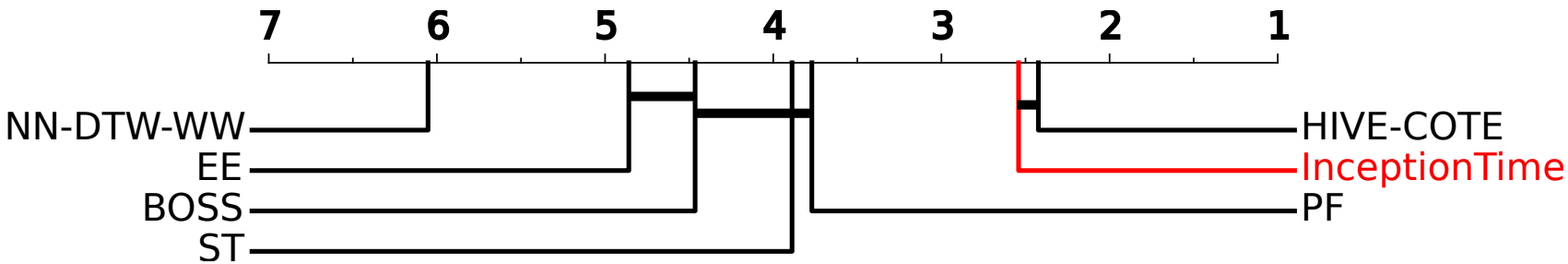
InceptionTime is an ensemble of five Inception models.

Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." 2020 Data Mining and Knowledge Discovery.

InceptionTime: Performance with Best Deep Learners



In 2020, InceptionTime became the state-of-the-art Deep Learning model for Time Series Classification on the UCR archive.

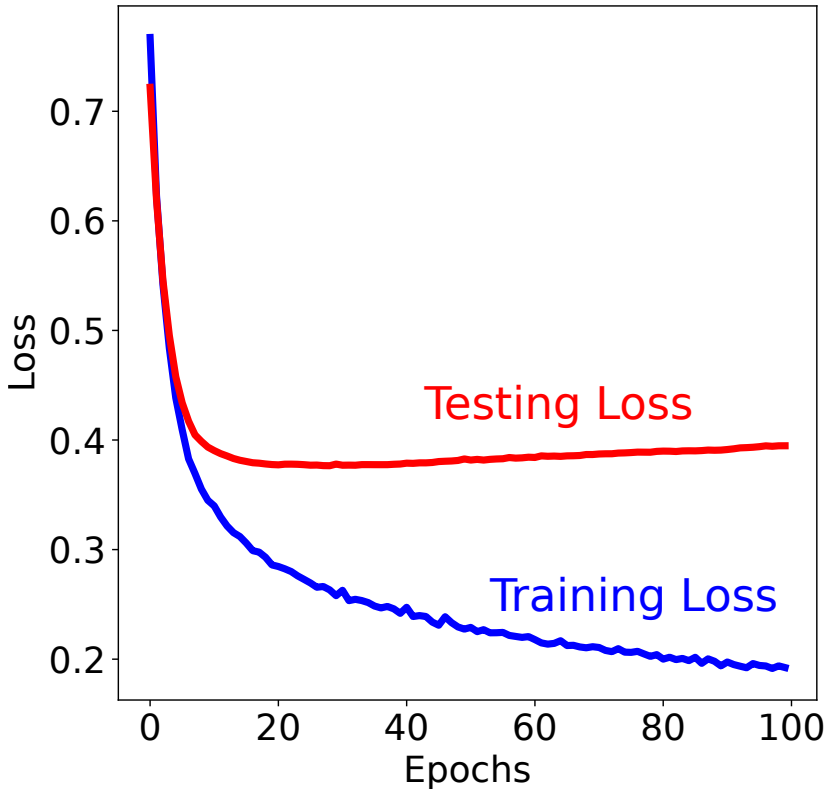


Compared to non-deep models, InceptionTime performed significantly better than most approaches with a no significance found in difference with HIVE-COTE.

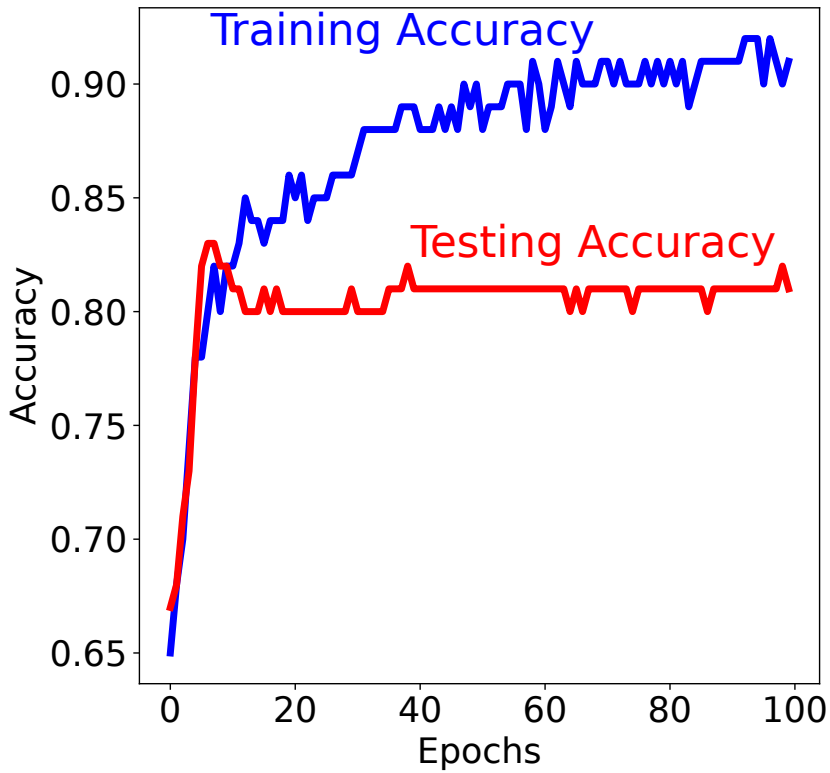
Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." 2020 Data Mining and Knowledge Discovery.

