A Benchmark for Early Time-Series Classification

Petro - Foti Kamberi, Evgenios Kladis, Charilaos Akasiadis

Institute of Informatics & Telecommunications
NCSR Demokritos

Early Time-Series Classification

- Train on labeled full-length time-series data.

- Predict class labels of unseen time-series data by observing only their **prefixes**.

- Balance trade-off between:
  → Accuracy / F1-Score: Predictive Performance,
  → Earliness: Required prefix length (Lower better).

- Example of application domains (among others):
  Life-sciences, Drug Discovery, Maritime, Energy, etc.

Early Time-Series Classification

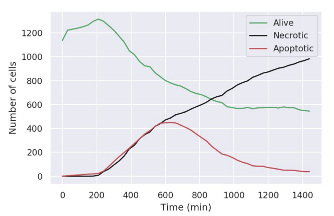
- Train on labeled full-length time-series data.

- Predict class labels of unseen time-series data by observing only their **prefixes**.

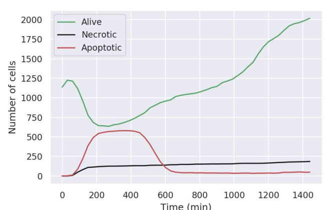
- Balance trade-off between:
  → Accuracy / F1-Score: Predictive Performance,
  → Earliness: Required prefix length (Lower better).

- Example of application domains (among others):
  Life-sciences, Drug Discovery, Maritime, Energy, etc.

Interesting:



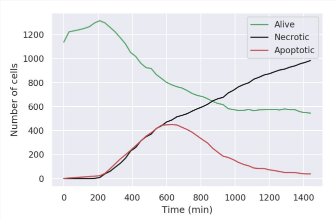
Not interesting:



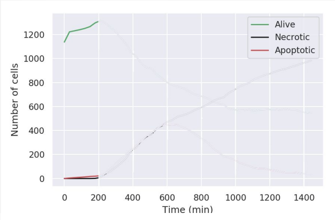
Early Time-Series Classification

- Train on labeled full-length time-series data.
- Predict class labels of unseen time-series data by observing only their **prefixes**.
- Balance trade-off between:
  → <u>Accuracy / F1-Score</u>: Predictive Performance,
  → <u>Earliness</u>: Required prefix length (Lower better).
- Example of application domains (among others):
  Life-sciences, Drug Discovery, Maritime, Energy, etc.

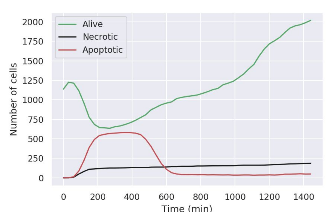
Interesting:



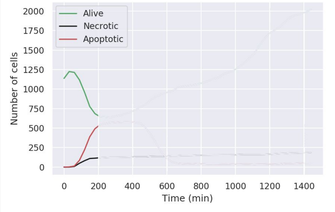
→



Not interesting:



→



Early Time-Series Classification

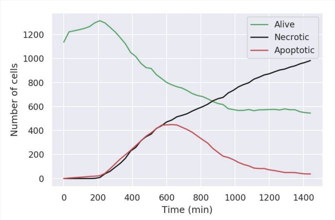
- Train on labeled full-length time-series data.

- Predict class labels of unseen time-series data by observing only their **prefixes**.

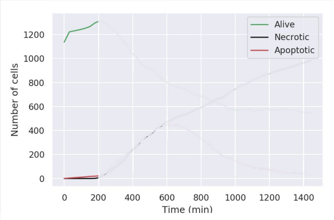
- Balance trade-off between:
  → Accuracy / F1-Score: Predictive Performance,
  → Earliness: Required prefix length (Lower better).

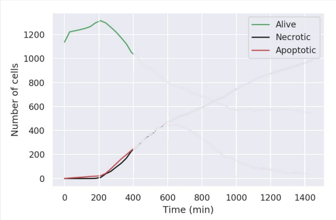
- Example of application domains (among others):
  Life-sciences, Drug Discovery, Maritime, Energy, etc.

Interesting:

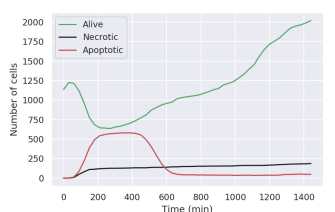


→

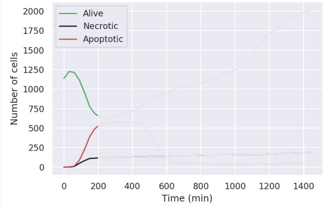


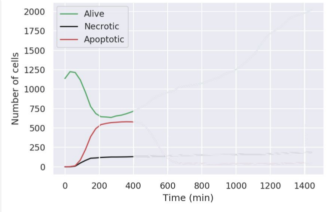


Not interesting:



→





Early Time-Series Classification

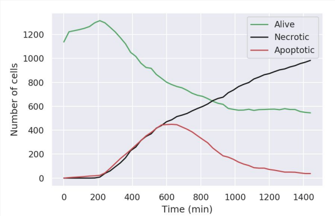
- Train on labeled full-length time-series data.

- Predict class labels of unseen time-series data by observing only their **prefixes**.

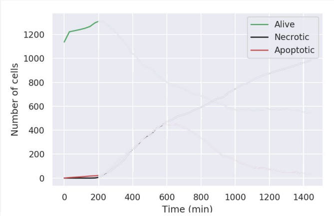
- Balance trade-off between:
  → Accuracy / F1-Score: Predictive Performance,
  → Earliness: Required prefix length (Lower better).

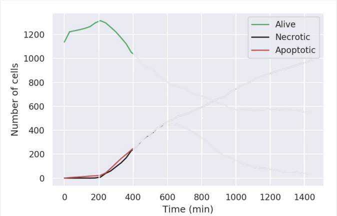
- Example of application domains (among others):
  Life-sciences, Drug Discovery, Maritime, Energy, etc.

