An Empirical Evaluation of Early Time-Series Classification Algorithms

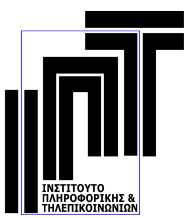
Evgenios Kladis (eukladis@iit.demokritos.gr), Charilaos Akasiadis, Evangelos Michelioudakis, Elias Alevizos, Alexandros Artikis
CER GROUP, SKEL LAB of Institute of Informatics and Telecommunications, NCSR DEMOKRITOS

SIMPLIFY WORKSHOP 2021

Tuesday 23rd March, 2021

Introduction

- Big data and the evolution of technology (e.g. the Internet of Things) increase our capacity for storing and using information in order to solve various problems.
- Although complex problems emerge, their solutions can bring about great change in the human activity.
- Among such problems is the **early detection of phenomena**, a challenge for many research fields.
- *Early Time-Series Classification (ETSC)* methods can contribute to this end.



skel
THE AI LAB

Contributions

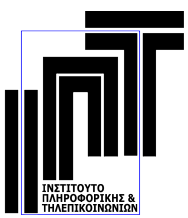
- Categorize the ETSC methods of the related literature.
- Bundled these methods in a framework.
- Utilized a publicly available dataset library (UCR) as well as two new datasets from the maritime and biology field.
- Evaluated the algorithms using various performance metrics.

Provide a publicly available benchmarking framework for ETSC algorithms, at https://github.com/Eukla/ETS.

Applications

ETSC can be used in a wide variety of research fields:

- Medical (Early Arrhythmia Detection, Predictive Medicine, Tumour Growth Detection).
- Maritime (Early Ship movement detection, for better naval route organization and ship smuggling events).
- Earthquake detection.
- Energy supply optimization.
- And many more.



skel
THE AI LAB

Existing Surveys on ETSC

- Most related work is based on plain time-series classification or time-series forecasting.
- Single related survey: Gupta et al. 2020 which conducts a review on Early Time-Series Classification methods.
- We do more: implement, test, evaluate.

Algorithm Organization

The methods concerning ETSC were grouped in 3 categories based on the core feature of their approaches:

- Sub-sequence
  - EDSC, *Xing et al. 2011*
  - ECEC, *Lv et al. 2019*
  - TEASER, *Schäfer et al. 2020*
  - DTEC, *Yao et al. 2019*
  - MCFEC, *He et al. 2015*
  - ECDIRE, *Morti et al. 2017*
- Clustering
  - ECTS, *Xing et al. 2012*
  - Trigger, *Dachraoui et al. 2015*
- Artificial Neural Networks
  - MLSTM-FCN, *Karim et al. 2019*
  - MDDNN, *Huang et al. 2018*

