Finding Foundation Models for Time Series Classification with a PreText Task

International Workshop on Temporal Analytics @PAKDD2024

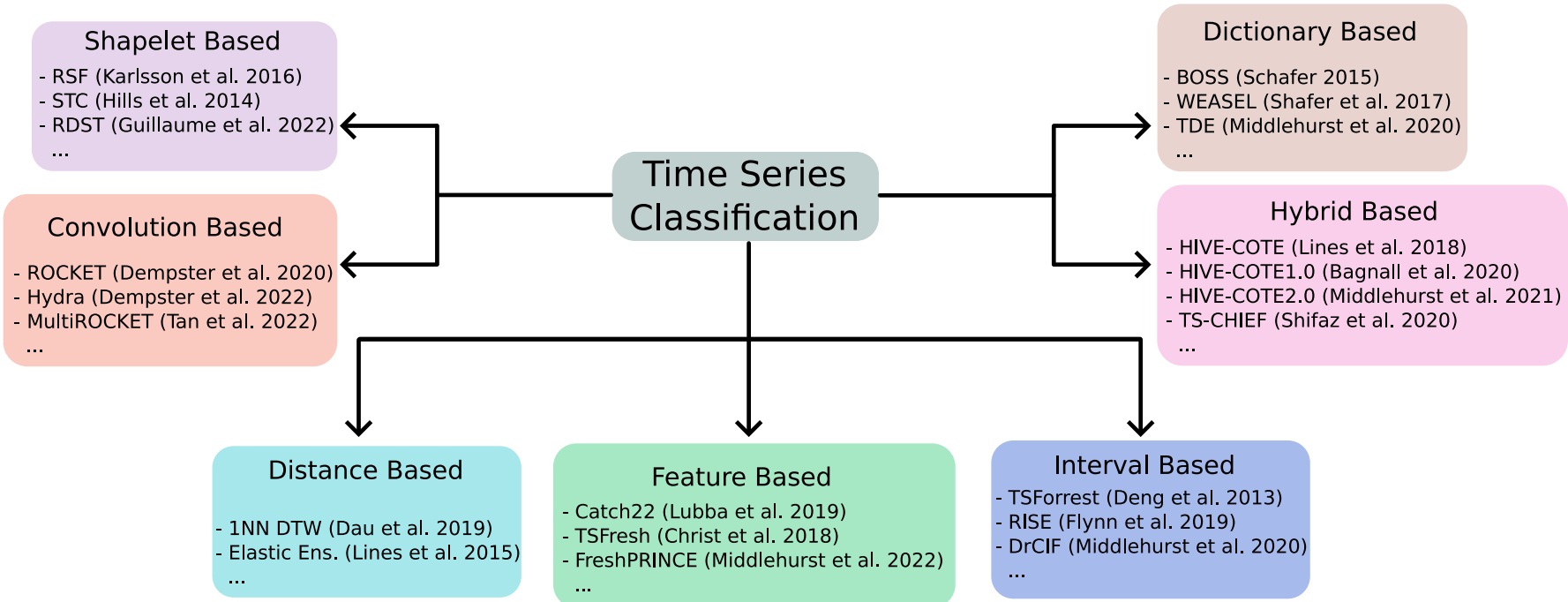
Ali Ismail-Fawaz[1], Maxime Devanne[1], Stefano Berretti[2], Jonathan Weber[1] and Germain Forestier[1,3]

[1]IRIMAS, Université Haute-Alsace, Mulhouse, France
[2]MICC, University of Florence, Florence, Italy
[3]DSAI, Monash University, Melbourne Australia

November 10, 2024

Time Series Classification (TSC)

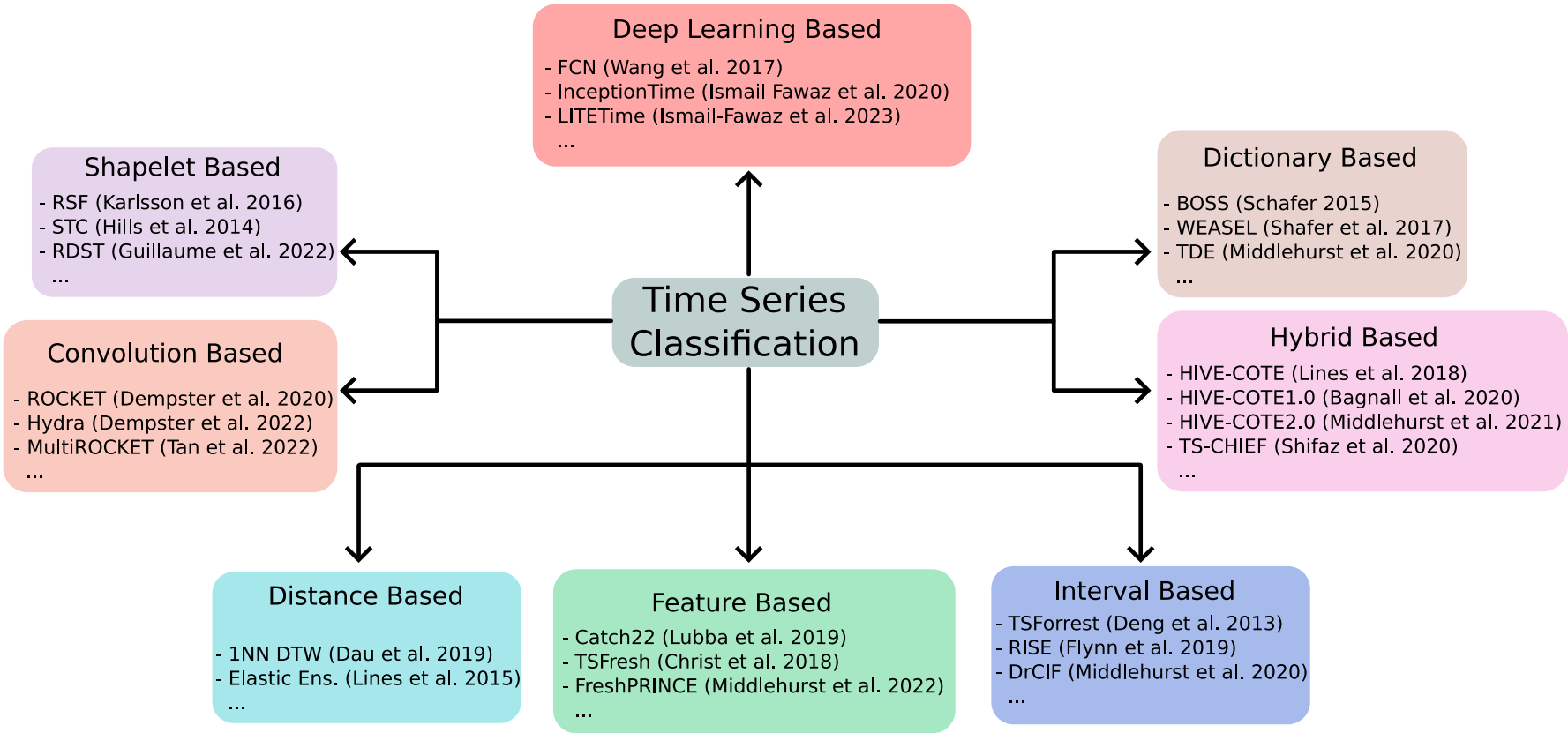
Various approaches are available for tackling TSC [1]



[1] 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 (TSC)