Contents

Regularization to Avoid Overfitting



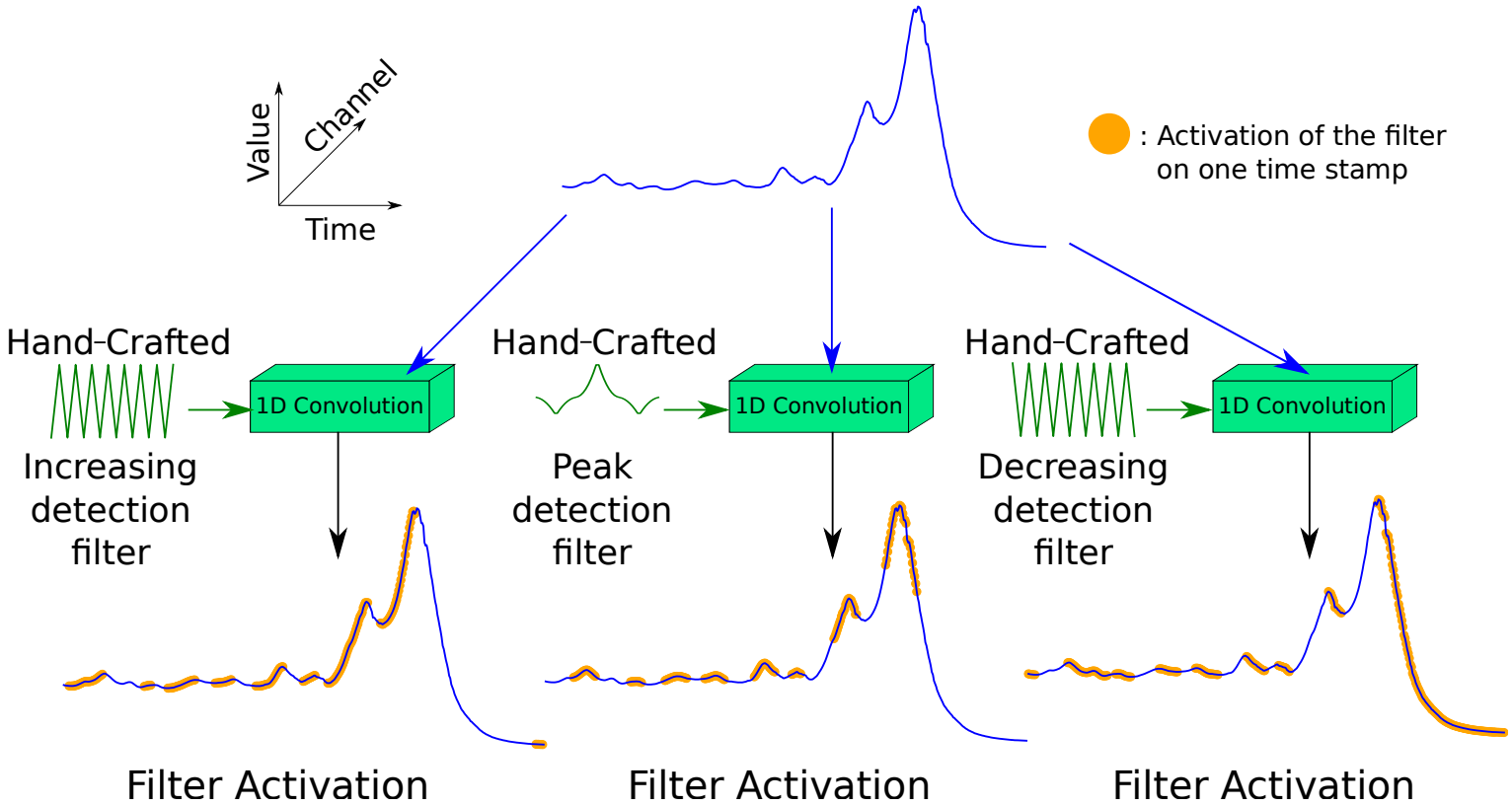
Classification Loss during training



Classification Accuracy during training

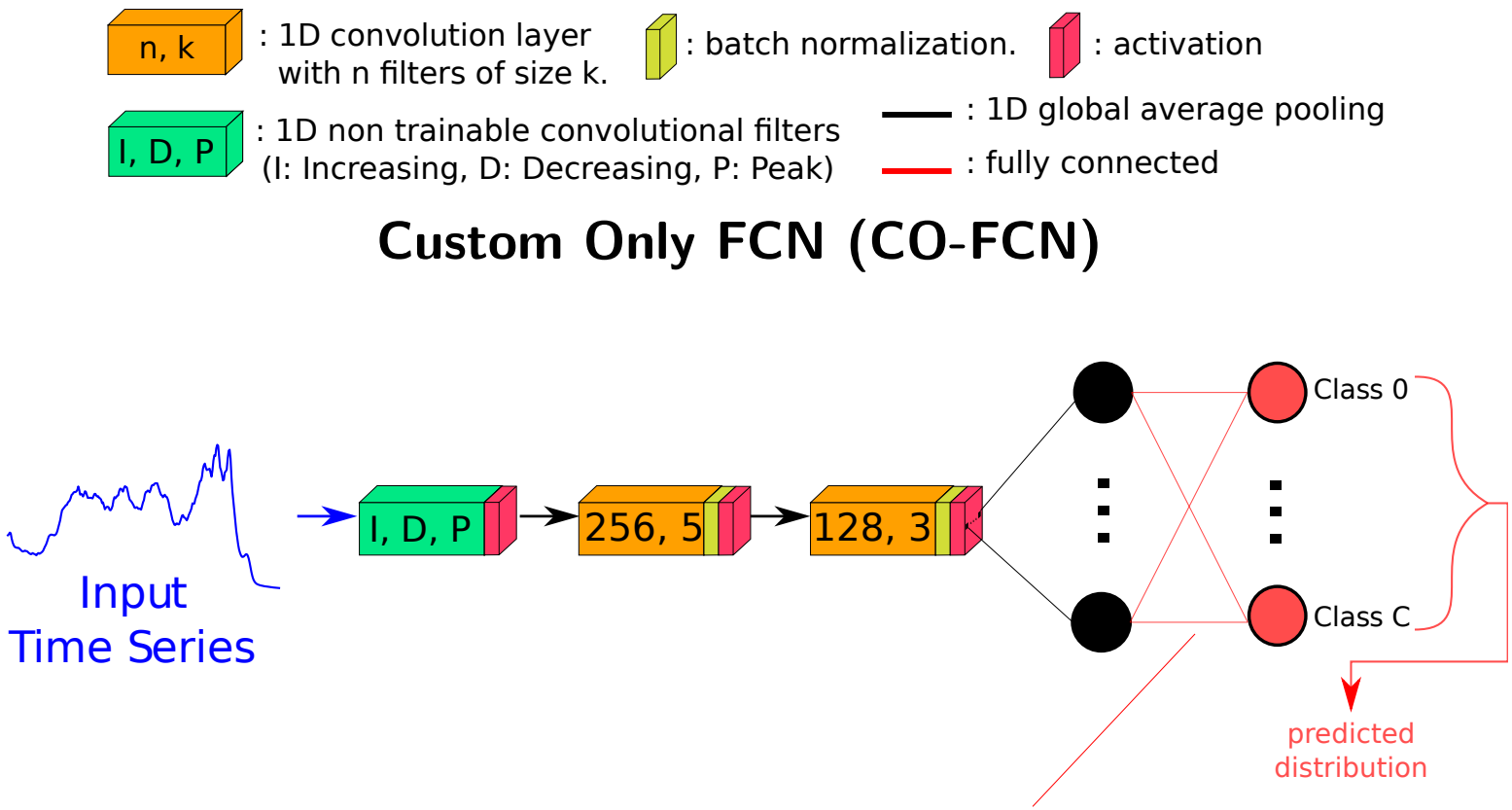
Hand-Crafted Convolutional Filters

In 2022, we proposed new hand-crafted convolutional filters.