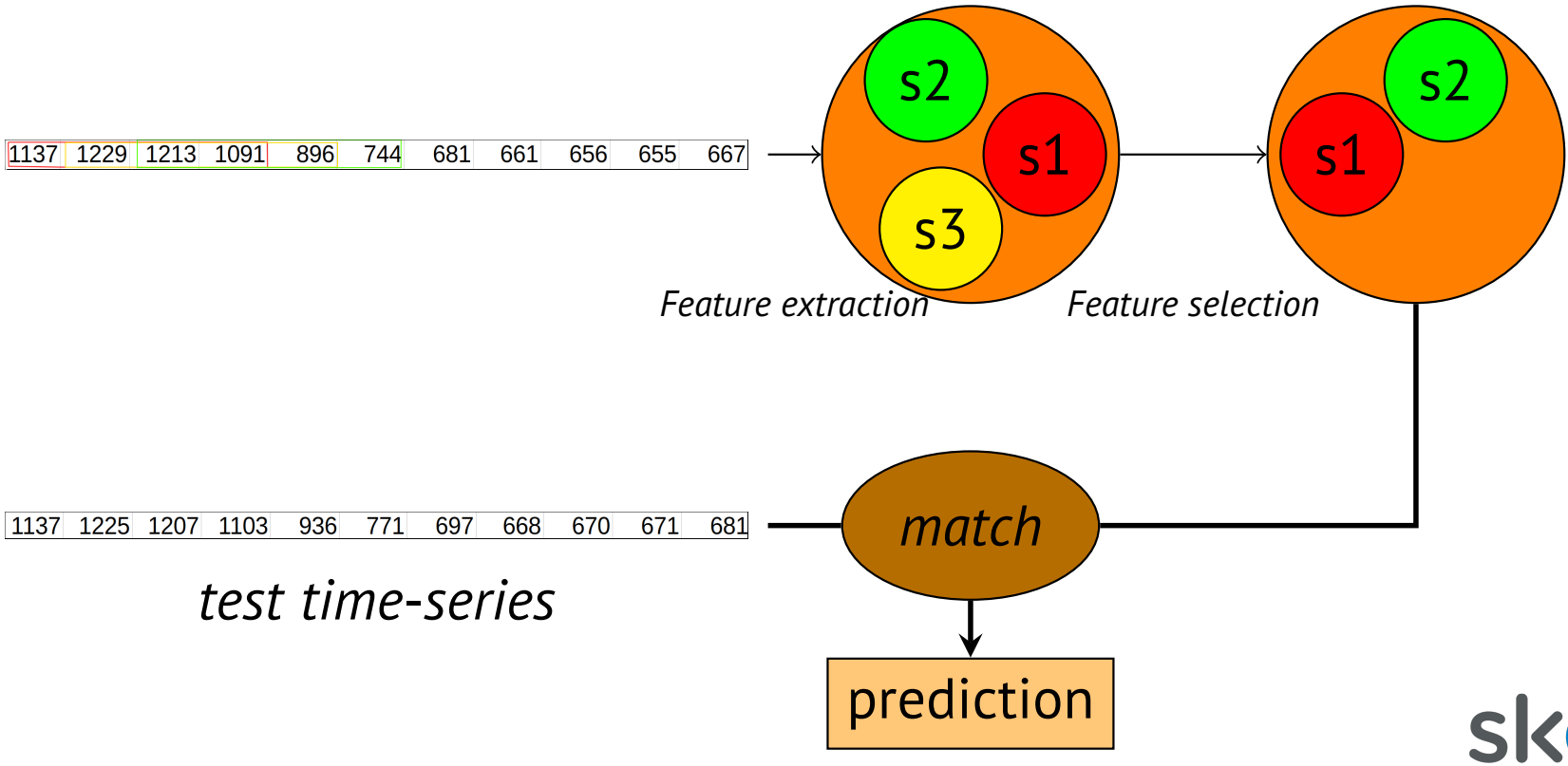
Algorithm Organization

The methods concerning ETSC were grouped in 3 categories based on the core feature of their approaches:

- Sub-sequence
  - EDSC, *Xing et al. 2011*
  - ECEC, *Lv et al. 2019*
  - TEASER, *Schäfer et al. 2020*
  - DTEC, *Yao et al. 2019*
  - MCFEC, *He et al. 2015*
  - ECDIRE, *Morti et al. 2017*
- Clustering
  - ECTS, *Xing et al. 2012*
  - Trigger, *Dachraoui et al. 2015*
- Artificial Neural Networks
  - MLSTM-FCN, *Karim et al. 2019*
  - MDDNN, *Huang et al. 2018*

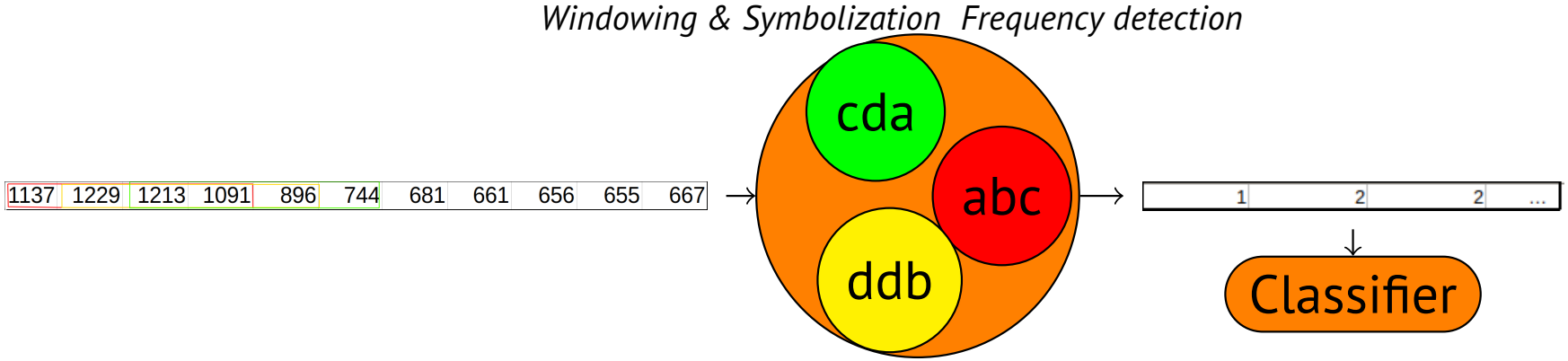
Sub-sequence I

- Algorithms with this logic, cut windows from the time-series, called *sub-series* and use it to aid to the classification task.
- The 1st approach *EDSC*:



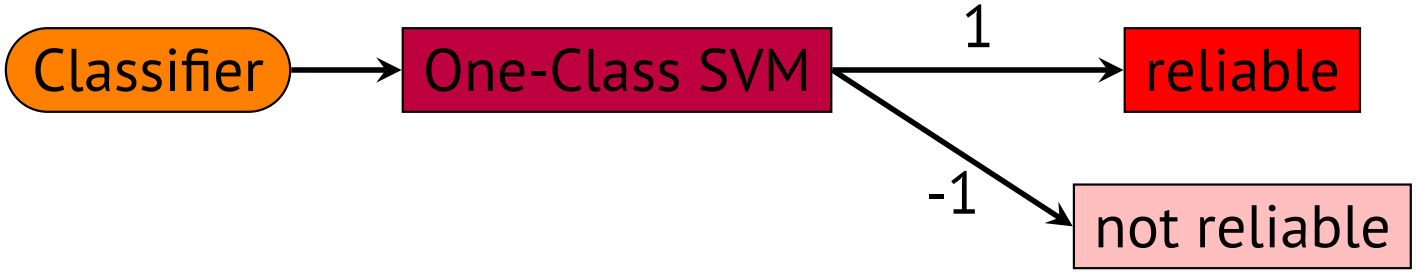
Sub-sequence II

- The 2nd approach followed by 2 methods *Teaser* and *ECEC*:

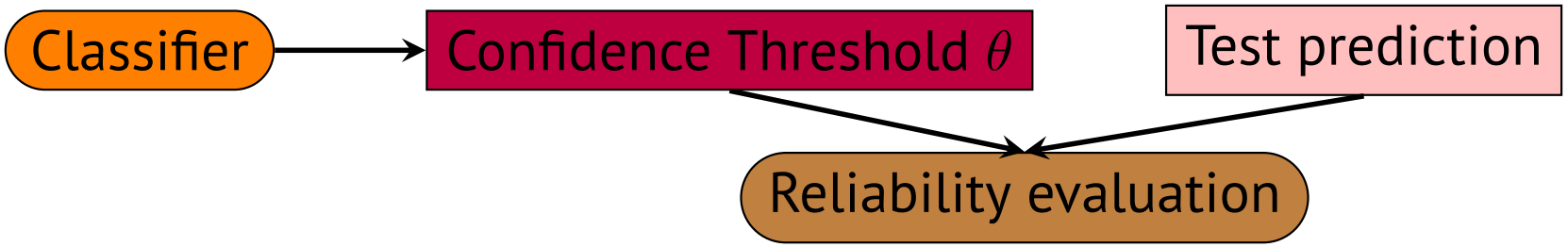


Sub-sequence III

- *Teaser*:

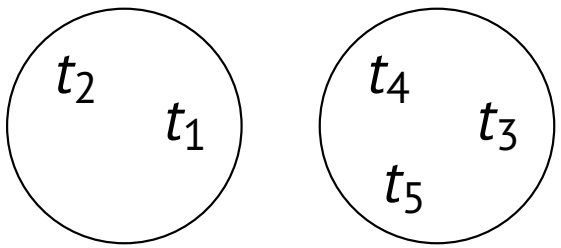


- *ECEC*:



Clustering

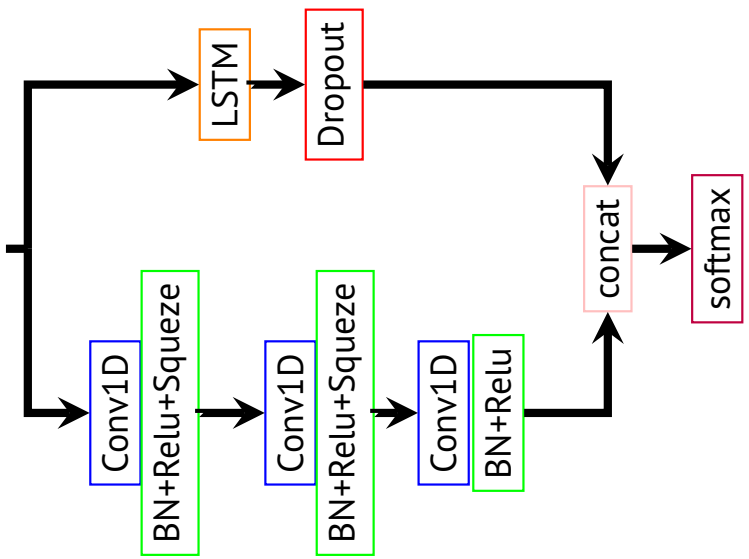
- Approaches in this category attempt to group time-series based on their similarity.
- A classic approach *ECTS* works as follows:
  - Makes clusters with hierarchical clustering based on Euclidean distance and the 1-Nearest Neighbor set of the time-series.
  - Each time-series, based on the cluster has an *MPL* ability, that signifies how many time-points are required for the time-series to act as a classifier.