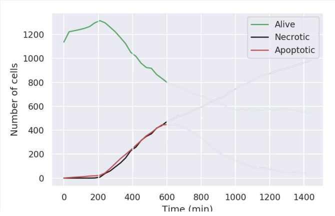
Interesting:



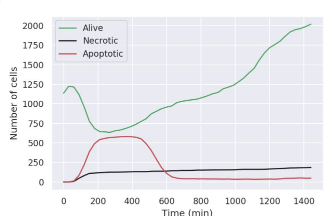
→



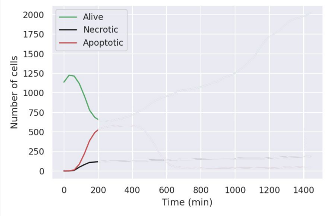


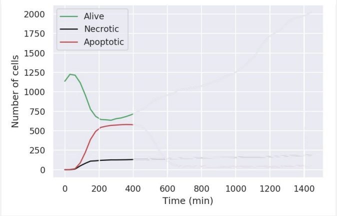


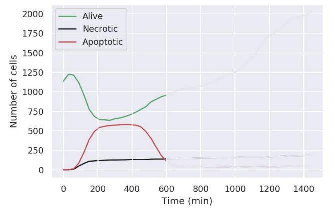
Not interesting:



→







The need for an ETSC Benchmark

- Lack of concise evaluation of ETSC methods in the literature:

  - Evaluation and comparison against a limited set of alternative algorithms.

- Selection of Datasets meaningful for ETSC, which meet the online operation requirements :

  - Time interval between consecutive measurements > Prediction Time,
  - No Z-normalization[1],
  - Temporal dimension.

[1]*R. Wu, A. Der, E. Keogh, When is early classification of time series meaningful, IEEE Transactions on Knowledge and Data Engineering (2021) 1–1, doi:10.1109/TKDE.2021.3108580.*

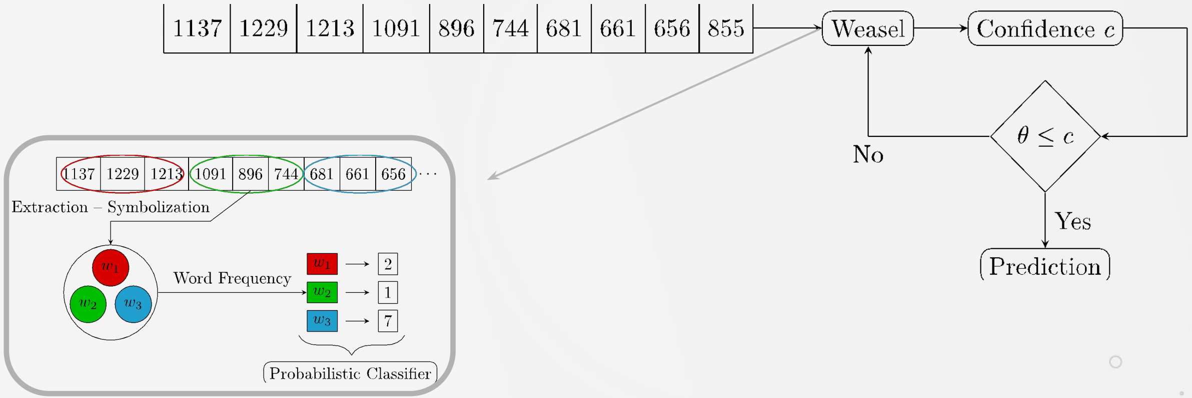
Key Features of this Work

- Open-Source and extensible framework for ETSC evaluation.
- Incorporation of 5 existing algorithms + 1 developed by our team (and some variations).
- Evaluation on 12 meaningful for ETSC datasets (10 from UEA & UCR repository + 2 novel datasets).
- Results and Analysis from an Empirical Comparison.

Incorporated Existing Algorithms - ECEC

Effective Confidence-based Early Classification (ECEC)[2] :

- Train N probabilistic classifiers (on N prefixes), and compute the posterior probability of each being correct.
- Rank classifiers and determine confidence thresholds.
- Accept a prediction if its confidence surpasses the threshold.
- Cubic complexity in terms of time-series length and linear in terms of dataset size.



[2] *J. Lv, X. Hu, L. Li, P.-P. Li, An effective confidence-based early classification of time series, IEEE Access 7 (2019), 96113–96124.*

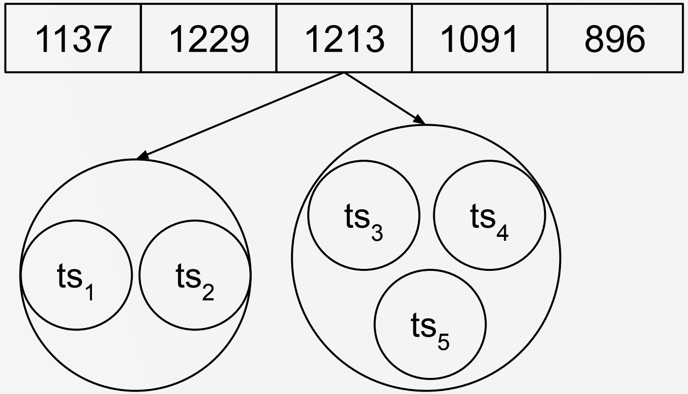
Incorporated Existing Algorithms - ECO-K

Economy-K (ECO-K)[3] :

- Clustering of full time-series.
- A Classifier for each time-point t.
- New observations are assigned to each cluster with a membership probability.
- Generate prediction at the time-point with the minimum f_τ:

$$f_\tau(t_{s_t}) = \sum_{c_k} P(c_k|t_{s_t}) \sum_{y} \sum_{\hat{y}} P_{t+\tau}(\hat{y}|y, c_k)(\hat{y}|y) + C(t+\tau)$$

- Linear complexity in terms of both time-series length and dataset size.