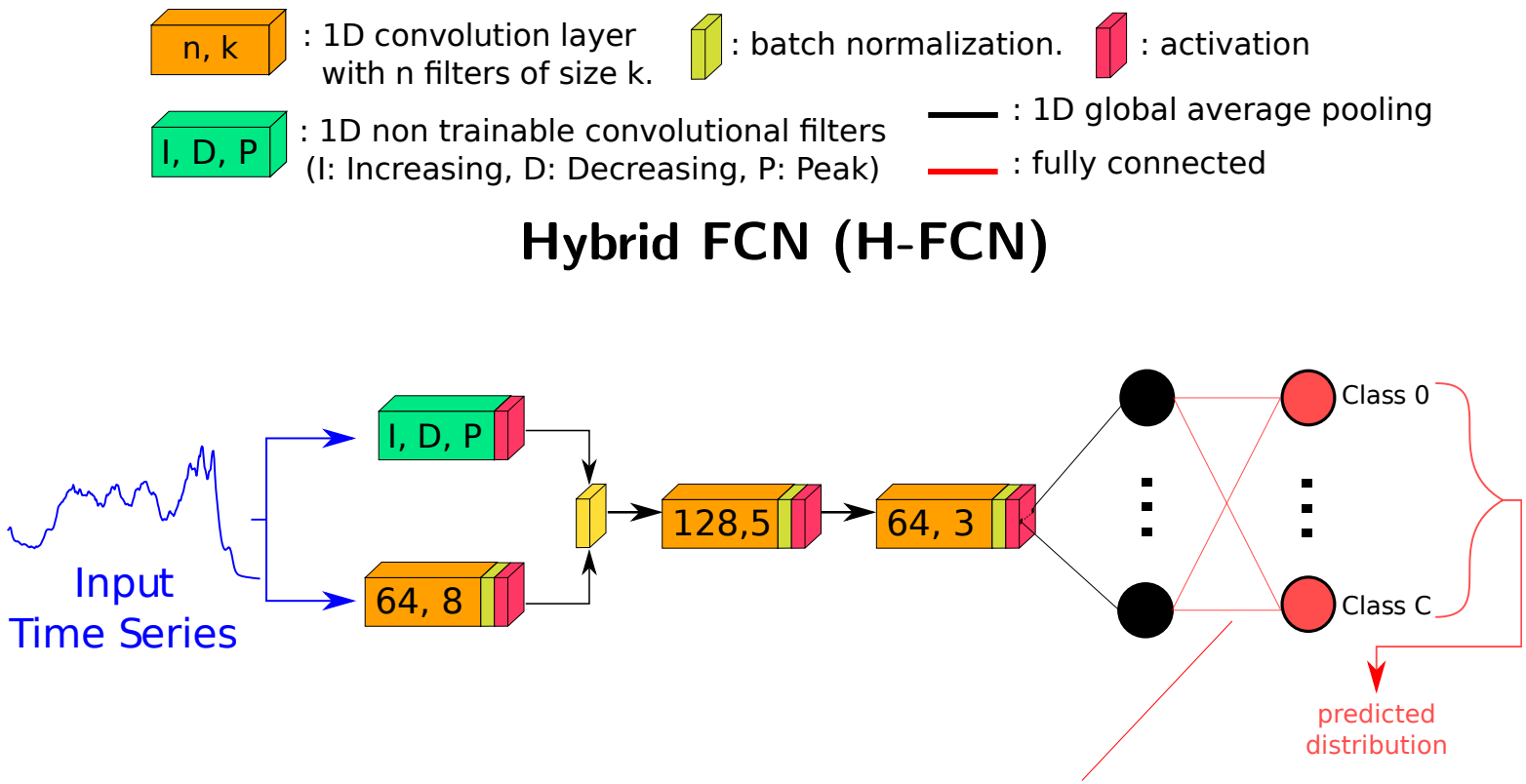
Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

Hand-Crafted Convolutional Filters - Architectures



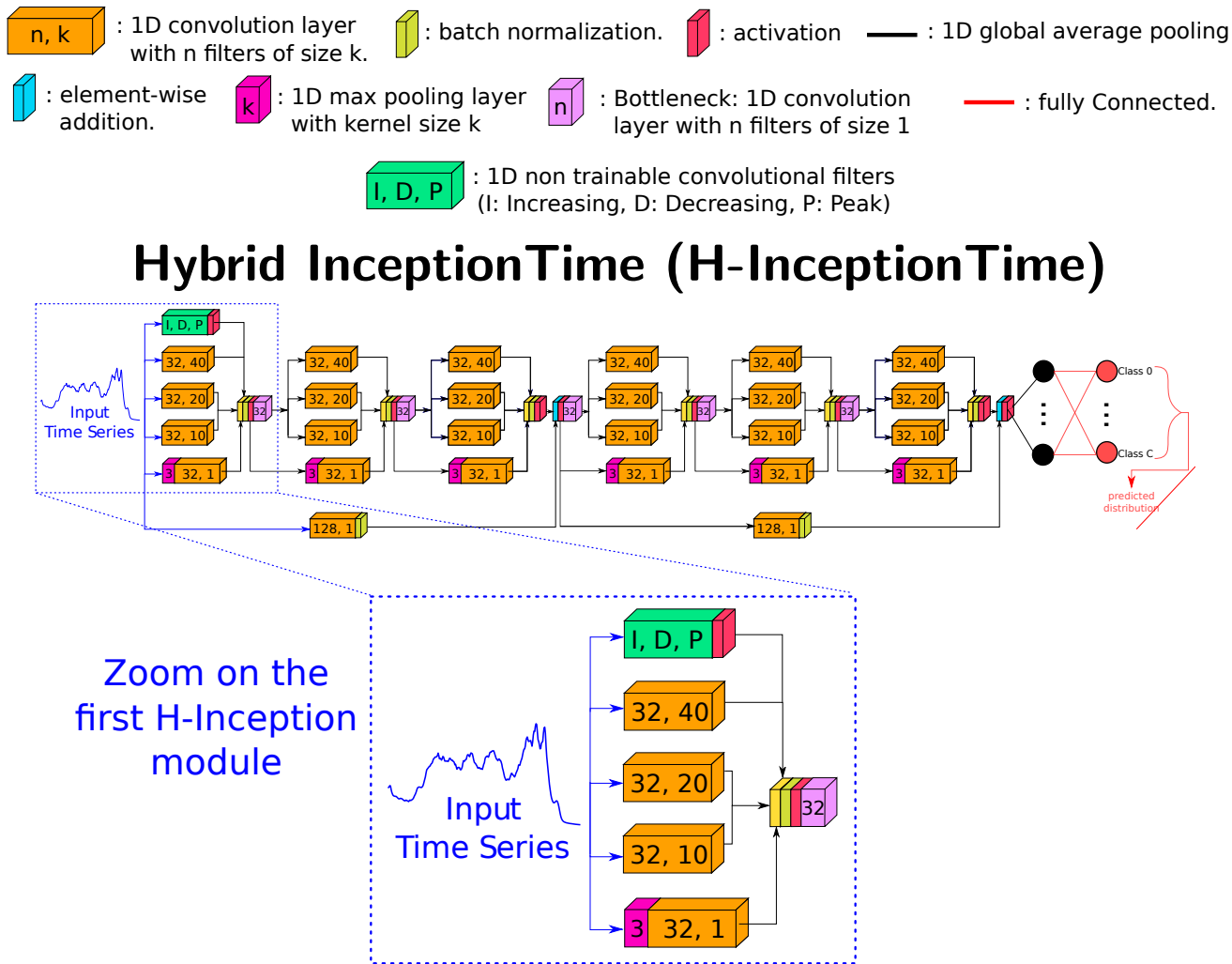
Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

Hand-Crafted Convolutional Filters - Architectures



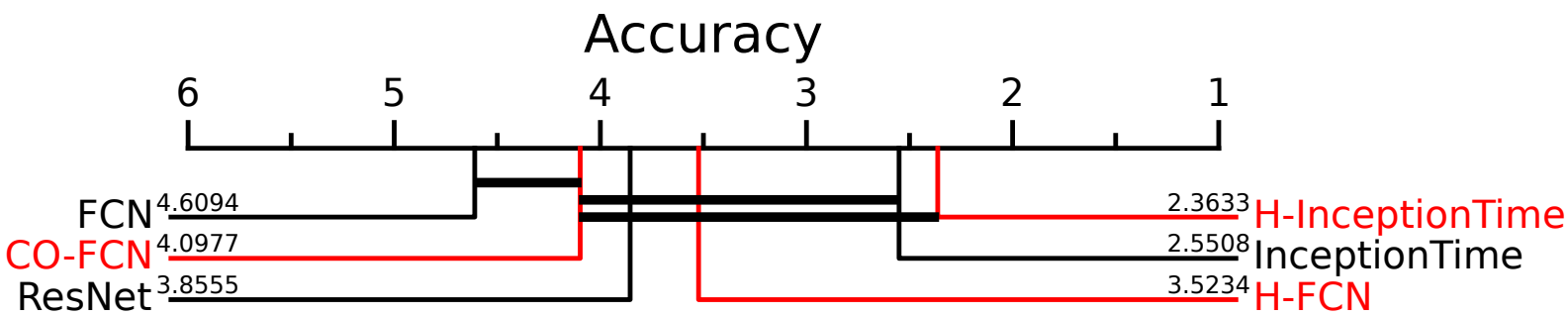
Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