Time-Series	MPL
t_1	2
t_2	3
t_3	4
t_4	4
t_5	4

Neural Networks

- These approaches use state of the art Neural Network models such as RNNs and CNNs, to conduct classifications with maximal accuracy.
- One of these neural networks algorithms is *MLSTM-FCN* which consists of 2 submodels:
  - The 1st consists of LSTM RNN model, the memory of which can detect intra-time-series dependencies.
  - The 2nd consists of three 1D CNNs, that simplify the time-series and detect their important parts.



Dataset UCR

- Publicly available time-series database from a wide variety of fields such as bio-medicine, food industry, astronomy and archaeology.
- 45 commonly used datasets were utilized for this work.
- Datasets are univariate.

Datasets Biological & Maritime

Two new datasets are also used:

- Biological
  - Cancer cell population after TNF administration.
  - Results from PhysiBoss v2 simulation experiments.
  - Classes *interesting* or *non-interesting*, based on whether TNF treatment was effective or not.
- Maritime
  - Ship coordinates, direction, and trajectory.
  - Classes *True* or *False* based on whether the ships enters a port or not.
- Both datasets are *Multivariate*.

Evaluation Metrics

- Accuracy

$$Accuracy = \frac{number\ of\ correct\ predictions}{number\ of\ test\ instances}$$

- Earliness

$$Earliness = \frac{consumed\ time\ series\ length}{total\ length\ of\ timeseries}$$

- Harmonic Mean

$$HarmonicMean = \frac{2 \cdot (1 - earliness) \cdot accuracy}{(1 - earliness) \cdot accuracy}$$

- F_1 Score

$$F_1\text{-}score = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$$

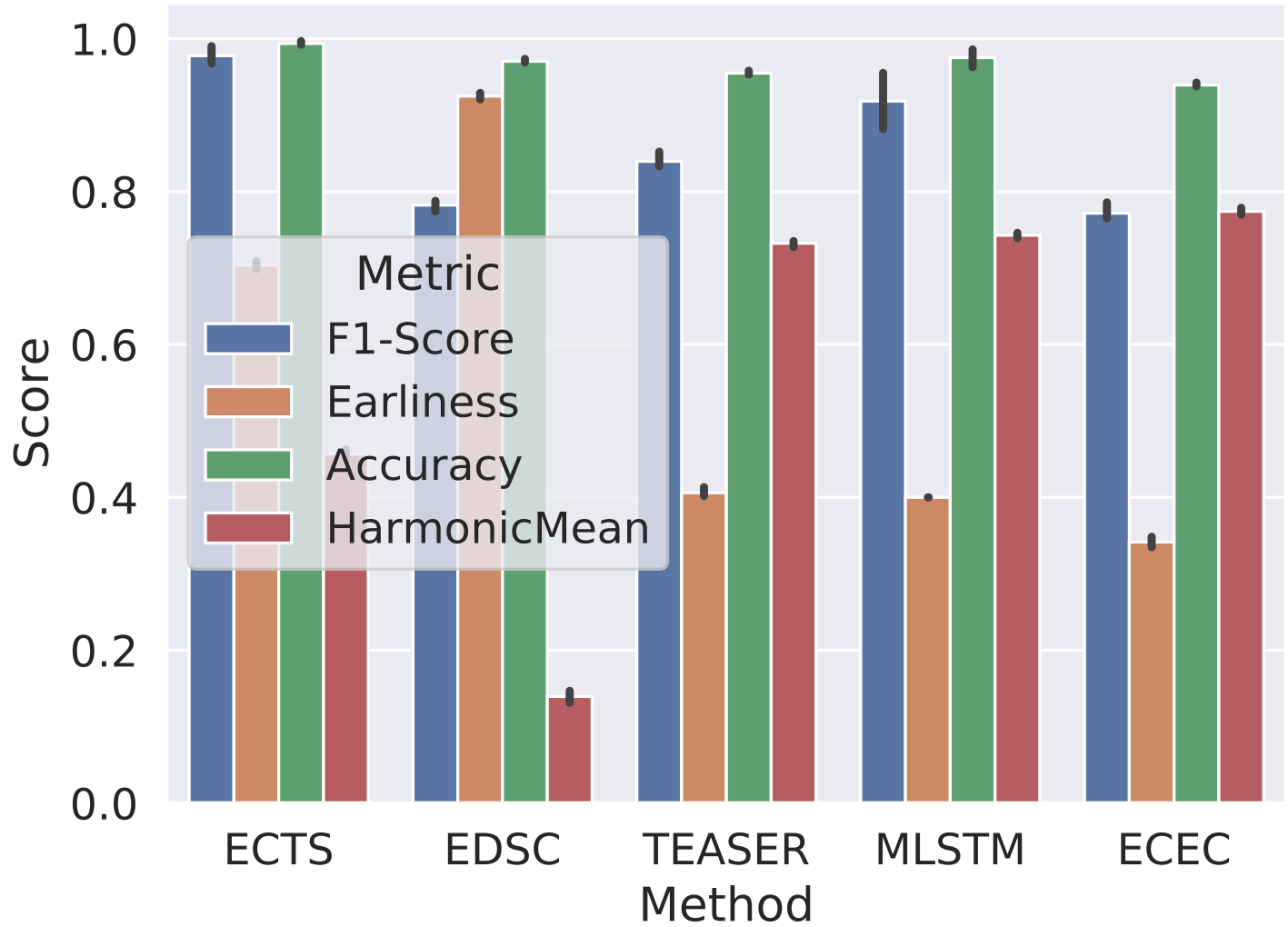
- Computational Complexity (Training and Testing times)