Cluster	Membership Probability
$cluster_1$	0.6
$cluster_2$	0.4

[3] *A. Dachraoui, A. Bondu, A. Cornuejols, Early classification of time series as a non myopic sequential decision making problem, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2015, pp. 433–447.*

Incorporated Existing Algorithms - ECTS

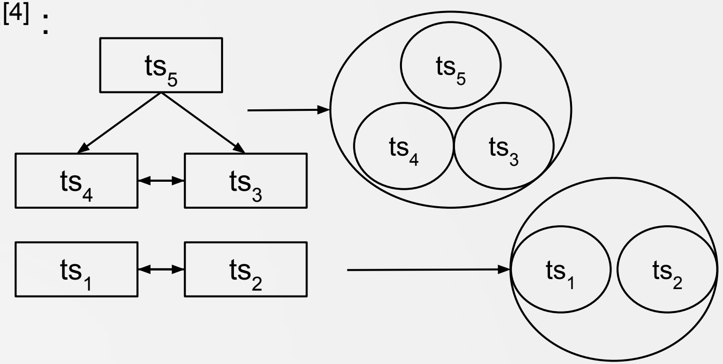
Early Classification of Time Series (ECTS)[4] :

Training:

- Find 1-NN sets for all prefixes and all time-series.
- Determine Reverse NN sets.
- Check from which time-point onward the RNN sets remain unchanged (Minimum prediction length - MPL).
- Group into clusters based on Euclidean distance.
- 1-NN and RNN sets should belong to the same cluster.
- Linear complexity in terms of time-series length and cubic in terms of dataset size.

Testing:

- Match new instances with their NN.
- If testing time-series length is larger than NN's MPL, then generate a prediction.



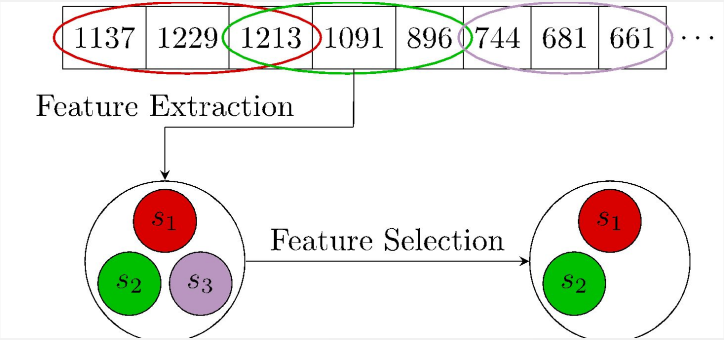
Time series	MPL (NN)	MPL (Clustering)
ts_1	2	2
ts_2	7	3
ts_3	6	4
ts_4	4	4
ts_5	4	4

[4] *Z. Xing, J. Pei, P. S. Yu, Early classification on time series, Knowledge and information systems 31 (1) (2012) 105–127.*

Incorporated Existing Algorithms - EDSC

Early Distinctive Shapelet Classification (EDSC)[5] :

- One of the first ETSC methods.
- User defines a range of subseries lengths.
- Create shapelets that are maximally representative of a class (*subseries, threshold, class label*).
- Threshold: minimum distance that another time-series should have to assign to the same class label -> find minimum distances of time-series with different class labels.
- Determine most discriminative shapelets for each class by checking their utility (F1-score with weighted Recall).
- Cubic complexity in terms of time-series length and quadratic complexity in terms of dataset size.

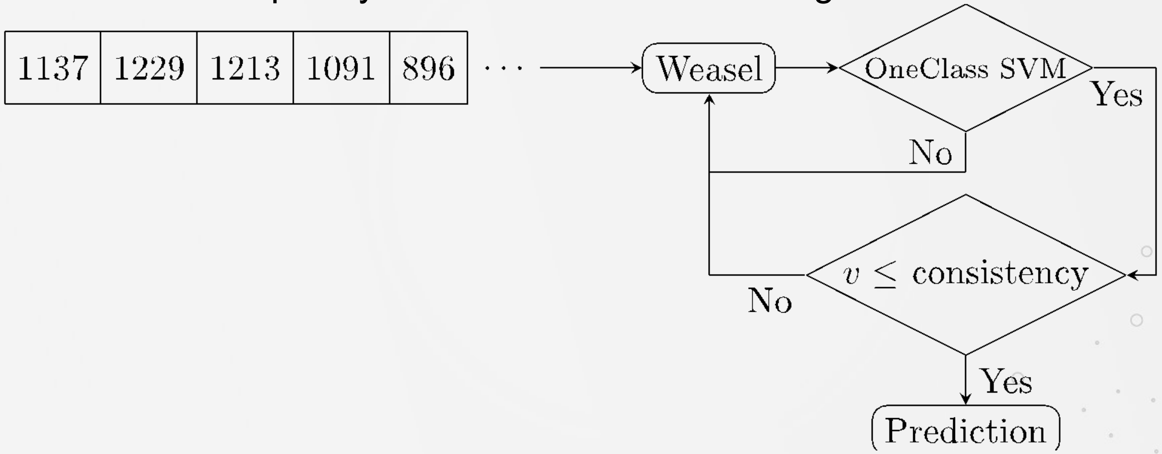


[5] *Z. Xing, J. Pei, P. S. Yu, K. Wang, Extracting interpretable features for early classification on time series, in: Proceedings of the 2011 SIAM international conference on data mining, SIAM, 2011, pp. 247–258.*

Incorporated Existing Algorithms - TEASER

Two-tier Early and Accurate Series Classifier (TEASER)[6] :

- Truncate the dataset to *N* overlapping prefixes.
- For each prefix train a WEASEL-Logistic Regression pair.
- For each prefix length train an One Class SVM that accepts correct predictions and rejects false ones.
- Check for consecutive and consistent predictions before terminating.
- Quadratic complexity in terms of time-series length.



[6] *P. Schäfer, U. Leser, TEASER: early and accurate time series classification, Data mining and knowledge discovery 34 (5) (2020) 1336–1362*

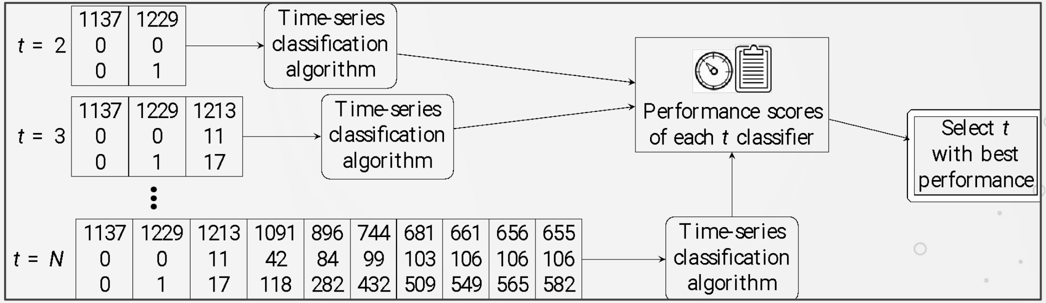
A new, simpler benchmark method: - STRUT

Selective Truncation of Time-series (STRUT):

- Split the training dataset to all possible prefixes.
- Train and validate full time-series classification algorithms (**Minirocket**[7], **MLSTM**[8], **WEASEL**[9]).
- Select minimum *t* where a user-defined metric *S(.)* is optimized

$$STRUT(X_{test}) = \min(\arg\max_t S(h^t(X_{test}^{1:t}|D_{train})))$$

- Faster Approximation Variant: Decide which time-points *t* to evaluate using bisection (essentially an iterative binary search approach).
- Linear complexity in terms of dataset size. For S-WEASEL quadratic complexity in terms of time-series length and linear fof S-MINI. The complexity of S-MLSTM is affected by the inherent complexity of LSTMs.