Hand-Crafted Convolutional Filters - Architectures



Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

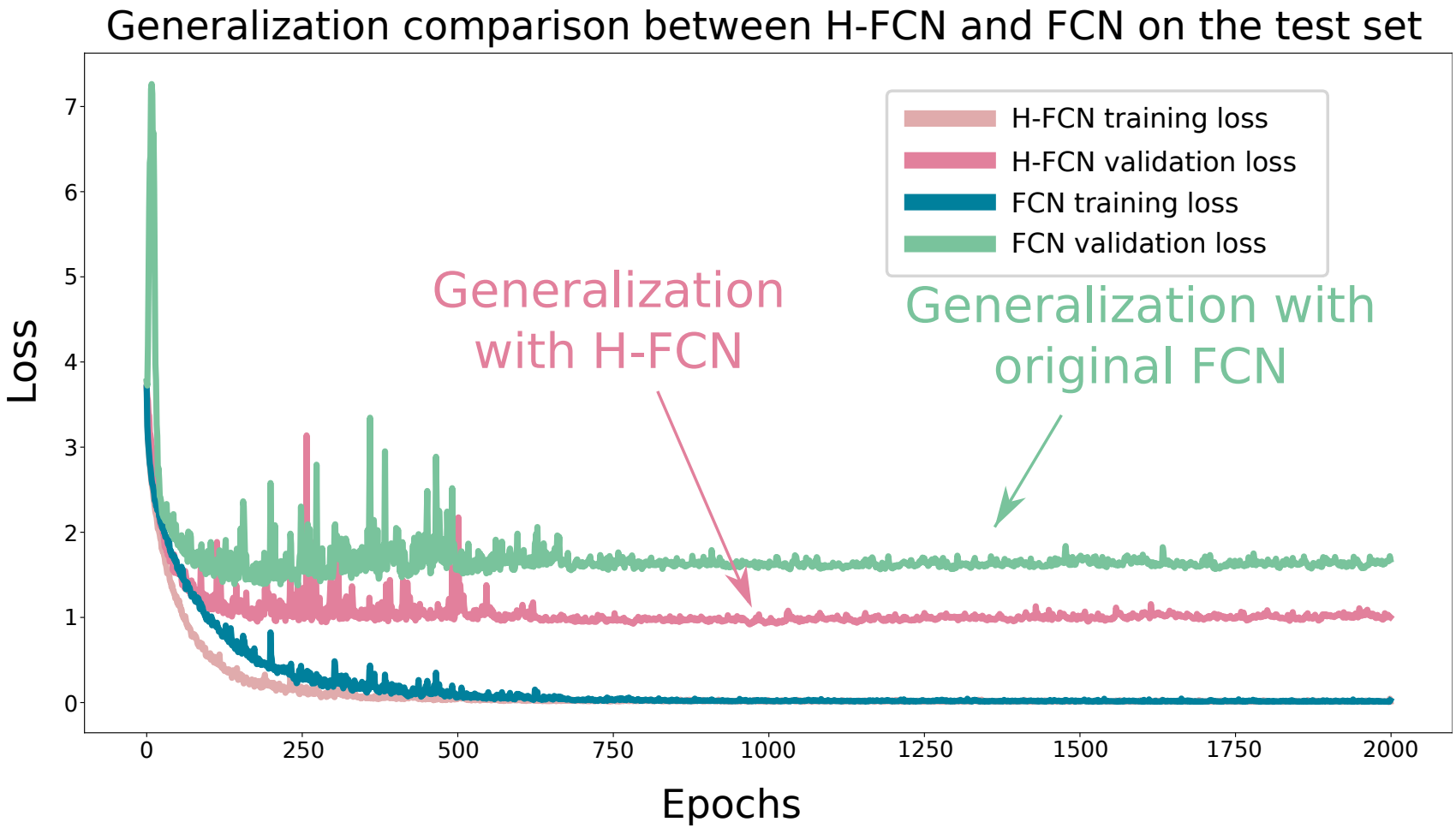
Hand-Crafted Convolutional Filters: Performance



- Both CO-FCN and H-FCN can beat FCN
- H-FCN is better than ResNet which means hand-crafted filters helped FCN generalize better than residual connections can
- H-InceptionTime beats InceptionTime, concluding that hand-crafted filters always help generalize

Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

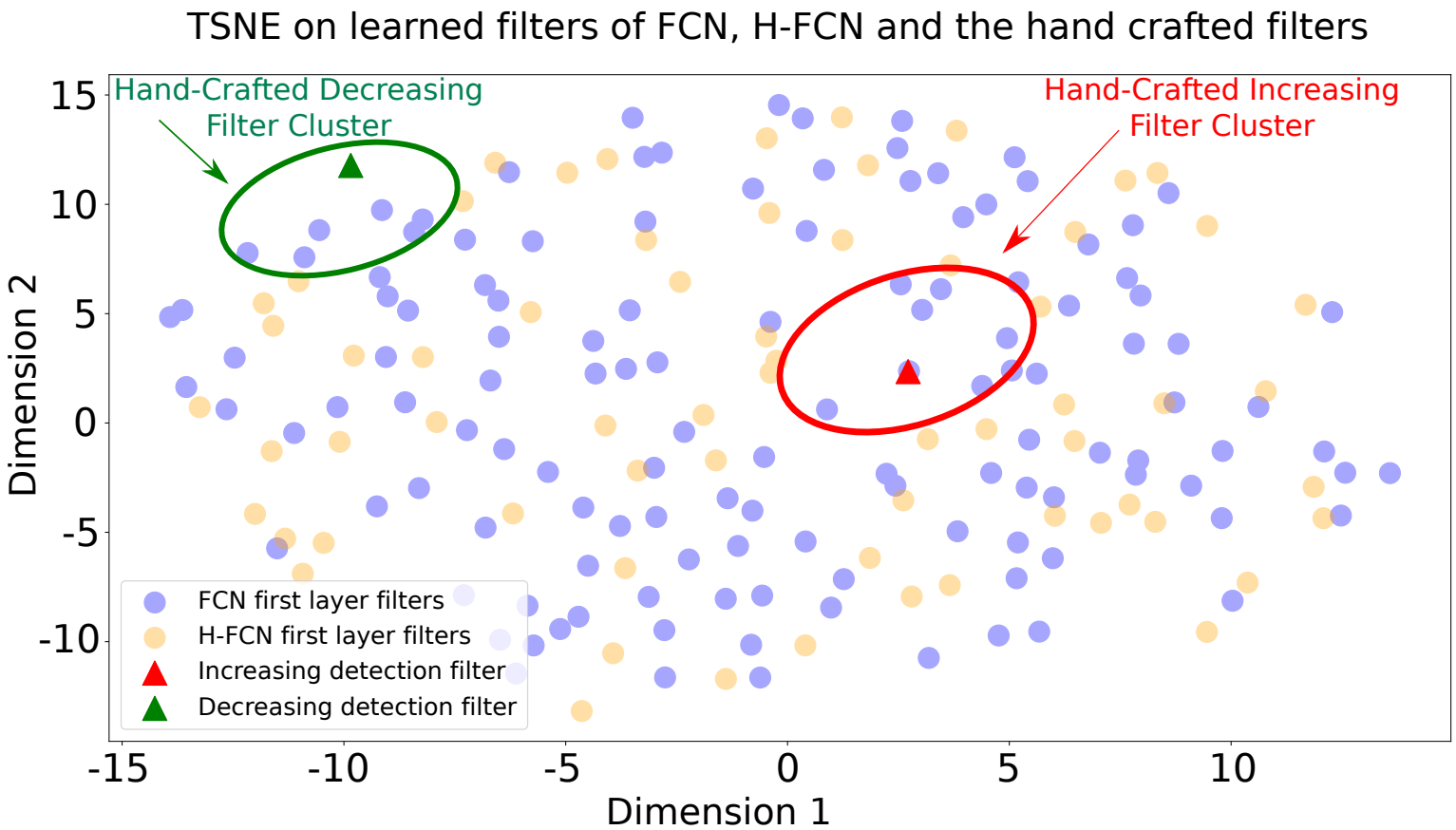
Hand-Crafted Convolutional Filters: Generalization



Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

Hand-Crafted Convolutional Filters: Kernel Space

Do the models learn the same filters again ?