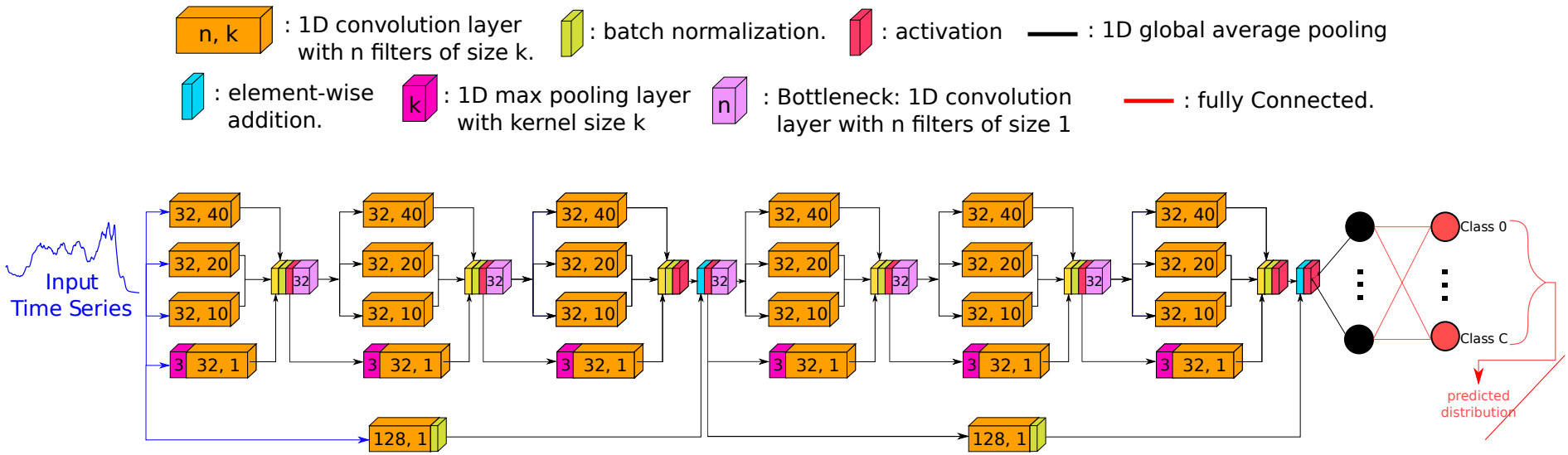
Various approaches are available for tackling TSC [1], we address this task using Deep Learning methods.



[1] 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.

Deep Learning for TSC

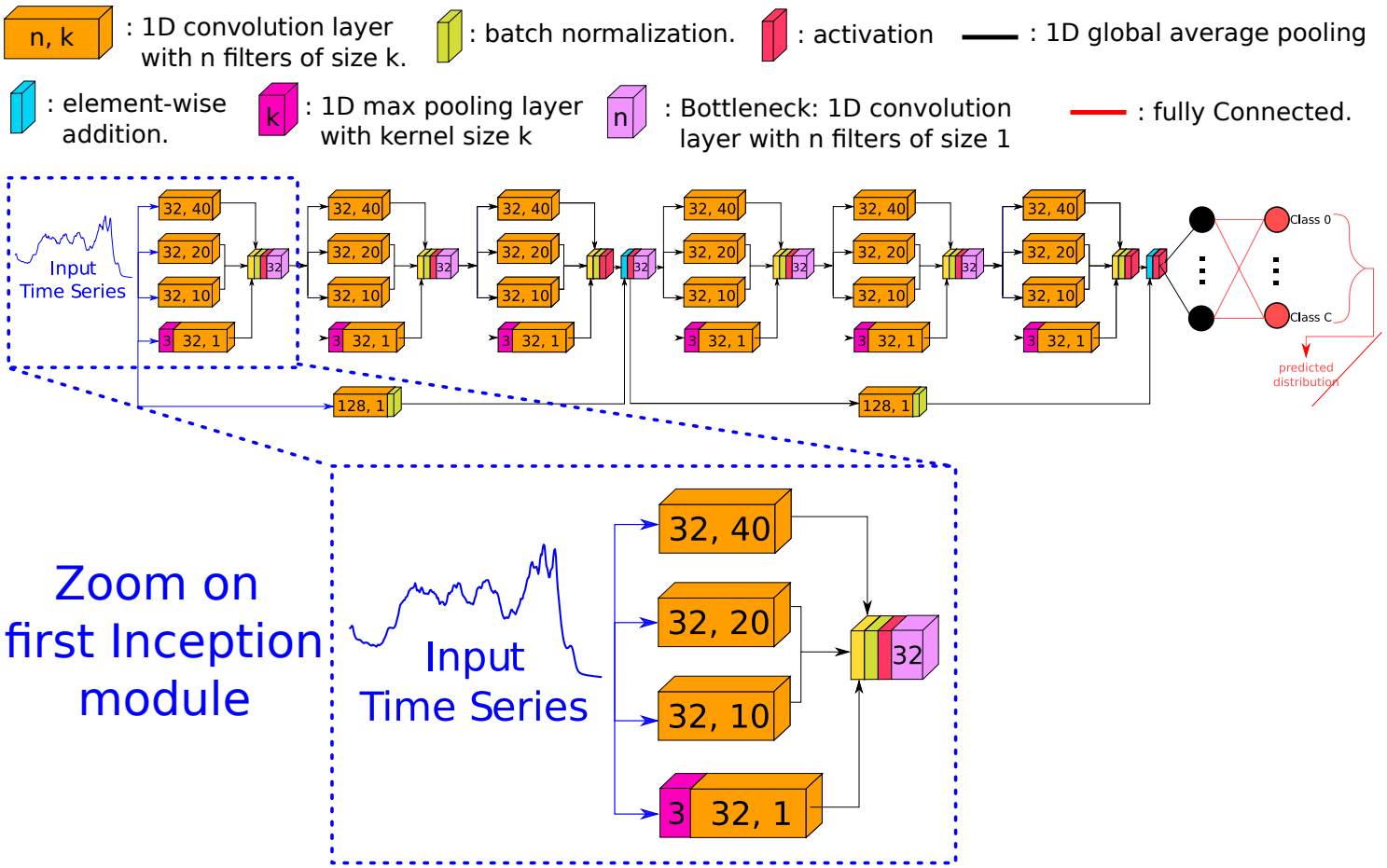
Until 2022, InceptionTime[2] (ensemble of five Inception models) was the state-of-the-art deep models for TSC



[2] Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." Data Mining and Knowledge Discovery 34.6 (2020): 1936-1962.

Deep Learning for TSC

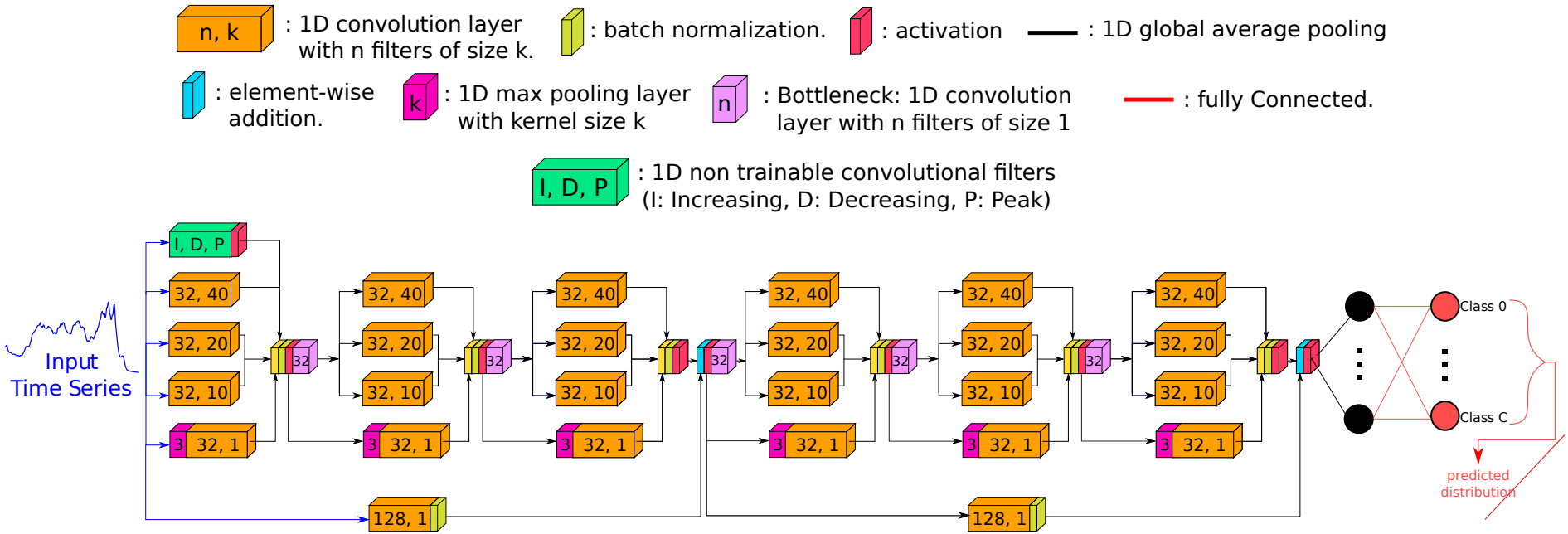
Until 2022, InceptionTime[2] (ensemble of five Inception models) was the state-of-the-art deep models for TSC



[2] Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." Data Mining and Knowledge Discovery 34.6 (2020): 1936-1962.