[7]A. Dempster, D. F. Schmidt, G. I. Webb, Minirocket: A very fast (almost) deterministic transform for time series classification, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 248–257.
[8]F. Karim, S. Majumdar, H. Darabi, S. Harford, Multivariate LSTM- FCNs for time series classification, Neural Networks 116 (2019) 237–245.
[9]P. Schäfer, U. Leser, Fast and accurate time series classification with WEASEL, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 637–646.

Datasets

- Ten meaningful for ETSC datasets from the UCR/UEA repository.
- Two new datasets:
    1. Life-sciences domain: large-scale simulations of drug treatments for cancer,
    2. Maritime domain, vessel AIS messages.
- Categorization according to characteristics.

Dataset Name	No. of Variables	No. of Instances (Height)	No. of Time-points (Length)	No. of Classes	Class Imbalance Ratio	Standart Deviation
BasicMotions	6	80	100	4	1	3.02
Biological	3	644	49	2	5.37	222.93
DodgerLoopDay	1	158	288	7	1.25	12.77
DodgerLoopGame	1	158	288	2	1.07	12.77
DodgerLoopWeekend	1	158	288	2	2.43	12.77
HouseTwenty	1	159	2,000	2	1.27	779.19
LSST	6	4,925	36	14	111	167.26
Maritime	7	80,591	30	2	4.21	37.08
PickupGestureWiimoteZ	1	100	361	10	1	0.2
PLAID	1	1,074	1,345	11	6.73	3.41
PowerCons	1	360	144	2	1	0.9
SharePriceIncrease	1	1,931	60	2	2.19	1.64

Group	Specifications	Datasets
Wide	Length $>$ 1,300	HouseTwenty, PLAID
Large	Height $>$ 1,000	LSST, Maritime, PLAID, SharePriceIncrease
Unstable	Coef. of Var. $>$ 1.08	BasicMotions, Biological, HouseTwenty, LSST, PLAID, SharePriceIncrease
Imbalanced	Class Imbalance Ratio $>$ 1.73	Biological, DodgerLoopWeekend, LSST, Maritime, PLAID, SharePriceIncrease
Multiclass	Number of Classes $>$ 2	BasicMotions, DodgerLoopDay, LSST, PickupGestureWiimoteZ, PLAID
Common	None of the above	BasicMotions, DodgerLoopGame, DodgerLoopWeekend, PickupGestureWiimoteZ, PowerCons
Univariate	One variable per instance	DodgerLoopDay, DodgerLoopGame, DodgerLoopWeekend, HouseTwenty, PickupGestureWiimoteZ, PLAID, PowerCons, SharePriceIncrease
Multivariate	More than one variable per instance	BasicMotions, Biological, LSST, Maritime

Top: Tested Datasets' Characteristics, **Bottom**: Dataset Groups and Categorization Criteria

The Framework

- Main language: Python 3.7
  - ECEC and TEASER implemented in Java
  - EDSC in C++

- Performance Metrics:
  - Accuracy
  - F1-Score
  - Earliness
  - Harmonic Mean of Accuracy and Earliness
  - Training and Testing times

- CLI with various options that allows to specify, e.g. :
  - Input files format (.arff and .csv)
  - Number of variables in the multivariate cases
  - Method for applying univariate algorithms to multivariate datasets (e.g. voting)
  - Target class for F1-Score calculation
  - Option for cross-validation