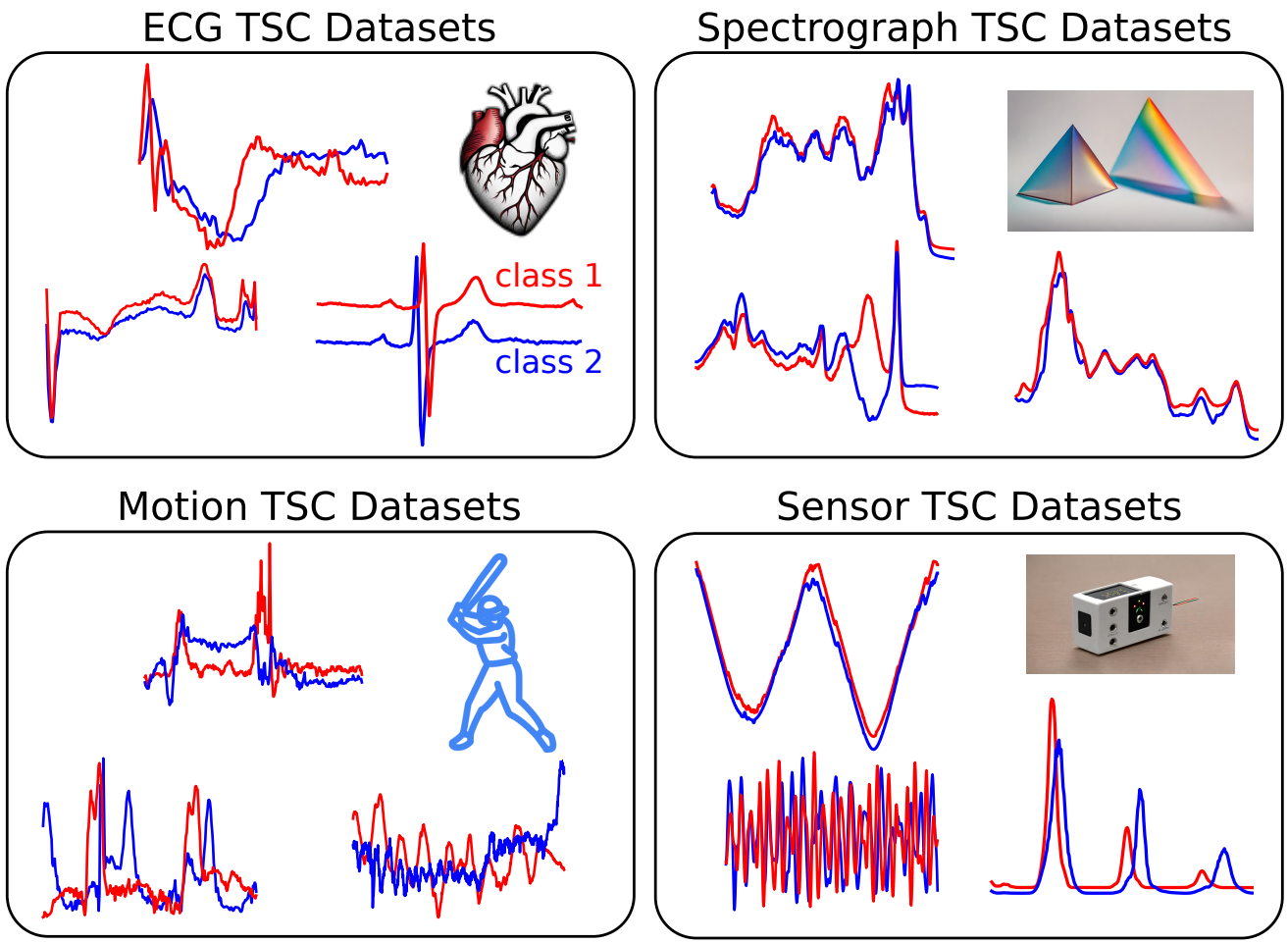


Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G. "Deep learning for time series classification using new hand-crafted convolution filters." In 2022 IEEE International Conference on Big Data.

Contents

Different Domains

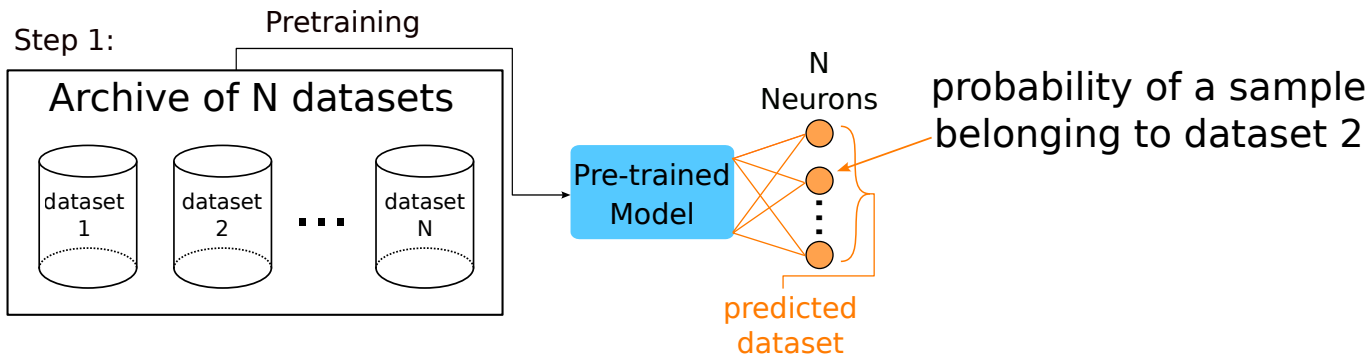
Usually, for each dataset we train a model from scratch to solve the classification task. However, we never leverage from the domain grouping



source : http://timeseriesclassification.com/
https://www.cs.ucr.edu/$\sim$Eeamonn/time_series_data_2018/