Deep Learning for TSC

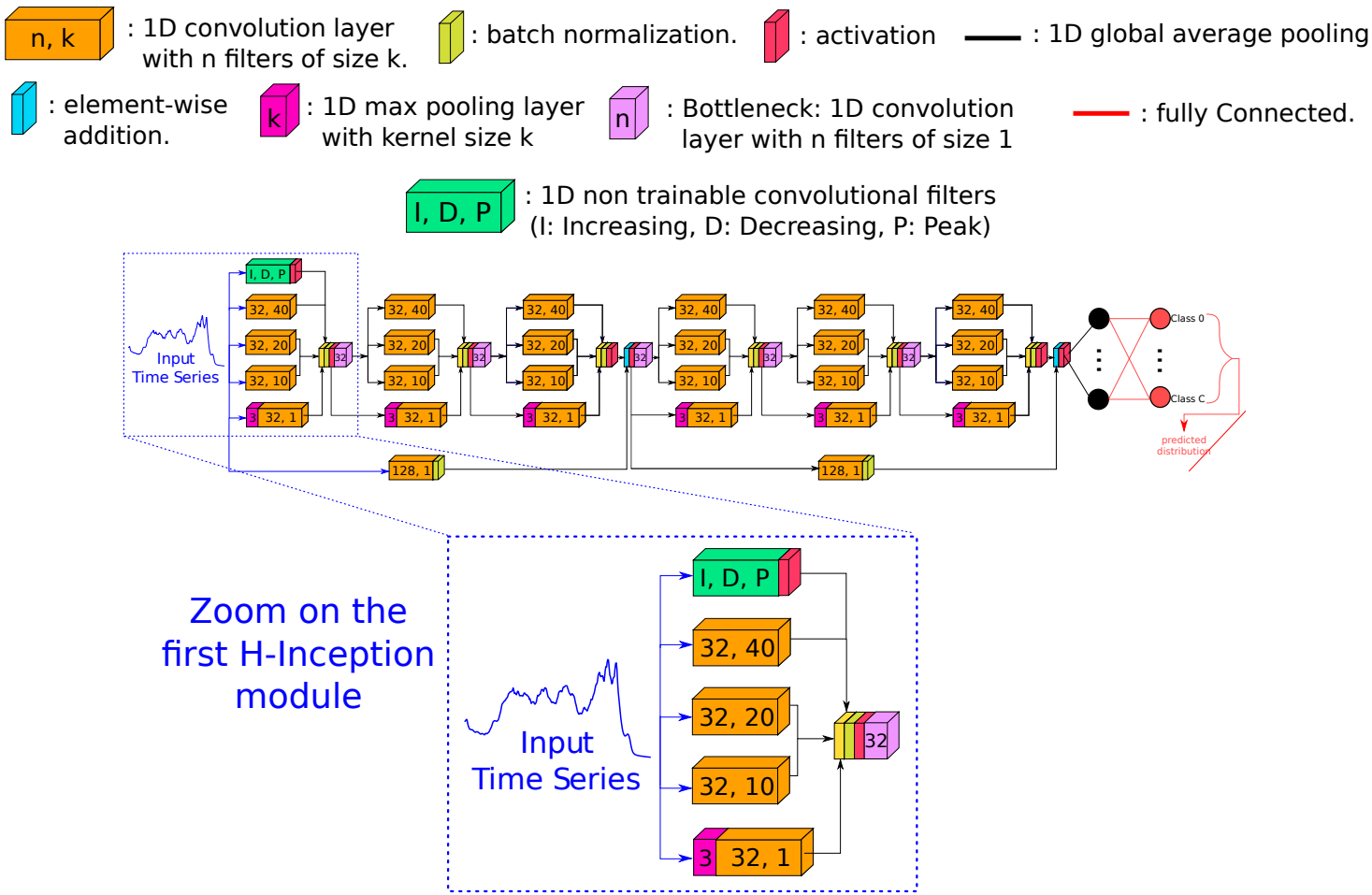
The recent state-of-the-art CNN model for TSC: Hybrid InceptionTime[3] (H-InceptionTime), an ensemble of five H-Inception models:



[3] Ismail-Fawaz, Ali, et al. "Deep learning for time series classification using new hand-crafted convolution filters." 2022 IEEE International Conference on Big Data (Big Data). IEEE, 2022.

Deep Learning for TSC

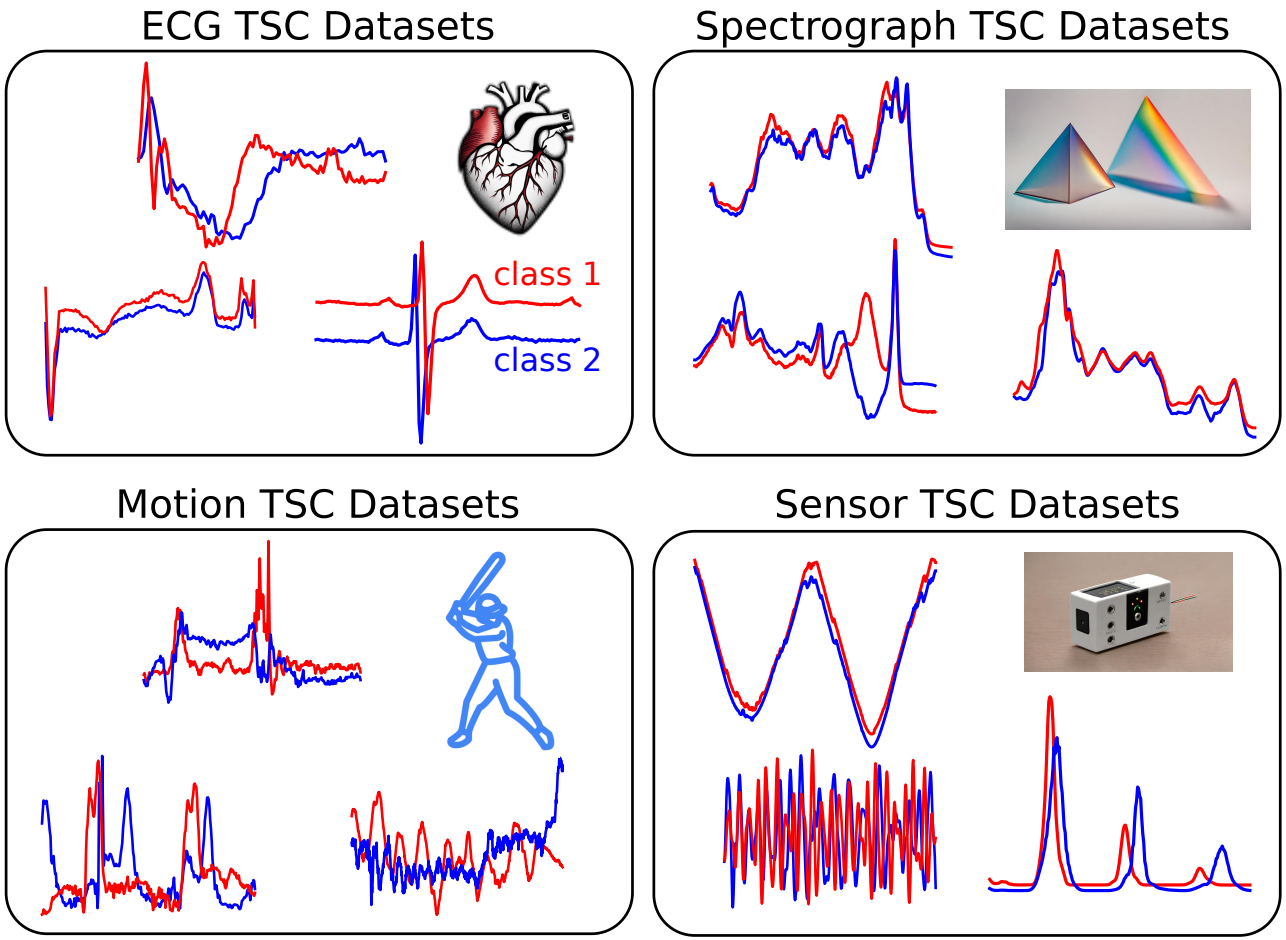
The recent state-of-the-art CNN model for TSC: Hybrid InceptionTime[3] (H-InceptionTime), an ensemble of five H-Inception models:



[3] Ismail-Fawaz, Ali, et al. "Deep learning for time series classification using new hand-crafted convolution filters." 2022 IEEE International Conference on Big Data (Big Data). IEEE, 2022.