Public GitHub Repository

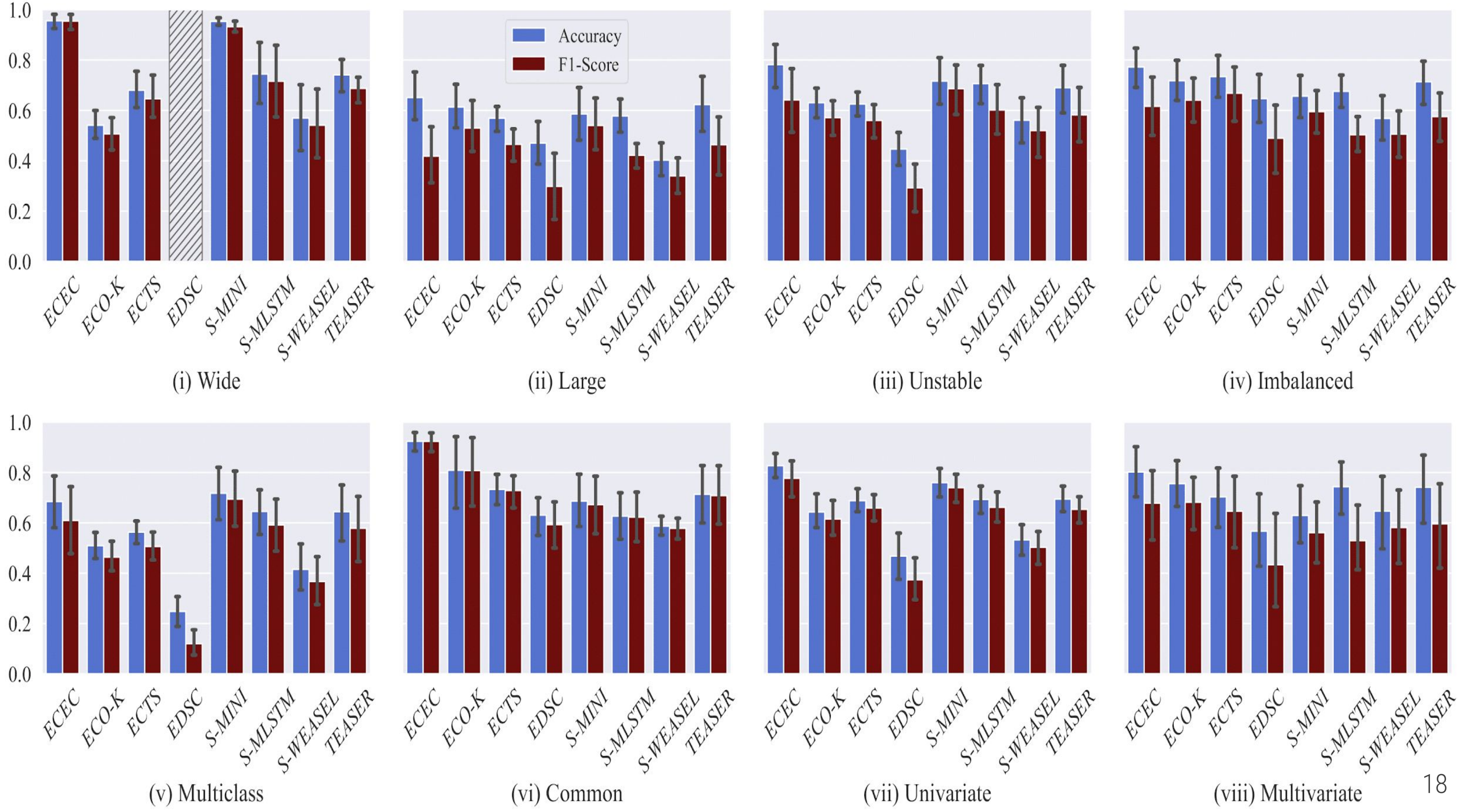
https://github.com/xarakas/ETSC

Framework Extensibility

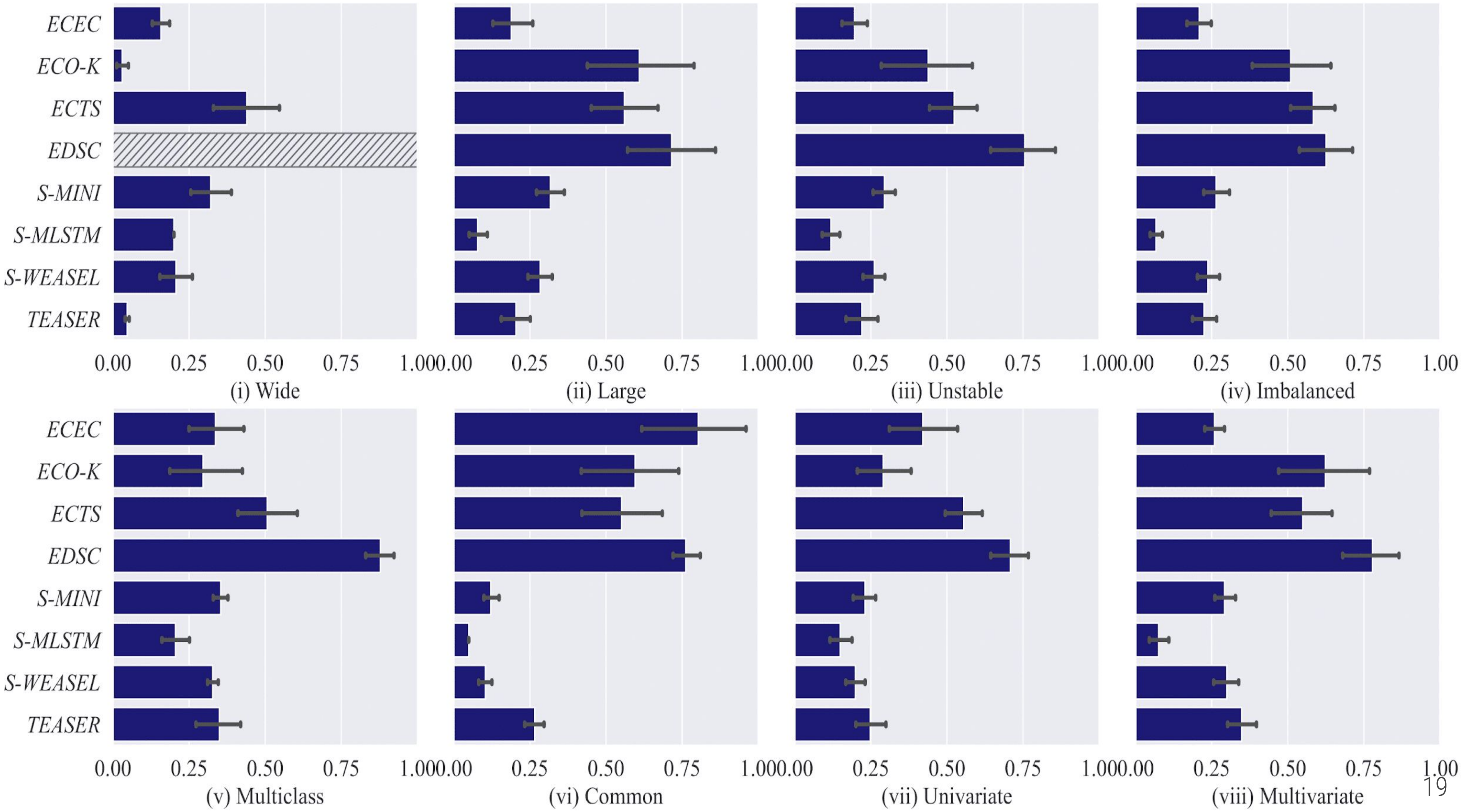
- Adding New Algorithms:

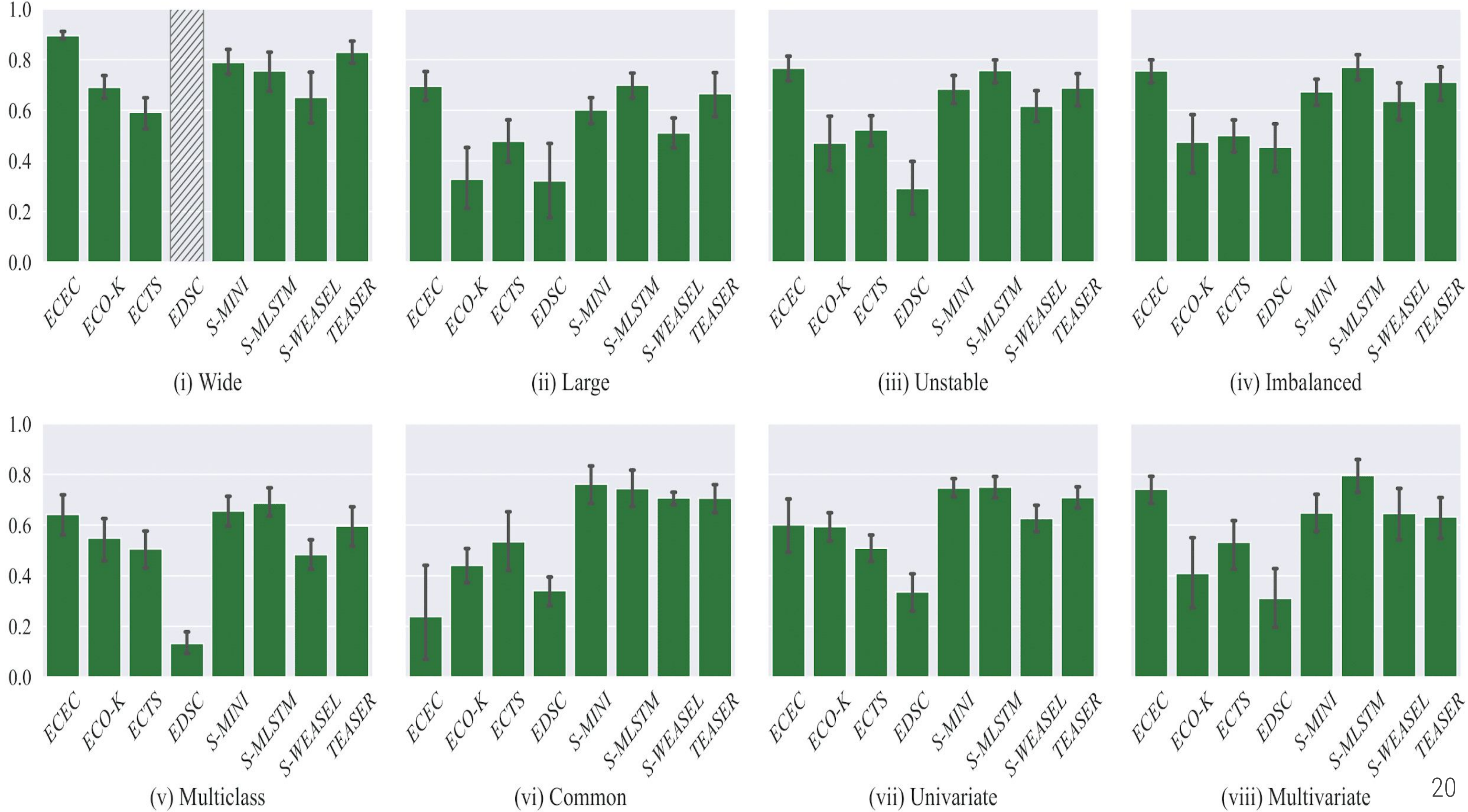
  - Create a Python interface implementing the abstract class EarlyClassifier.
  - Create a python script that Implements the train and predict methods.
  - Extend the file CLI to:
    (a) Import the new algorithm implementation,
    (b) Define input parameters, execution options, and algorithm invocation.
  - Note: Algorithms can be implemented in any language, provided a Python wrapper is available.

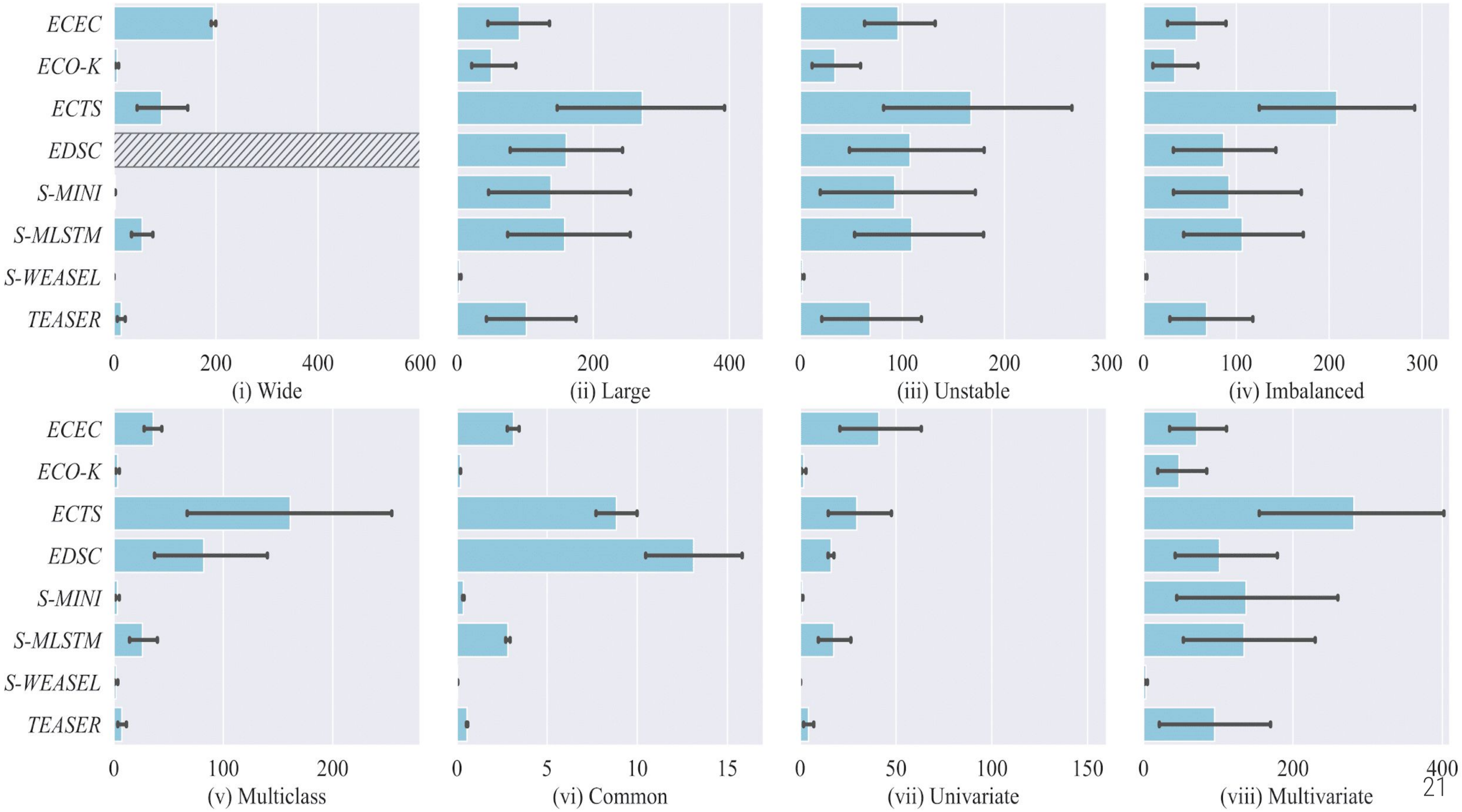
- Enriching Datasets
  - Datasets should be in .csv file format.
  - Each row should represent a time-series instance and the first value should represent the assigned class label.