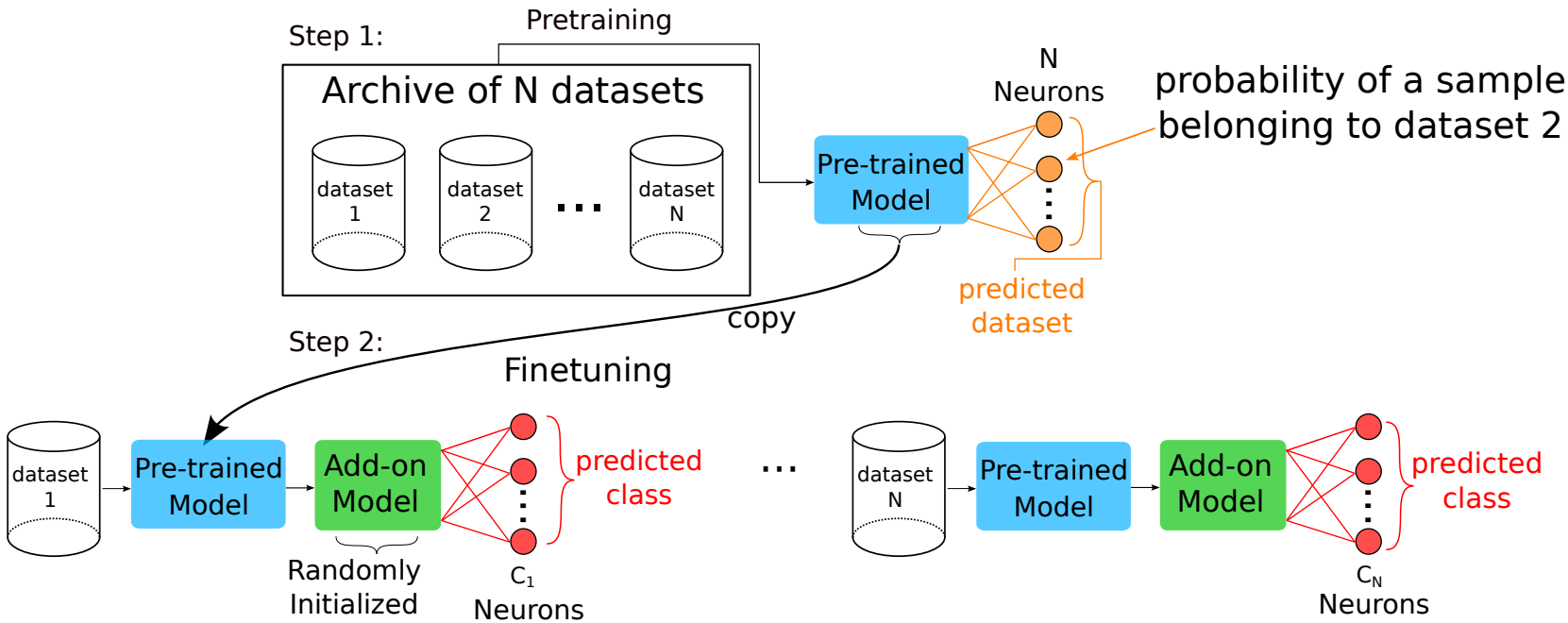
Learning from Different Datasets

We propose the following setup to be applied on each group of datasets per domain:



Learning from Different Datasets

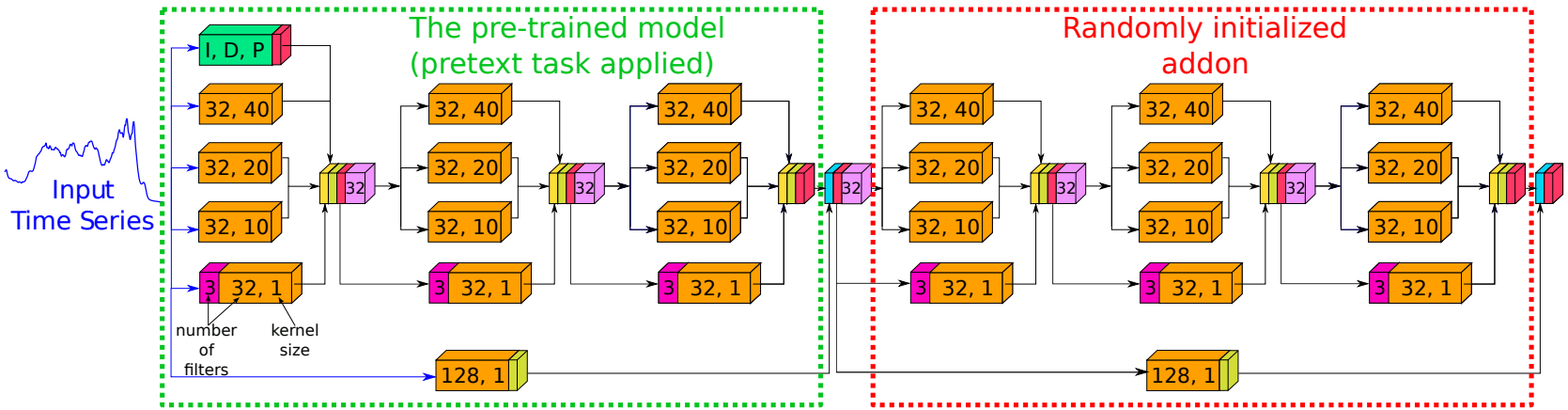
We propose the following setup to be applied on each group of datasets per domain:



- **Ismail-Fawaz, A.**, Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Splitting the Architecture

We utilize the state-of-the-art deep model for TSC: H-Inception

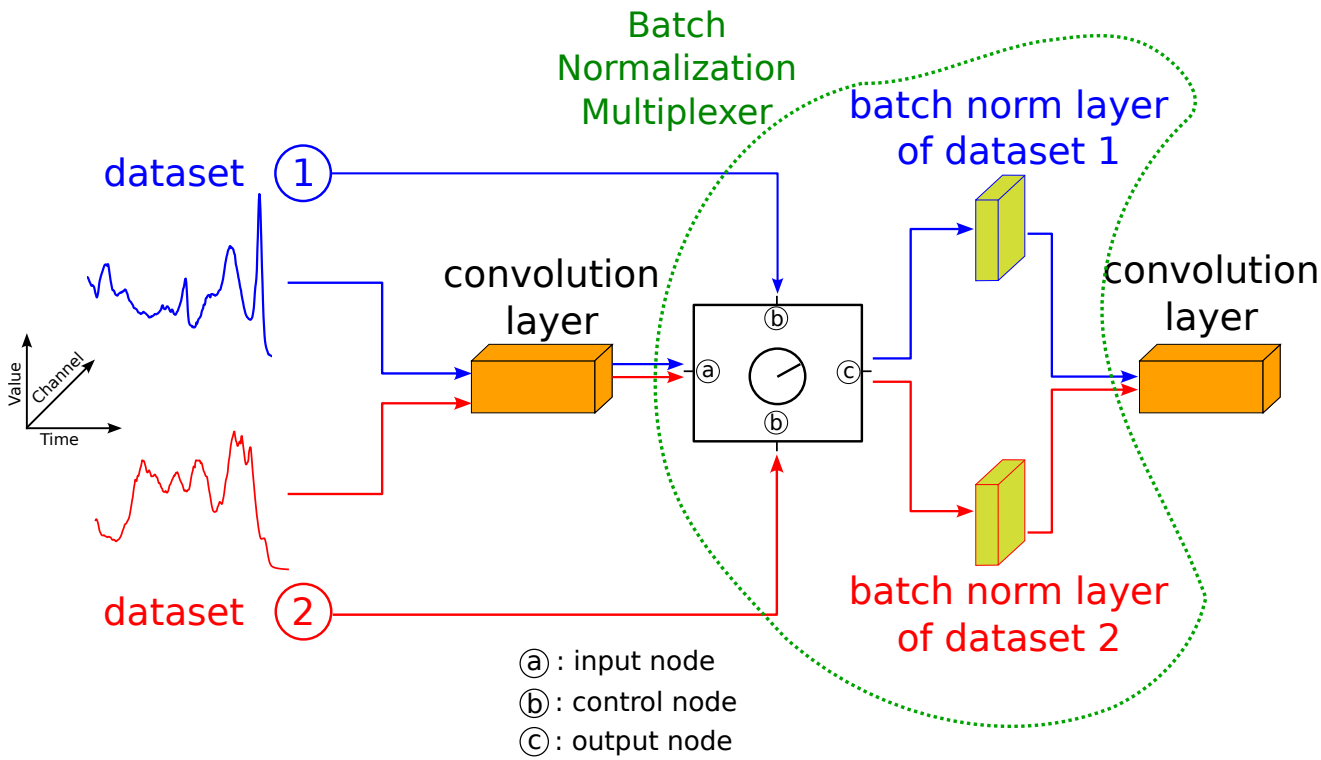


By ensembling different trained models, we propose the Pre-trained Hybrid InceptionTime (PHIT)

- **Ismail-Fawaz, A.**, Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Fixing the Batch Normalization Issue

