

## CSC 8740 - Advanced Data Mining

### Homework Assignment 3

Time series classification is an important supervised learning task in data mining. The overall goal is to identify the label (or the target feature) of a time series instance as coming from one of possibly many predefined classes/groups. This is of course done using labeled training data. That is, in this setting we conduct supervised learning, where the different time series sources are considered known.

Recent interest in time series classification resulted in many open-source implementations of time series classification algorithms. These include `sktime`<sup>1</sup>, `tslearn`<sup>2</sup>, and `aeon`<sup>3</sup> libraries. In this homework, you are free to use and experiment with any of these libraries. Note here that `aeon` is the preferred library for this homework but feel free to use the other ones.

In each of those libraries, you will find benchmark dataset integration/downloaders. See for example:

- `tslearn`.datasets module – [https://tslearn.readthedocs.io/en/stable/gen\\_modules/tslearn.datasets.html](https://tslearn.readthedocs.io/en/stable/gen_modules/tslearn.datasets.html)
- `Sktime`'s datasets module – [https://www.sktime.net/en/stable/api\\_reference/datasets.html](https://www.sktime.net/en/stable/api_reference/datasets.html)
- `Aeon`'s tutorial – [https://www.aeon-toolkit.org/en/stable/examples/datasets/load\\_data\\_from\\_web.html](https://www.aeon-toolkit.org/en/stable/examples/datasets/load_data_from_web.html)

Please take a look at the time series dataset downloaders and make use of them in this homework.

In the literature, the **time series classification** algorithms are often **classified** in **eight broad** and at times overlapping categories. These include:

1. **Distance-based Classifiers** (time domain and differential)
2. **[Canonical] Feature-based Classifiers**
3. **Dictionary-based Classifiers**
4. Shapelet-based Classifiers
5. Interval-based Classifiers
6. **Convolution-based Classifiers**
7. Deep learning-based Classifiers
8. Hybrid Classifiers

In this homework, you will be training **univariate time series classifiers**. Please note that the above mentioned libraries have implemented a number of classification algorithms but as expected, libraries do not necessarily cover a comprehensive list of algorithms.

**Task 1:** Read the Bagnall et al. (2017) paper [<https://arxiv.org/abs/1602.01711>]. (repeat this step until you understand) and Faouzi (2022) for further reference on time series classification method categorization [<https://inria.hal.science/hal-03558165/document>]

**Task 2:** Pick **10 univariate time series datasets** from UCR/UAE archives for classification (See <http://www.timeseriesclassification.com/> for their information). You are expected to use the downloaders from libraries to avoid formatting issues but you can choose to download and load

---

<sup>1</sup> <https://www.sktime.net/>

<sup>2</sup> <https://tslearn.readthedocs.io>

<sup>3</sup> <https://www.aeon-toolkit.org>

## CSC 8740 - Advanced Data Mining

### Homework Assignment 3

your data as well. You can pick relatively smaller datasets to make sure your training times are reasonable.

**Task 3:** Pick **four classifiers** from four different categories mentioned above (namely, one from each). For **each classifier**, train a model **using ten datasets** you have picked in the earlier step. When training, please use **10-fold cross validation** and get the **mean accuracy**.

**Task 4:** Compare the performance of each selected model using **critical difference diagrams**.

See an example library for CD diagrams: <https://github.com/hfawaz/cd-diagram> and another one from aoe toolkit library

([https://www.aeon-toolkit.org/en/stable/examples/visualisation/plotting\\_results.html](https://www.aeon-toolkit.org/en/stable/examples/visualisation/plotting_results.html))

#### Notes:

- You are expected to use the Python programming language and use the mentioned libraries above. If you find another library or would like to implement your own classifiers, please feel free, but consult with the instructor first.
- You can show your work using a Jupyter notebook. Please state clearly which time series classification category-algorithm you are choosing.
- Some datasets are large and some algorithms are computationally very expensive. Considering you will need to train ten models for each of the ten datasets using four different algorithms, it may take a considerable amount of time. Please be aware of this potential time requirement and choose your algorithms (hyperparameters) and datasets wisely if needed.