Accuracy
F1-Score
1.0
0.8
0.6
0.4
0.2
0.0
ECEC
ECO-K
ECTS
EDSC
S-MINI
S-MLSTM
S-WEASEL
TEASER
(i) Wide
(ii) Large
(iii) Unstable
(iv) Imbalanced
(v) Multiclass
(vi) Common
(vii) Univariate
(viii) Multivariate


ECEC
ECO-K
ECTS
EDSC
S-MINI
S-MLSTM
S-WEASEL
TEASER
0.00
0.25
0.50
0.75
1.00
(i) Wide
(ii) Large
(iii) Unstable
(iv) Imbalanced
(v) Multiclass
(vi) Common
(vii) Univariate
(viii) Multivariate


1.0
0.8
0.6
0.4
0.2
0.0
ECEC
ECO-K
ECTS
EDSC
S-MINI
S-MLSTM
S-WEASEL
TEASER
(i) Wide
(ii) Large
(iii) Unstable
(iv) Imbalanced
(v) Multiclass
(vi) Common
(vii) Univariate
(viii) Multivariate


ECEC
ECO-K
ECTS
EDSC
S-MINI
S-MLSTM
S-WEASEL
TEASER
0
200
400
600
(i) Wide
0
200
400
(ii) Large
0
100
200
300
(iii) Unstable
0
100
200
300
(iv) Imbalanced
0
100
200
(v) Multiclass
0
5
10
15
(vi) Common
0
50
100
150
(vii) Univariate
0
200
400
(viii) Multivariate

Key Takeaways

- Introduction of an **open-source framework** for evaluating and objectively comparing ETSC algorithms.
- The **framework** is highly **extensible**, allowing for the seamless incorporation of new **algorithms** and the enrichment with new **datasets** (that should be meaningful for ETSC).
- **Benchmark Method - STRUT**: We present a benchmark method called STRUT, which serves as an essential baseline for the monitoring of the ETSC performance without a dynamic component to declare earliness.

Ongoing and Future Work

We are currently extending the framework's algorithms with:

- **T-SMOTE**[10]: An algorithm that focuses on addressing class imbalance in time-series data to improve classification.
- **SDRE[11]**: Incorporating Spectral Density Ratio Estimation for more robust ETSC solutions.
- **MOO-ETSC[12]**: Multi-Objective Optimization in ETSC, exploring diverse performance criteria.

In the future, we plan to incorporate Automated Machine Learning (**AutoML**)[13] tools for optimizing hyperparameters in ETSC algorithms.

[10]*P. Zhao, C. Luo, B. Qiao, L. Wang, S. Rajmohan, Q. Lin, and D. Zhang. 2022. T-SMOTE: temporal-oriented synthetic minority oversampling technique for imbalanced time series classification. In Proceedings of IJCAI.*

[11]*A. F. Ebihara, T. Miyagawa, K. Sakurai, and H. Imaoka. 2023. Toward Asymptotic Optimality: Sequential Unsupervised Regression of Density Ratio for Early Classification. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 1–5.*

[12]*U. Mori, A. Mendiburu, I.M. Miranda, and J.A. Lozano. 2019. Early classification of time series using multi-objective optimization techniques. Information Sciences 492 (2019), 204–218. https://doi.org/10.1016/j.ins.2019.04.024.*