Evaluated setup

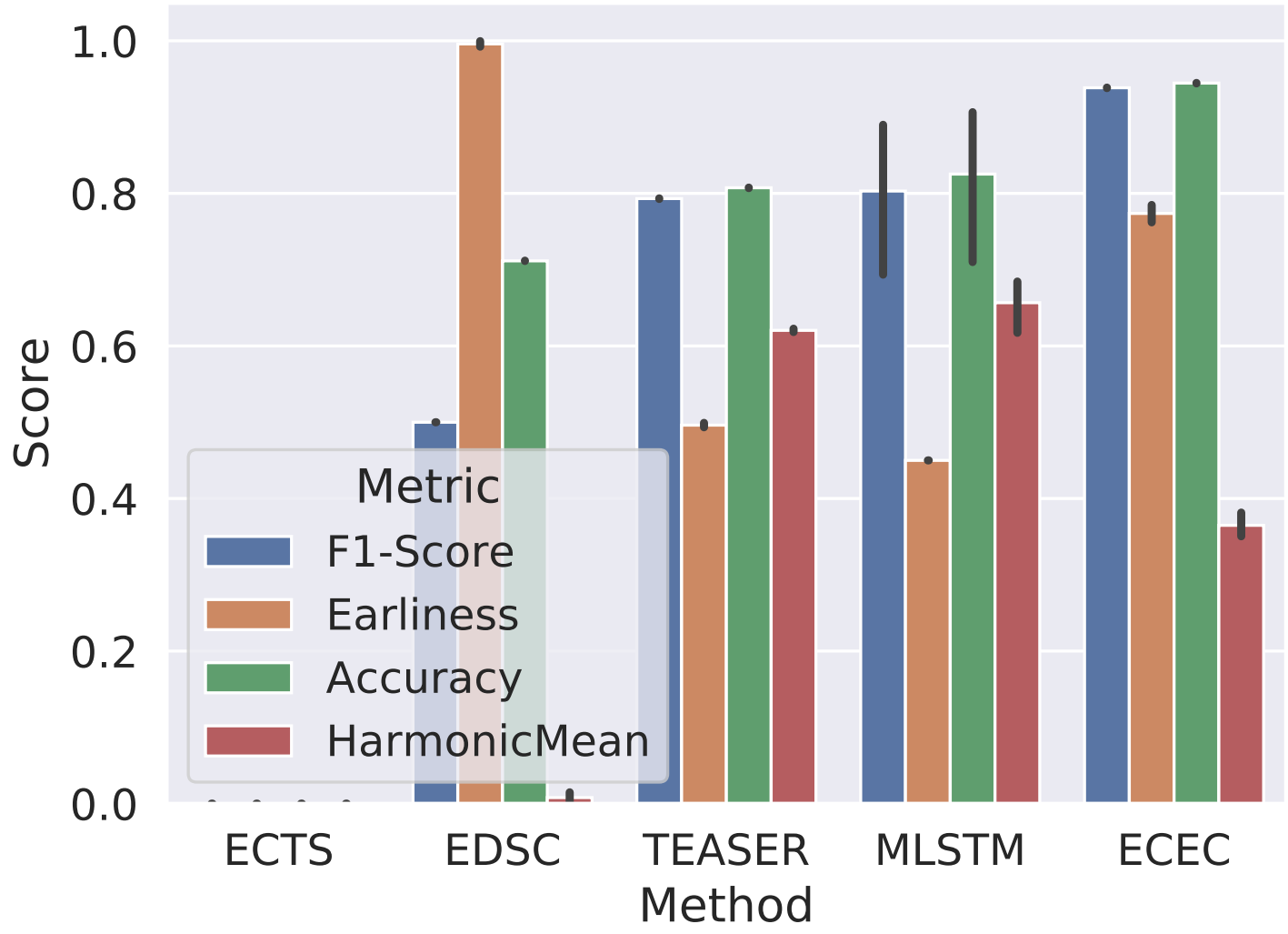
Among these algorithms only *ECTS* was implemented by us.