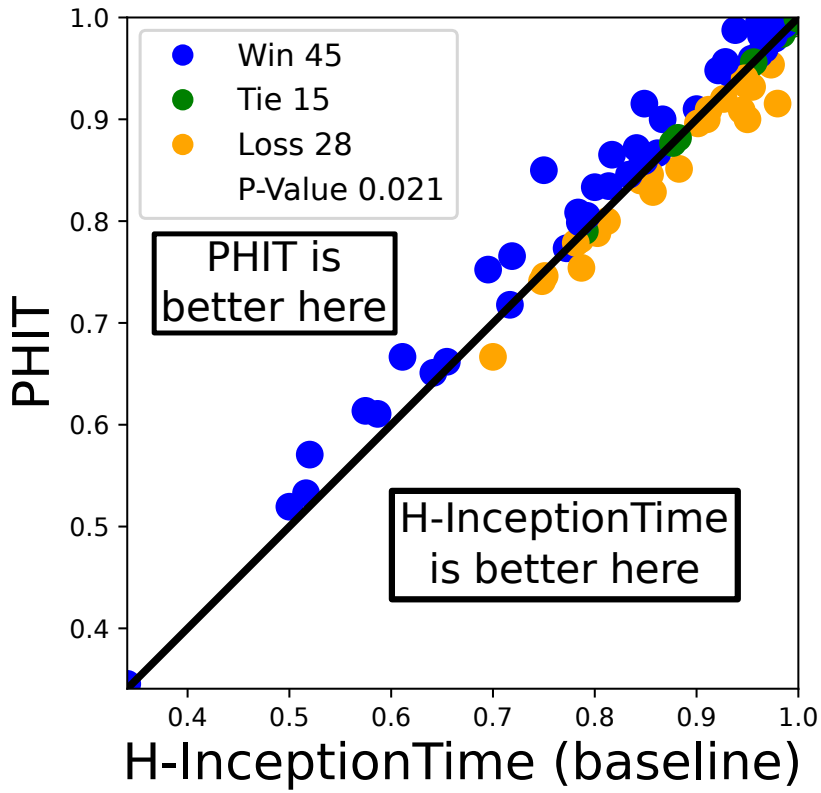
Every convolution layer is followed by a Batch Normalization that is fully dependent on the dataset distribution, given we train on multiple datasets we propose the Batch Normalization Multiplexer:



Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Results

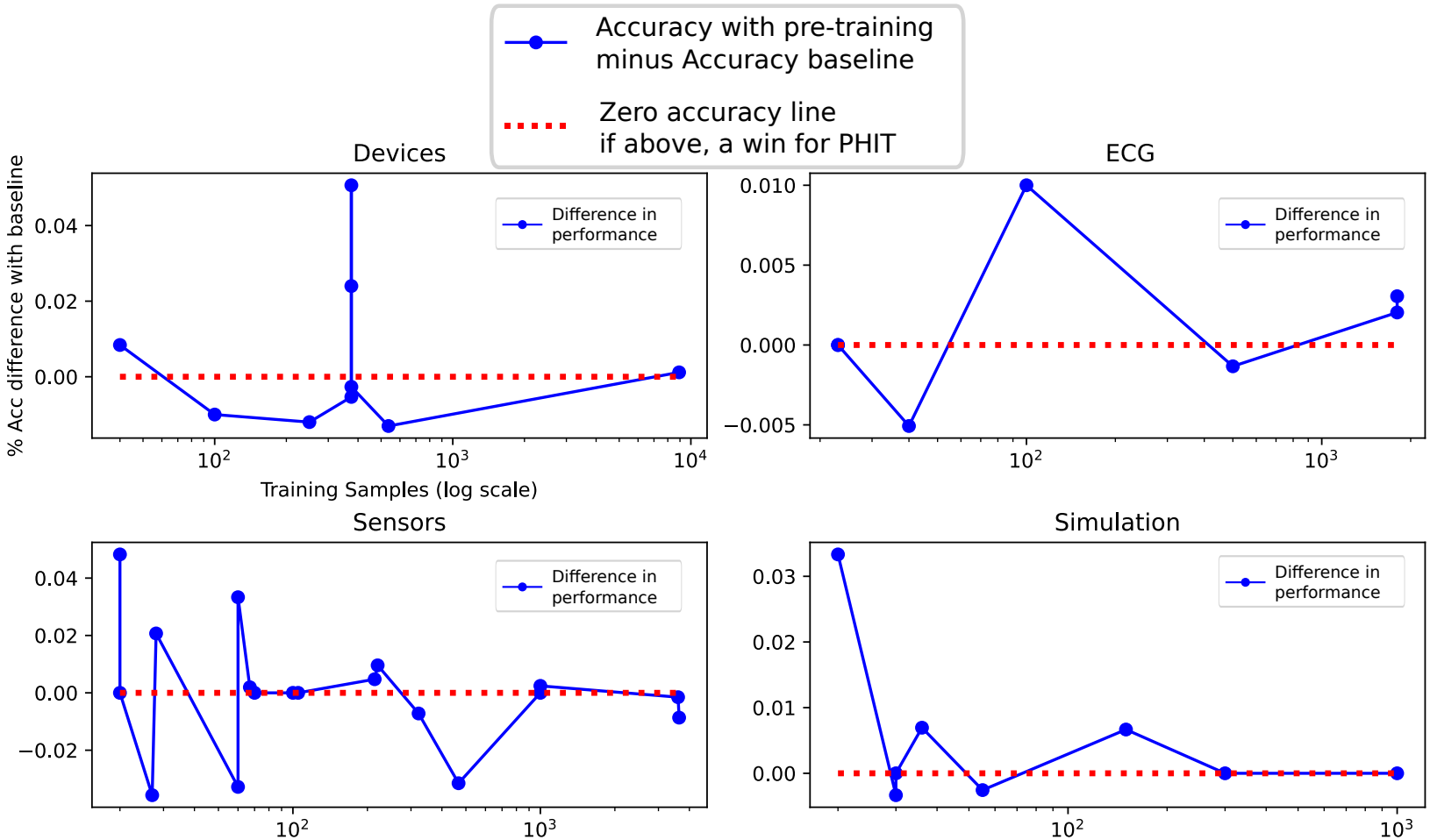
We utilize the UCR archive for our experiments:



- **Ismail-Fawaz, A.**, Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Analysis

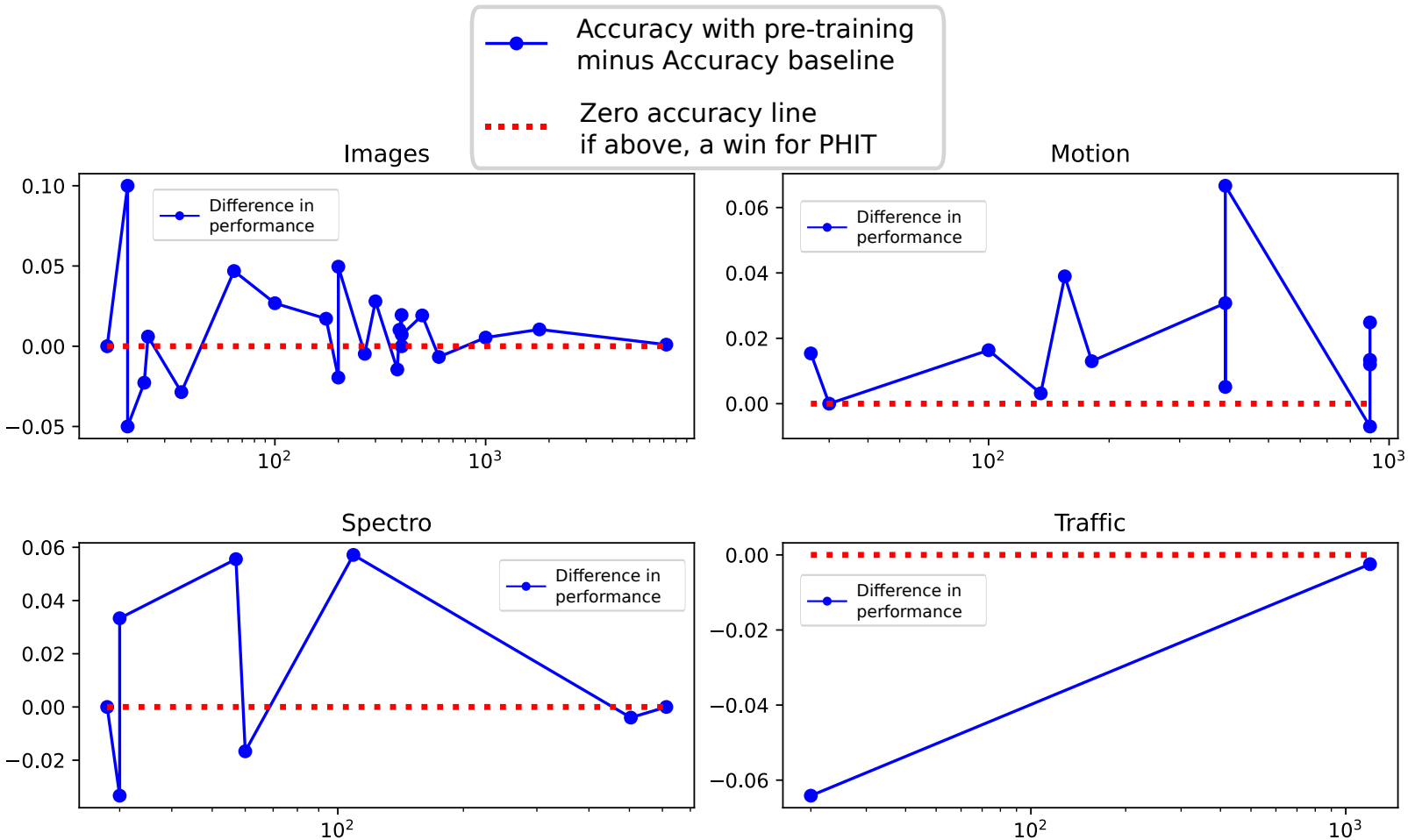
Presenting performance change with respect of training size