Different Domains

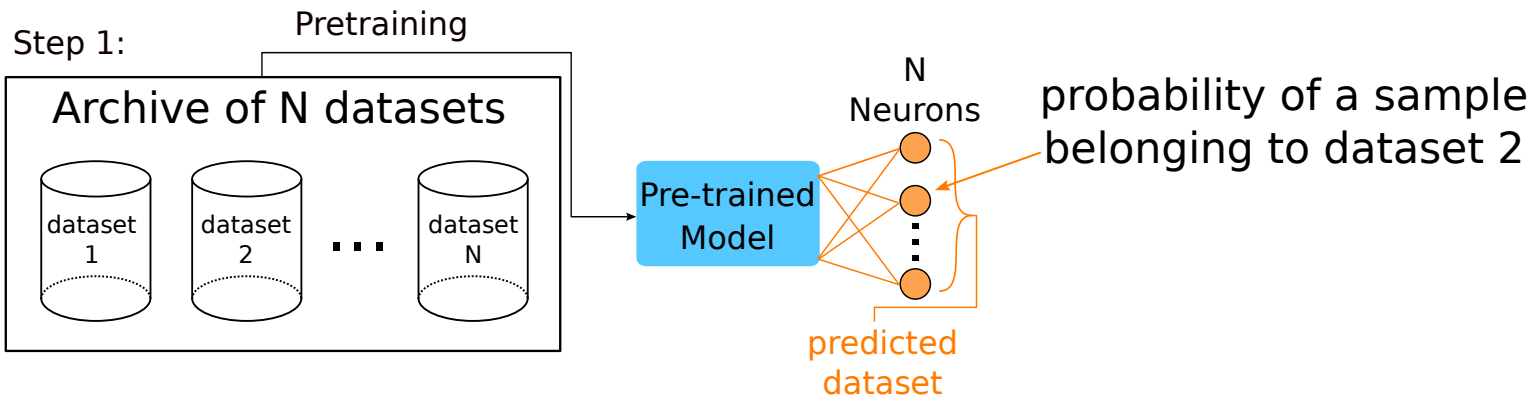
Usually, for each dataset we train a model from scratch to solve the classification task. However, we never leverage from the domain grouping [4]



[4] https://www.cs.ucr.edu/∼Eeamonn/time_series_data_2018/ http://timeseriesclassification.com/.