Method	Language	Configuration
EDSC	C++	CHE, k=3, minLen=5,maxLen = $\frac{L}{2}$
ECTS	Python	support=0
TEASER	Java	S=10 for New datasets, S=20 for UCR
ECEC	Java	N = 20, a=0.8
MLSTM	Python	Attention-LSTM, LSTM cells = 8

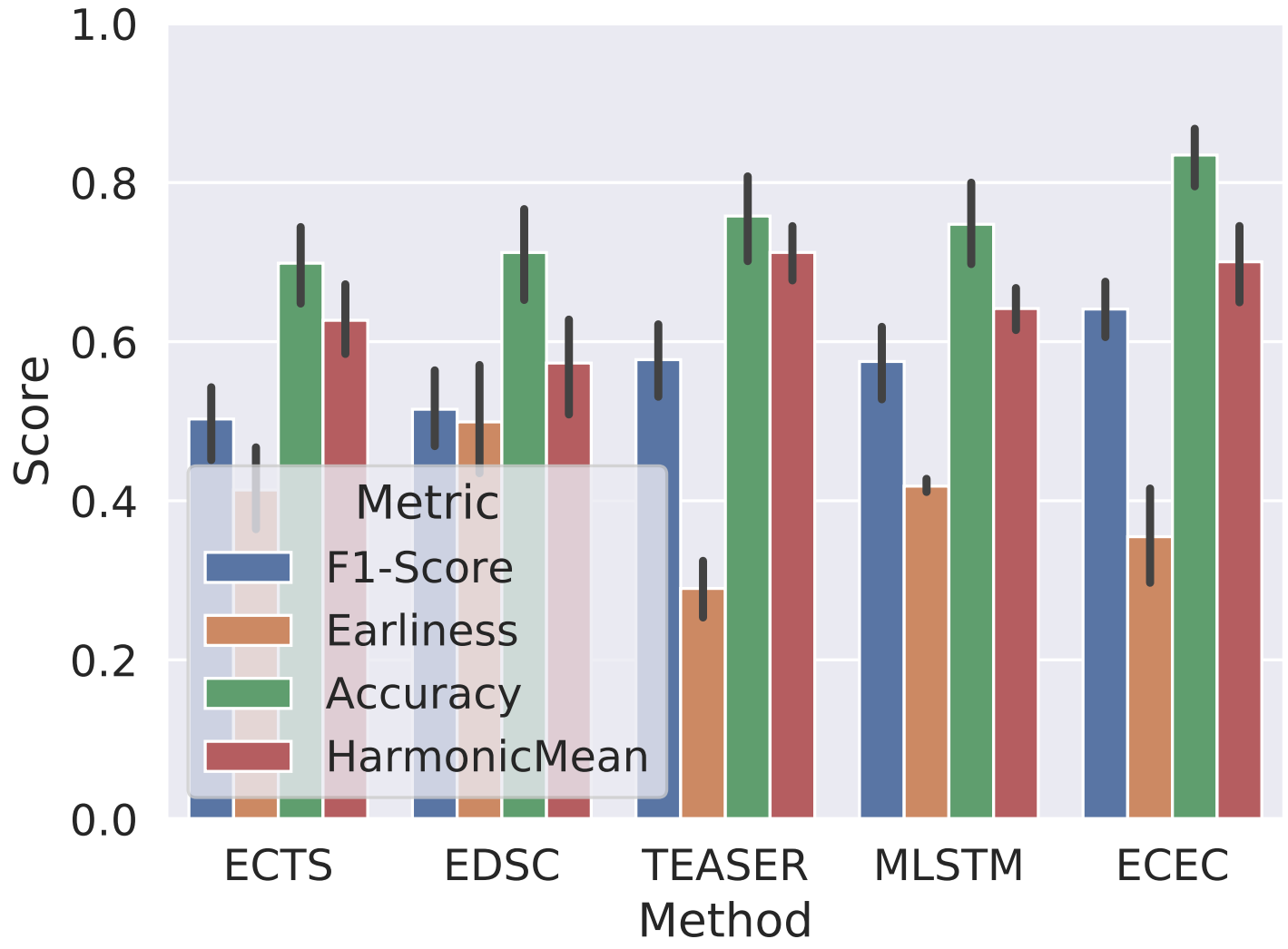
Results for the Biological



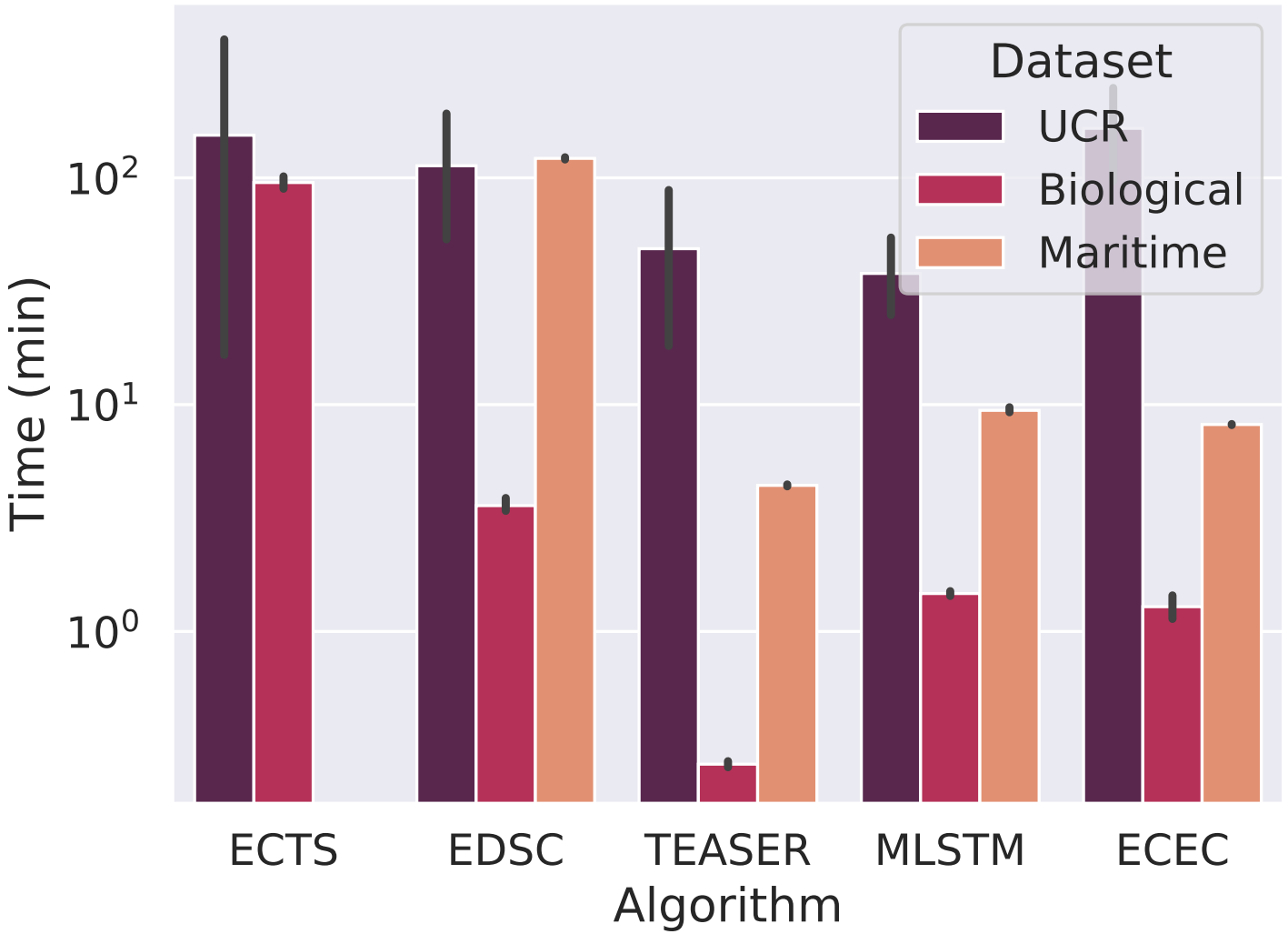
Results for the Maritime



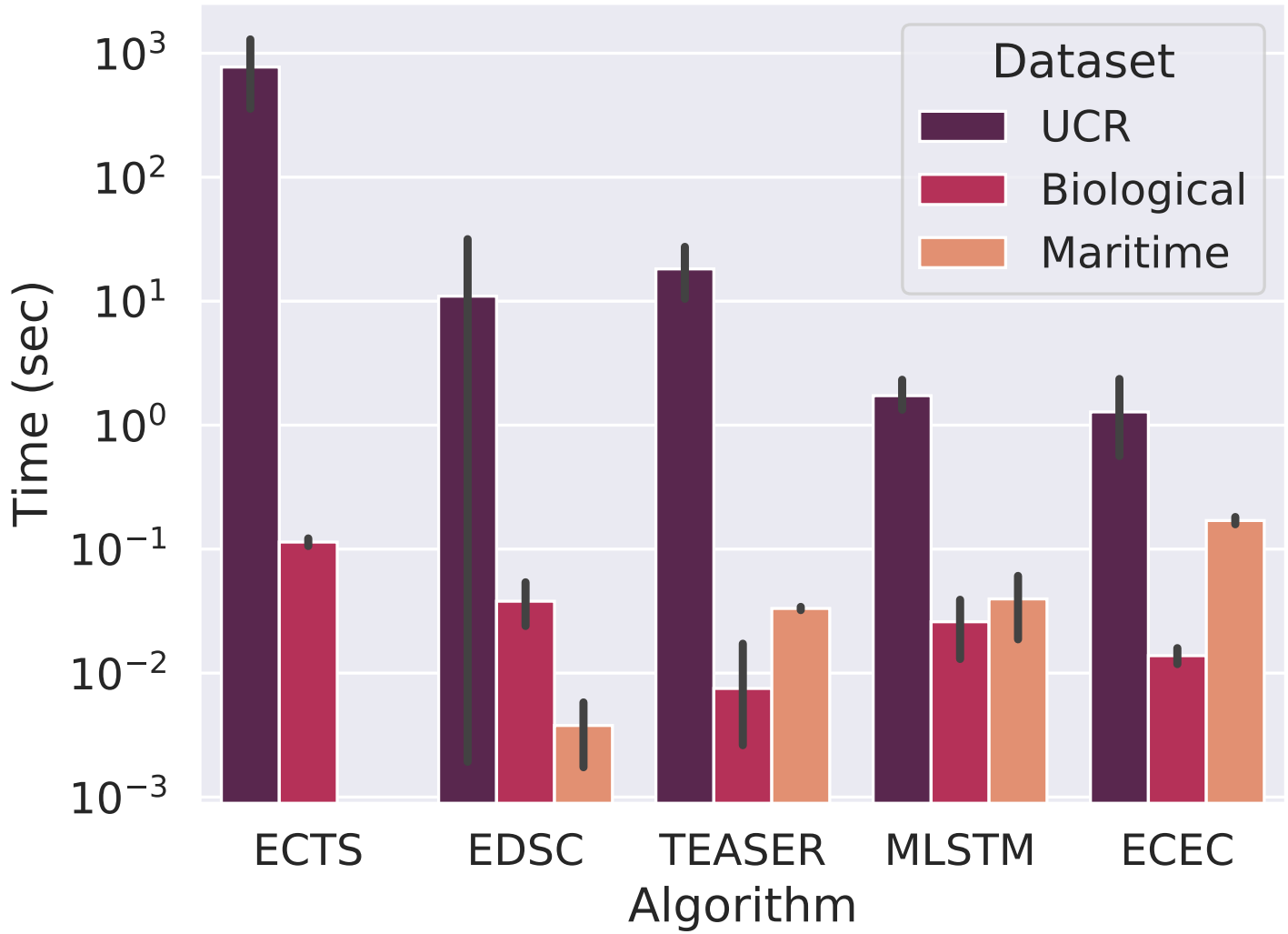
Results UCR


1.0
0.8
0.6
0.4
0.2
0.0
Score
Metric
F1-Score
Earliness
Accuracy
HarmonicMean
ECTS
EDSC
TEASER
MLSTM
ECEC
Method

Results IV


Dataset
UCR
Biological
Maritime
Time (min)
10^2
10^1
10^0
ECTS
EDSC
TEASER
MLSTM
ECEC
Algorithm

Results V


Dataset
UCR
Biological
Maritime
Time (sec)
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
ECTS
EDSC
TEASER
MLSTM
ECEC
Algorithm

Summary & Future work

- Created a publicly available framework for ETS **https://github.com/Eukla/ETS**.
- There is no clear winner among the algorithms since for each dataset they behave differently.
- The 3 most recent approaches Teaser, ECEC, and MLSTM are shown to be more promising with quality results and low data usages.

Aims for the future:

- Extend the framework with a new algorithm Trigger.
- Feel free to extend this open-source repository.

skel
THE AI LAB

Thank You for Your Attention!