- **Ismail-Fawaz, A.**, Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Analysis

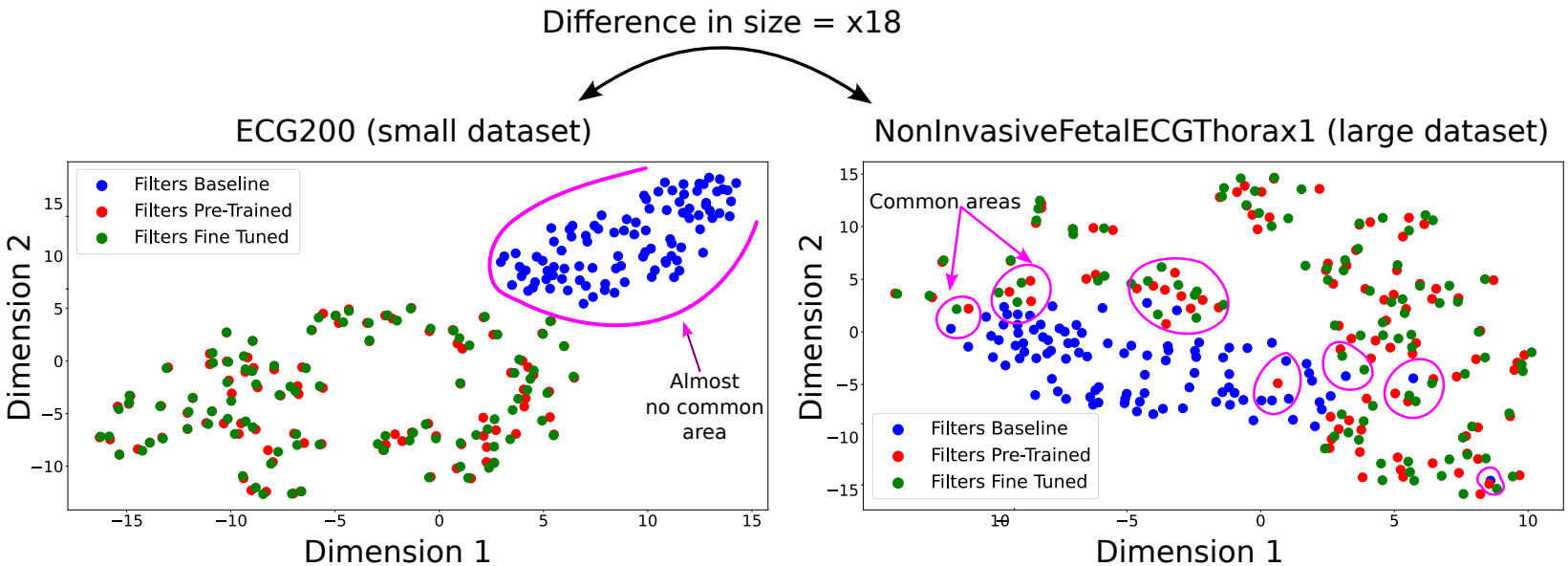
Presenting performance change with respect of training size



- **Ismail-Fawaz, A.**, Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Analysis

Visualization of the filters 2D space using t-SNE coupled with Dynamic Time Warping similarity measure: showcase that large dataset help small datasets



Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Other Ways to Construct PHIT

We do not claim that the method used in this work is the best. Other ways to construct PHIT:

- Train on the whole UCR at the same time regardless of the domain
- Domain Transfer i.e. train on one domain and fine tune on another
- Transfer learning within each domain
- Concatenate all possible classes and simply classify directly

Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J., Forestier, G. "Finding foundation models for time series classification with a pretext task." In 2024 PAKDD International Workshop on Temporal Analytics.

Contents

Other Work

Work that was not presented

- Self Representation Learning for Time Series Data [1]
- Decreasing Complexity of Deep Models for Time Series Data [2]
- Finding Stable Approaches for Comparing Models [3]

- **Ismail-Fawaz, Ali**, et al. "Enhancing Time Series Classification with Self-Supervised Learning." In 2023 International Conference on Agents and Artificial Intelligence.
- **Ismail-Fawaz, Ali** et al. "LITE: Light Inception with boosTing tEchniques for Time Series Classification." In 2023 IEEE International Conference on Data Science and Advanced Analytics/
- **Ismail-Fawaz, Ali**, et al. "An approach to multiple comparison benchmark evaluations that is stable under manipulation of the comparate set." arXiv preprint arXiv:2305.11921 (2023).

Other Work

Work that was not presented

- Self Representation Learning for Time Series Data [1]
- Decreasing Complexity of Deep Models for Time Series Data [2]
- Finding Stable Approaches for Comparing Models [3]

Hot Topics

- Data Augmentation for Time Series Data
- Self Representation Learning
- Foundation Models for Time Series Data
- Transformers for Time Series Classification
- Transfer Learning, studying conditions for a dataset to benefit from transfer learning
- Multivariate Time Series Classification

- **Ismail-Fawaz, Ali**, et al. "Enhancing Time Series Classification with Self-Supervised Learning." In 2023 International Conference on Agents and Artificial Intelligence.
- **Ismail-Fawaz, Ali** et al. "LITE: Light Inception with boosTing tEchniques for Time Series Classification." In 2023 IEEE International Conference on Data Science and Advanced Analytics/
- **Ismail-Fawaz, Ali**, et al. "An approach to multiple comparison benchmark evaluations that is stable under manipulation of the comparate set." arXiv preprint arXiv:2305.11921 (2023).

Contents

Reproducibility

- All our codes are publically available on our MSD-IRIMAS github page starting 2022.
- All DL4TSC models are now implemented in aeon [1], a python package dedicated to time series machine learning. Non-deep learning papers are also implemented now in aeon.

[1] https://github.com/aeon-toolkit/aeon

Takeway


Stefano Berretti


Maxime Devanne


Germain Forestier


Ali Ismail-Fawaz


Hassan Ismail Fawaz


Jonathan Weber

- Deep Learning for Time Series Analysis: hot topic open for innovation
- Finding Foundation Models for TSC is still a challenging task: the goal should be to address generalization
- We believe that the UCR is the best way to benchmark different approaches in TSC, however more application datasets should be used for specific use-case
- **Personal Web-page: https://hadifawaz1999.github.io/**
- **Contact: ali-el-hadi.ismail-fawaz@uha.fr**
- **Team Web-page: https://msd-irimas.github.io/**