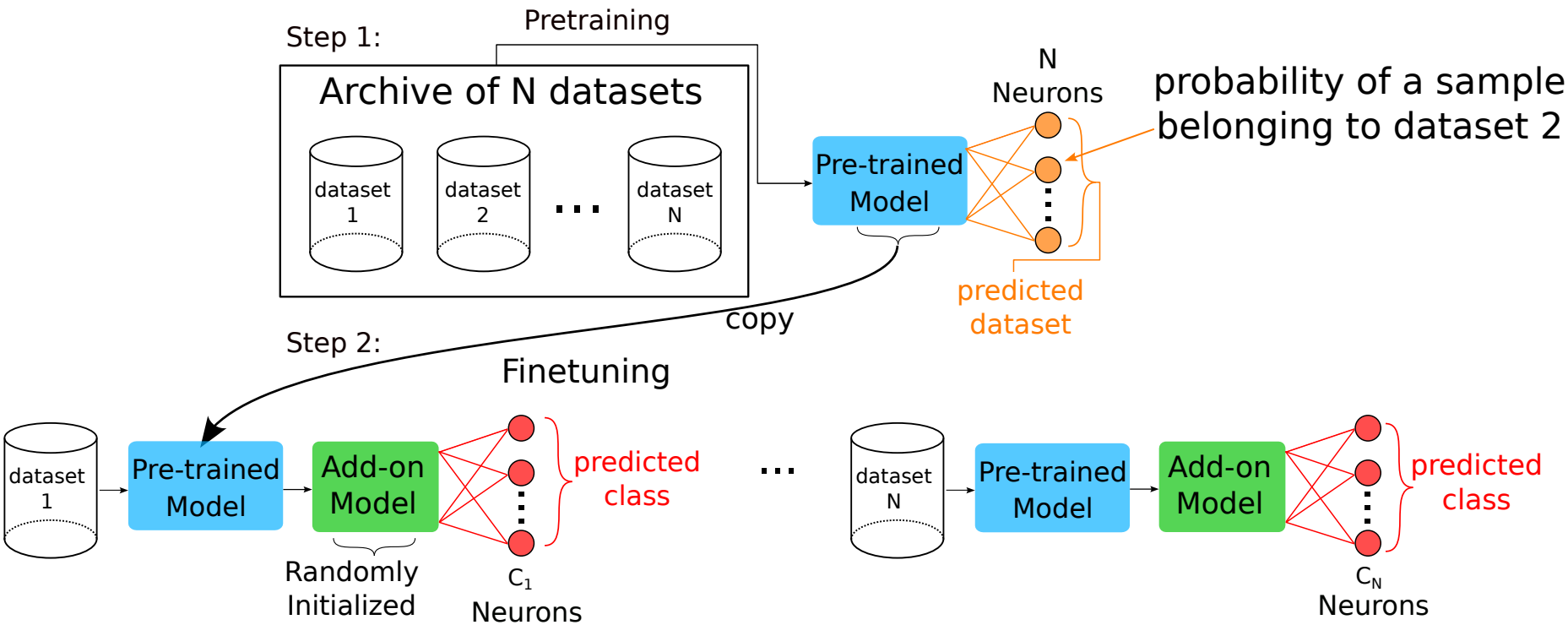
Learning from Different Datasets

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



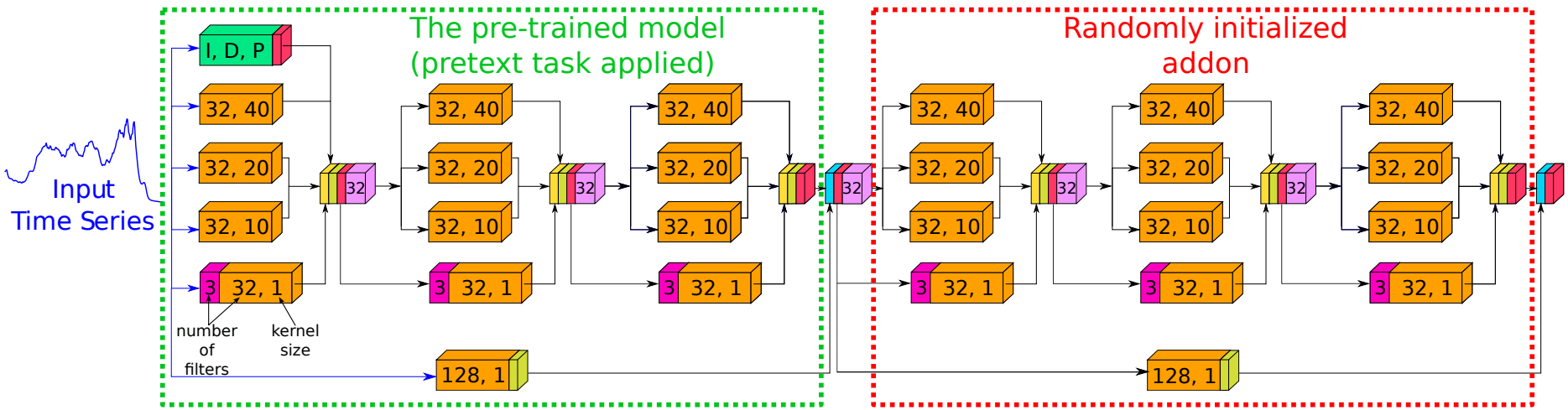
Learning from Different Datasets

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



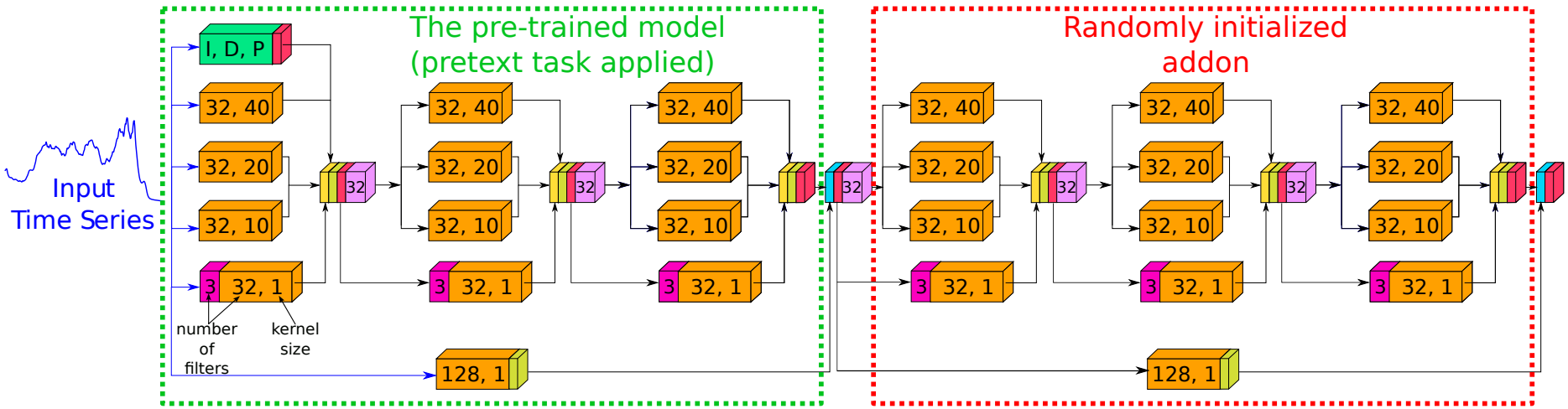
Splitting the Architecture

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



Splitting the Architecture

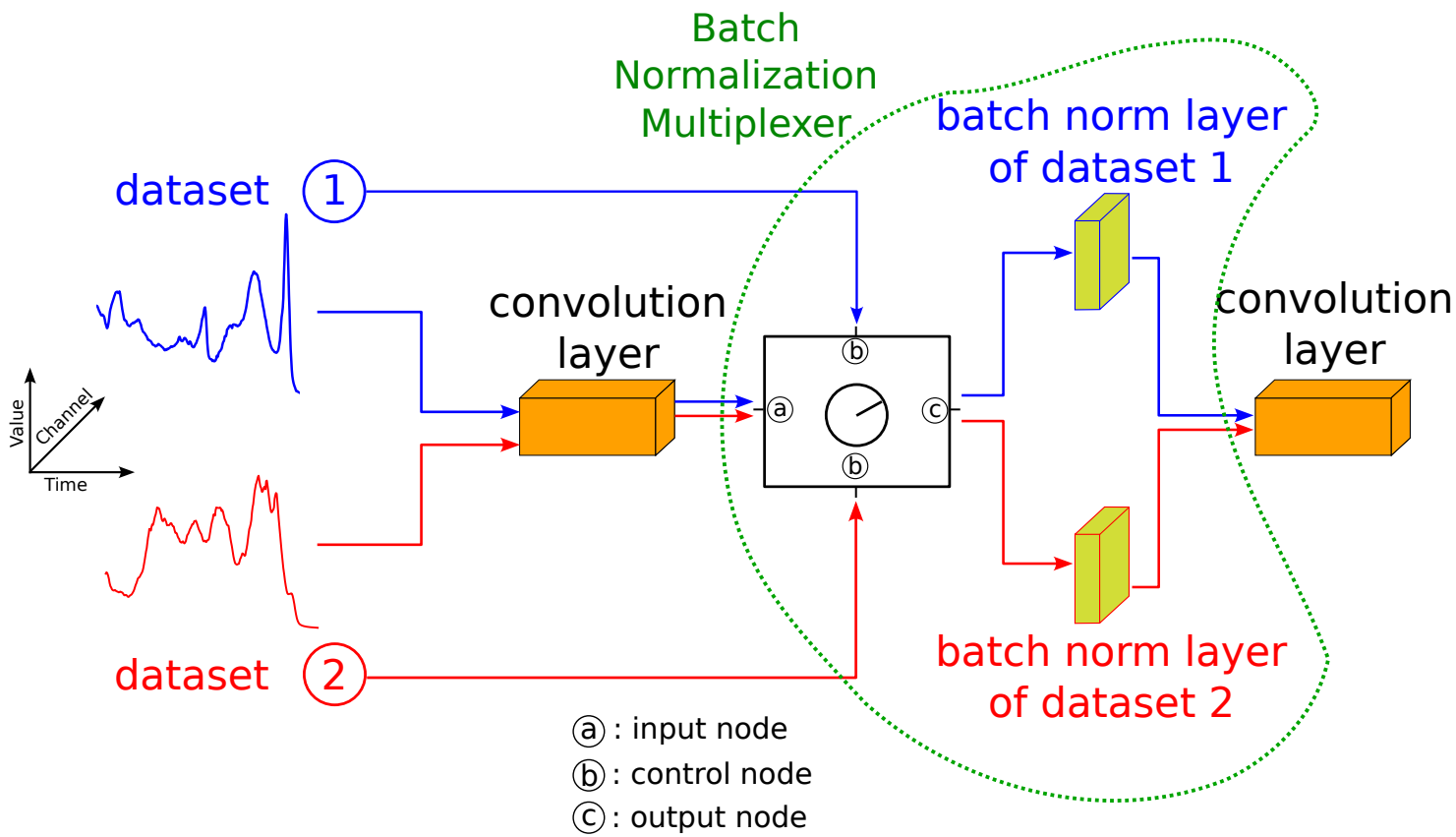
We utilize the state-of-the-art deep model for TSC: H-Inception



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

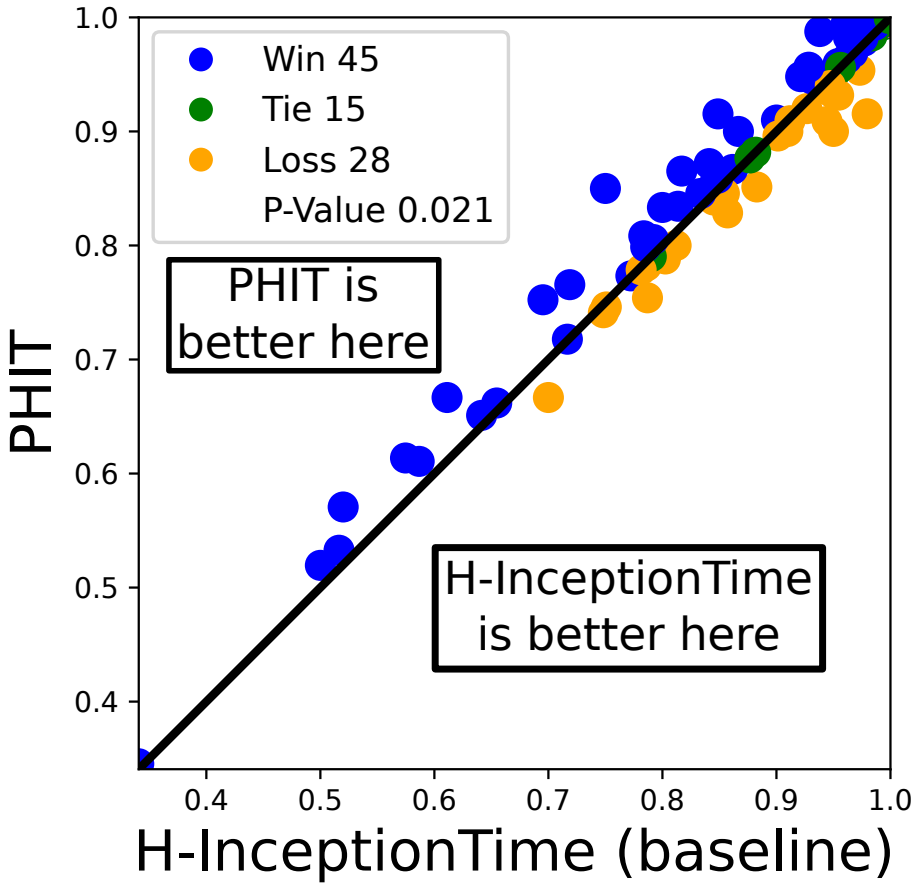
Fixing the Batch Normalization Issue

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:



Results

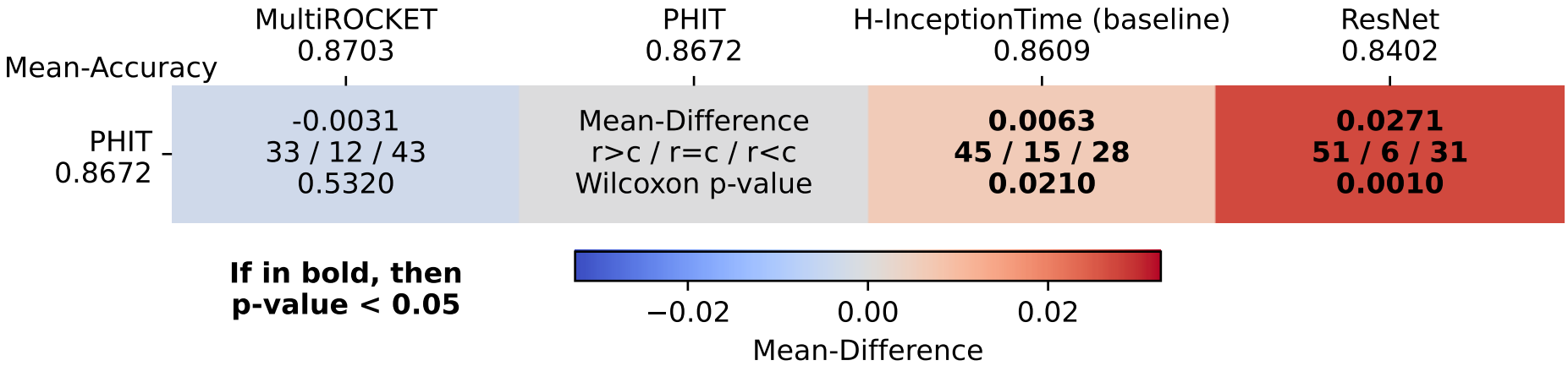
We utilize the UCR archive [5] for our experiments:



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

Results

Using the Multi-Comparison Matrix[6] to compare PHIT with H-InceptionTime, ResNet and MultiROCKET



- ResNet[7] and H-InceptionTime: two state-of-the-art deep models for TSC
- MultiROCKET[8]: a non-deep model state-of-the-art for TSC

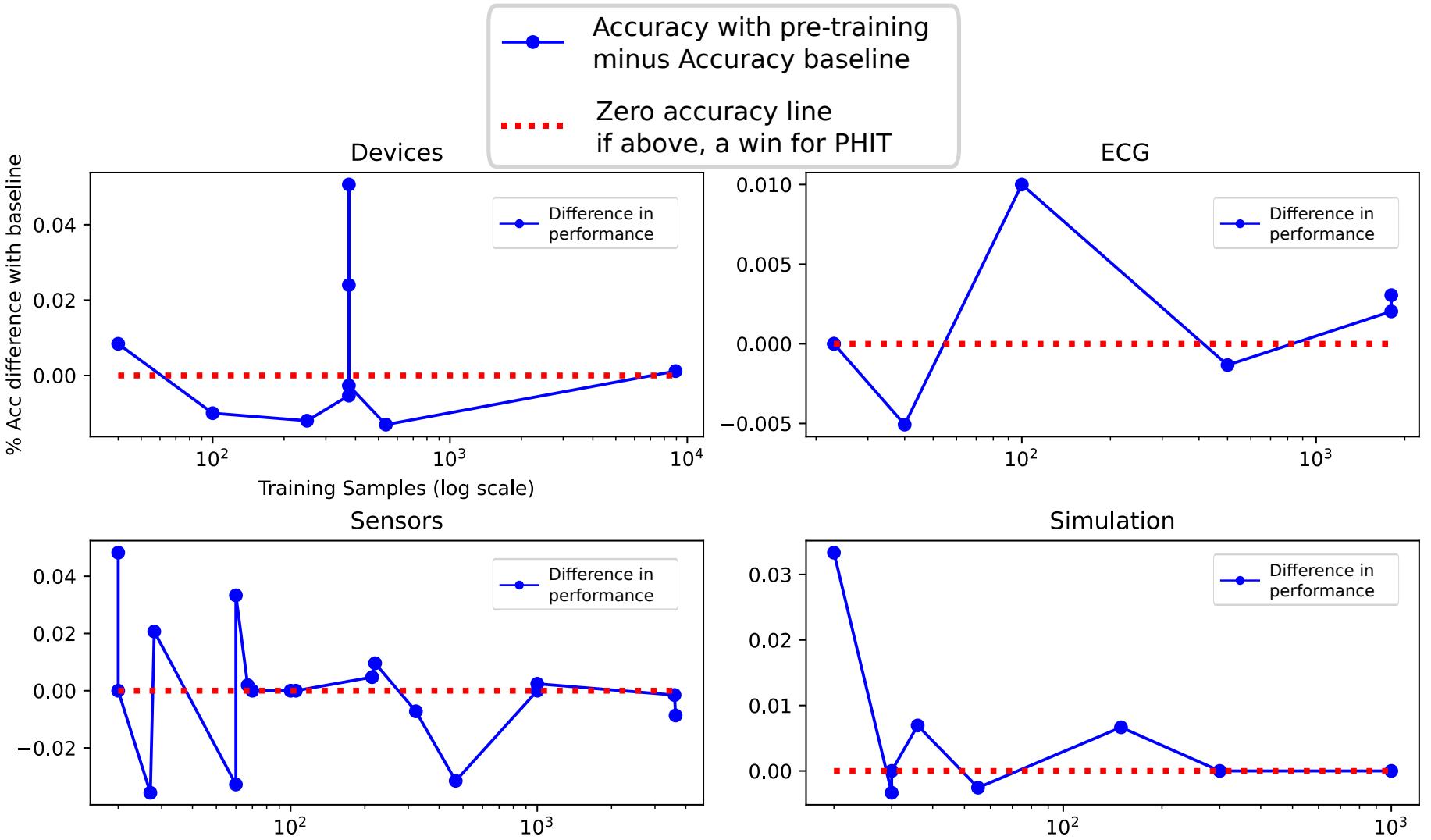
[6] Ismail-Fawaz, Ali, et al. "An approach to multiple comparison benchmark evaluations that is stable under manipulation of the comparate set." arXiv preprint arXiv:2305.11921 (2023).

[7] Wang, Zhiguang, Weizhong Yan, and Tim Oates. "Time series classification from scratch with deep neural networks: A strong baseline." 2017 International joint conference on neural networks (IJCNN). IEEE, 2017.

[8] Tan, Chang Wei, et al. "MultiRocket: multiple pooling operators and transformations for fast and effective time series classification." Data Mining and Knowledge Discovery 36.5 (2022): 1623-1646.

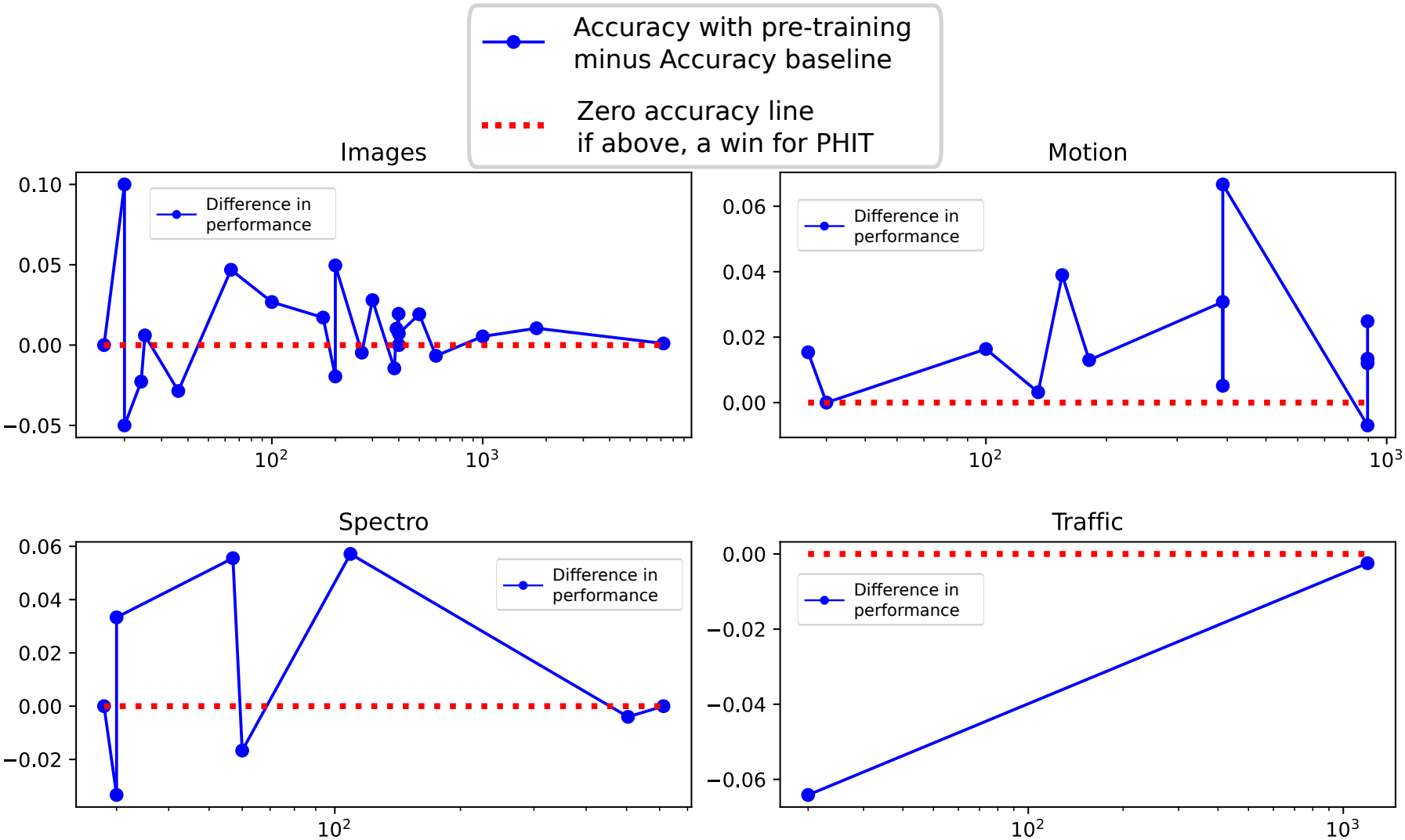
Analysis

Presenting performance change with respect of training size



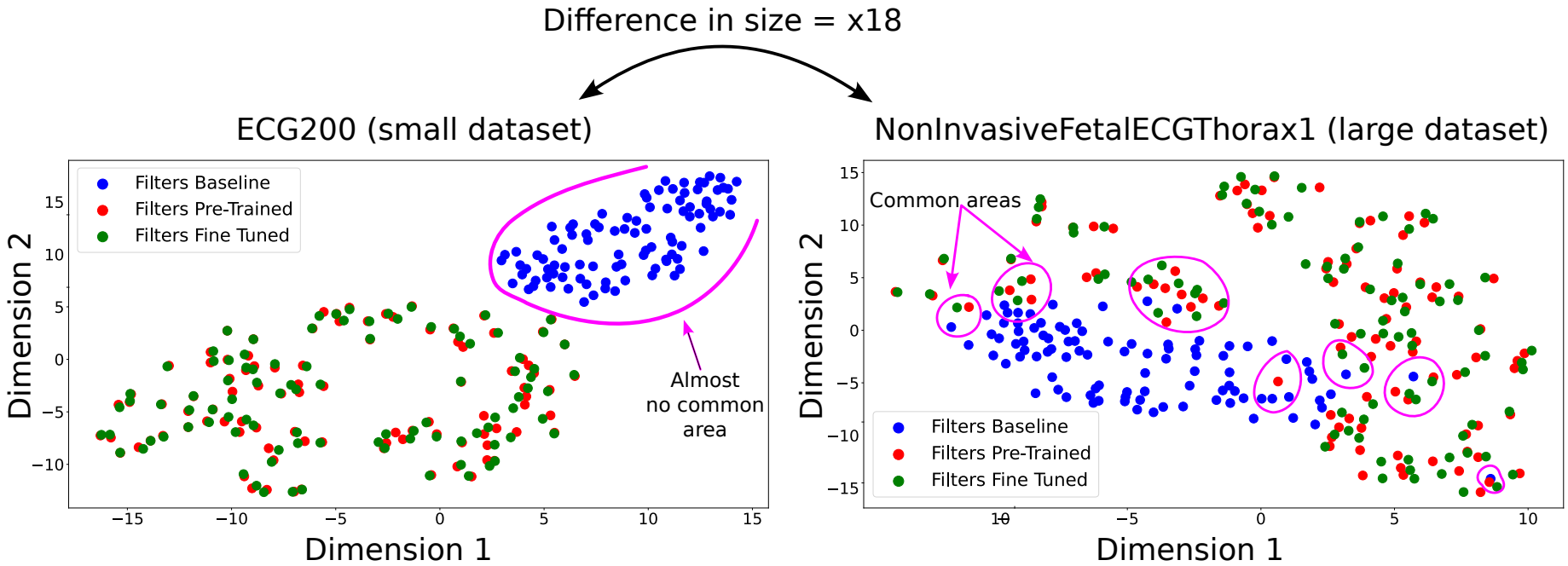
Analysis

Presenting performance change with respect of training size



Analysis

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

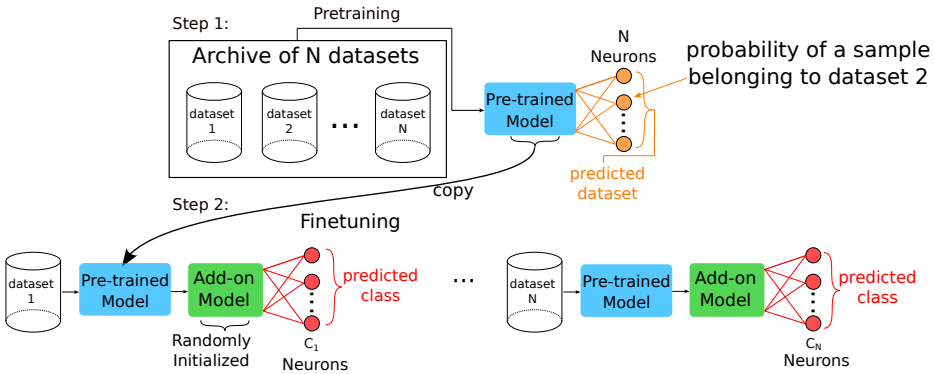


Other Ways to Construct PHIT

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

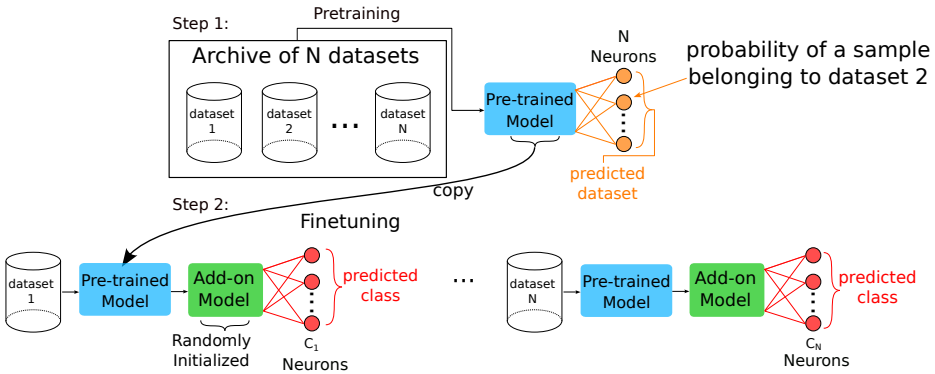
- Train on the whole UCR at the same time regardless of the domain
- Domain Transfer i.e. train on one domain and fine tune on another
- Transfer learning within each domain
- Concatenate all possible classes and simply classify directly

Takeaway Message



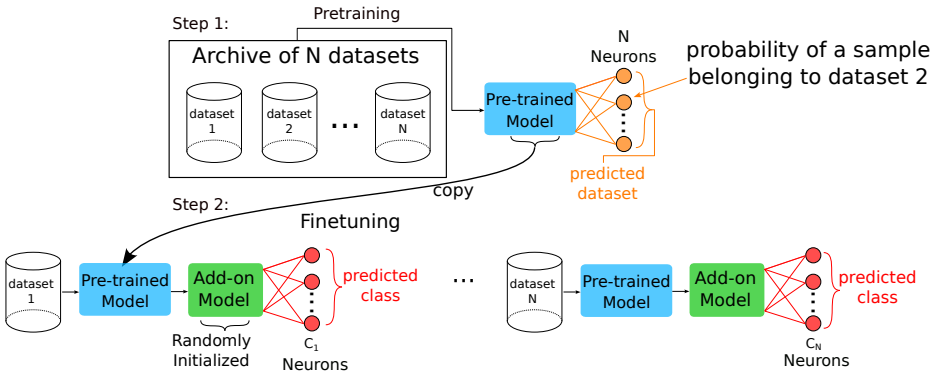
- Finding Foundation Models for Time Series Classification with a PreText Task

Takeaway Message



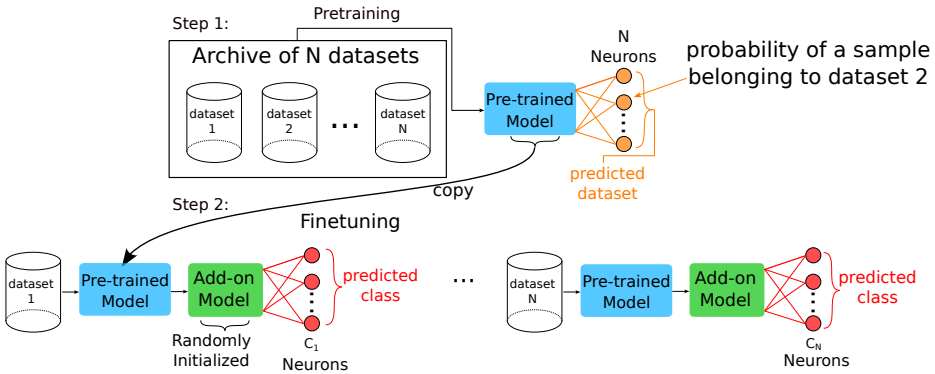
- Finding Foundation Models for Time Series Classification with a PreText Task
- Foundation Models are a hot topic, not only in time series, and a challenging one

Takeaway Message



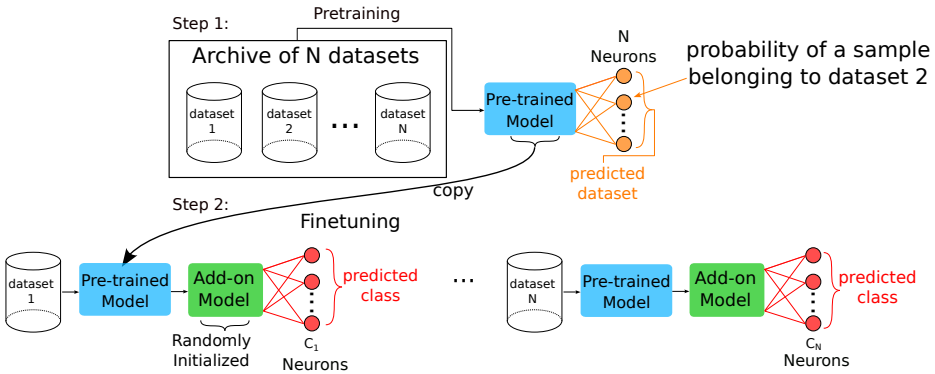
- Finding Foundation Models for Time Series Classification with a PreText Task
- Foundation Models are a hot topic, not only in time series, and a challenging one
- PHIT is a foundation model defined for each unique domain available for time series data in the UCR archive

Takeaway Message



- Finding Foundation Models for Time Series Classification with a PreText Task
- Foundation Models are a hot topic, not only in time series, and a challenging one
- PHIT is a foundation model defined for each unique domain available for time series data in the UCR archive
- PHIT outperforms the baseline model (without pre-training) on almost all domains (7 out of 8 domains)

Takeaway Message



- Finding Foundation Models for Time Series Classification with a PreText Task
- Foundation Models are a hot topic, not only in time series, and a challenging one
- PHIT is a foundation model defined for each unique domain available for time series data in the UCR archive
- PHIT outperforms the baseline model (without pre-training) on almost all domains (7 out of 8 domains)
- e-email: **ali-el-hadi.ismail-fawaz@uha.fr**
- website: **https://hadifawaz1999.github.io**
- team website: **https://msd-irimas.github.io**
- github repo: **https://github.com/MSD-IRIMAS/DomainFoundationModelsTSC**

Question: How does the PHIT model adapt to data distribution changes within a single domain?

- Utilizes a Batch Normalization Multiplexer (BNM) to handle data from multiple distributions.
- Each convolution layer applies separate batch normalization for each dataset.
- This allows dynamic adaptation to varying distributions over time within the same domain.
- The BNM aligns batch normalization processes with the dataset-specific characteristics.
- Adaptation is achieved by linking each sample to its corresponding normalization process.

Impact of Pre-training Dataset Size

Question: What is the impact of pre-training dataset size on PHIT's performance?

- Pre-training combines data from multiple datasets to generalize across different sizes.
- The pretext task involves predicting the dataset origin, which aids in handling variance in dataset sizes.
- Generalized features learned during pre-training are robust to size variations of the datasets.
- Fine-tuning adjusts the model to specific characteristics and scales of each dataset.
- This dual-phase training helps optimize performance while accommodating dataset size impacts.

Preservation of Dataset-Specific Features

Question: How does PHIT preserve unique dataset features while benefiting from cross-domain pre-training?

- Uses a Batch Normalization Multiplexer to maintain the ability to adapt to specific dataset characteristics.
- Pre-training with a general model is combined with an add-on model for fine-tuning specific tasks.
- Ensures that unique features of each dataset are preserved during the fine-tuning phase.
- Allows the model to effectively handle dataset-specific nuances and complexities.
- The approach balances generalization with the need to adapt to particular dataset details.

Modifications to InceptionTime Architecture

Question: What enhancements are made to the InceptionTime model in PHIT?

- Integrates hand-crafted convolutional filters and a Batch Normalization Multiplexer.
- Designed to handle diverse challenges of different time series classification (TSC) domains.
- BNM allows effective handling of different distributions, crucial for varied domains.
- Hand-crafted filters capture domain-specific patterns, enhancing domain generalization.
- Bottleneck layers reduce dimensionality and parameter count, optimizing the model for efficiency.

Avoiding Over-parameterization in PHIT

Question: How does PHIT avoid over-parameterization and ensure practical efficiency?

- Utilizes a hybrid architecture dividing InceptionTime into pre-trained and fine-tuning sections.
- Balances complexity and performance, managing the model's parameter count effectively.
- Bottleneck layers within the architecture help reduce the number of parameters.
- Ensures efficient processing and avoids the pitfalls of over-parameterization.
- Maintains performance while being computationally efficient for practical applications.