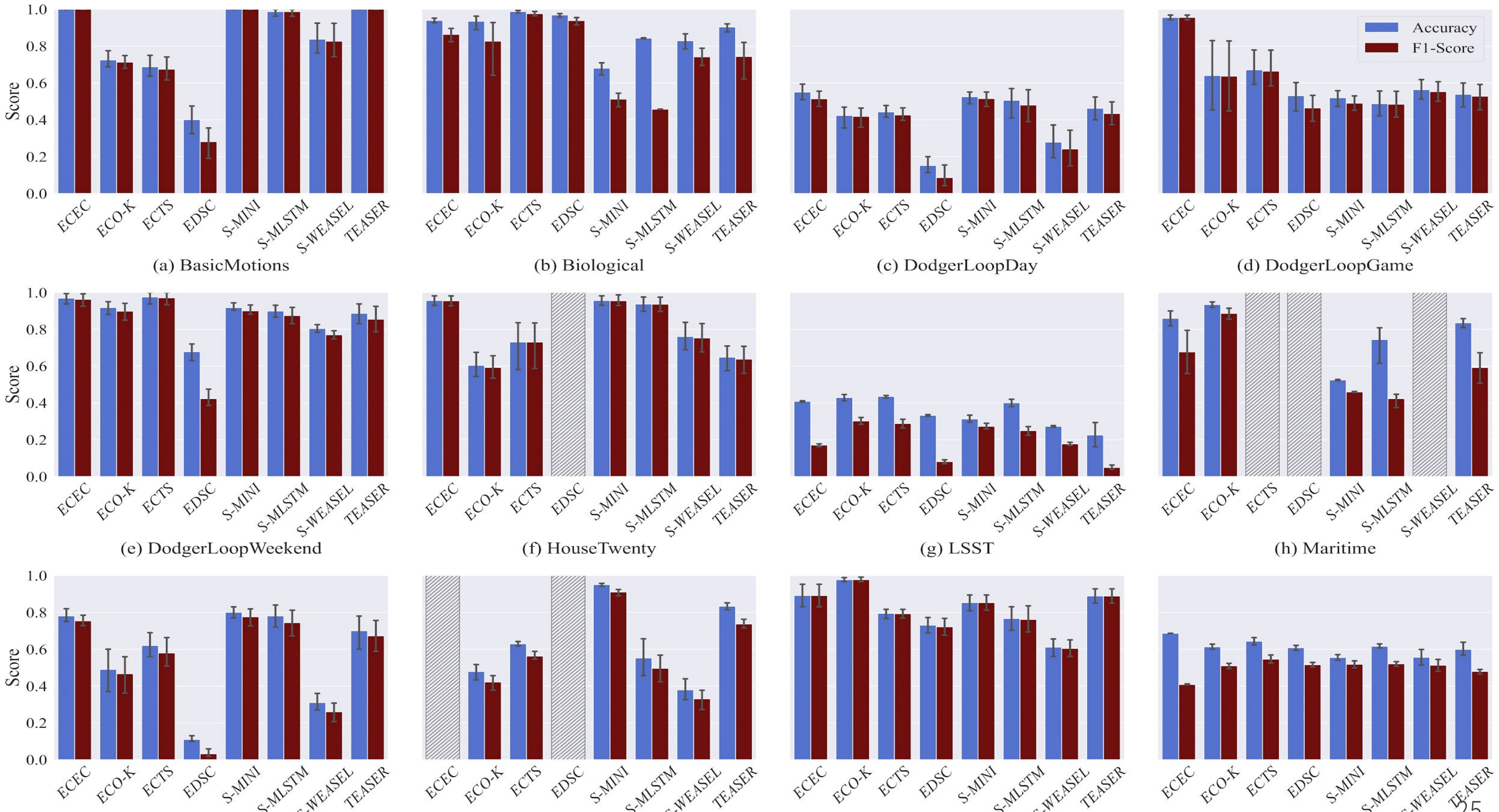
[13]*G. Ottervanger, M. Baratchi, and H. H. Hoos. 2021. MultiETSC: automated machine learning for early time series classification. Data Mining and Knowledge Discovery (2021), 1–53.*

Thank you!

Contact: cakasiadis@iit.demokritos.gr
pkamperi@iit.demokritos.gr

More info: https://cer.iit.demokritos.gr/

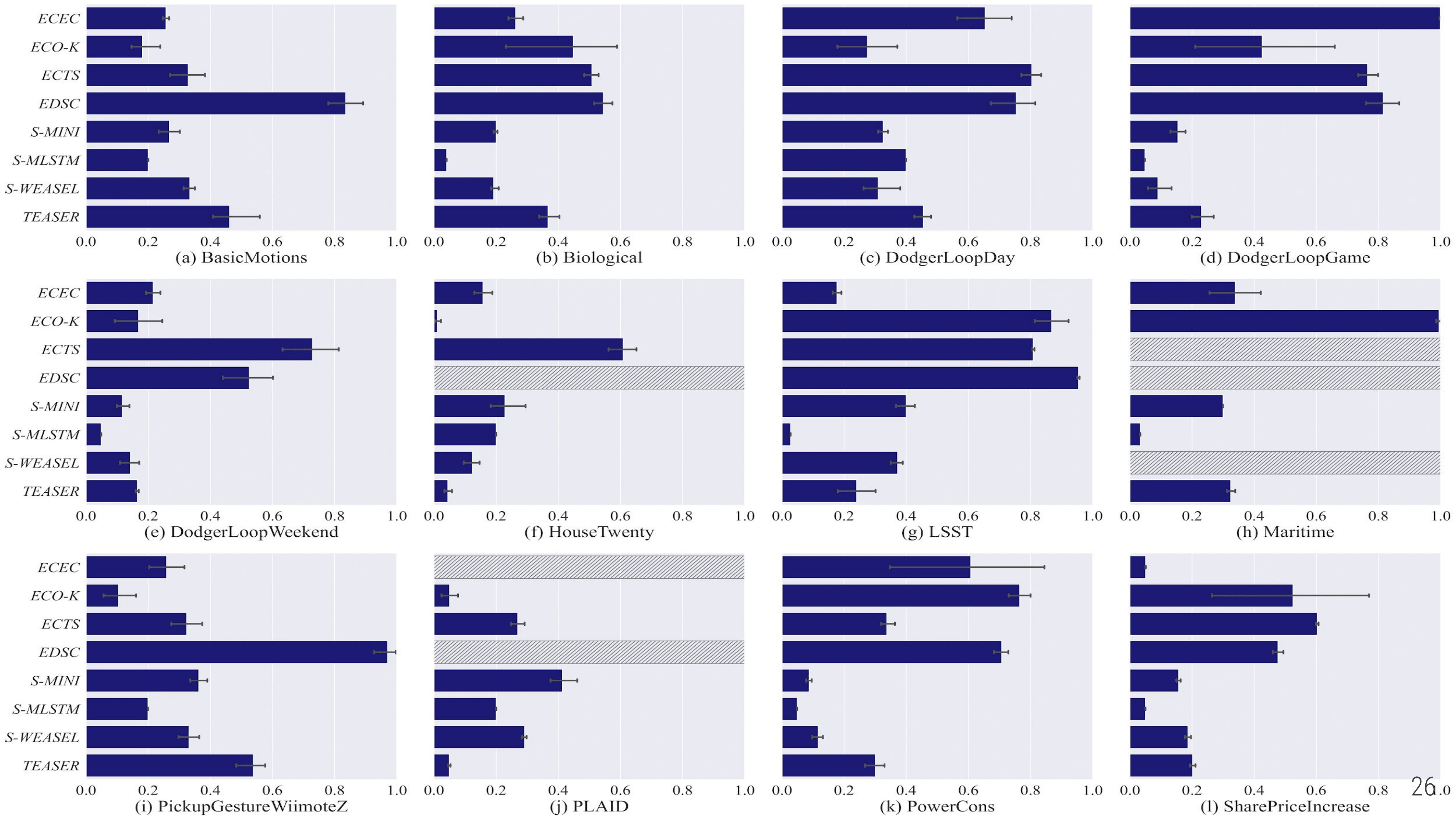
Comments/Questions?

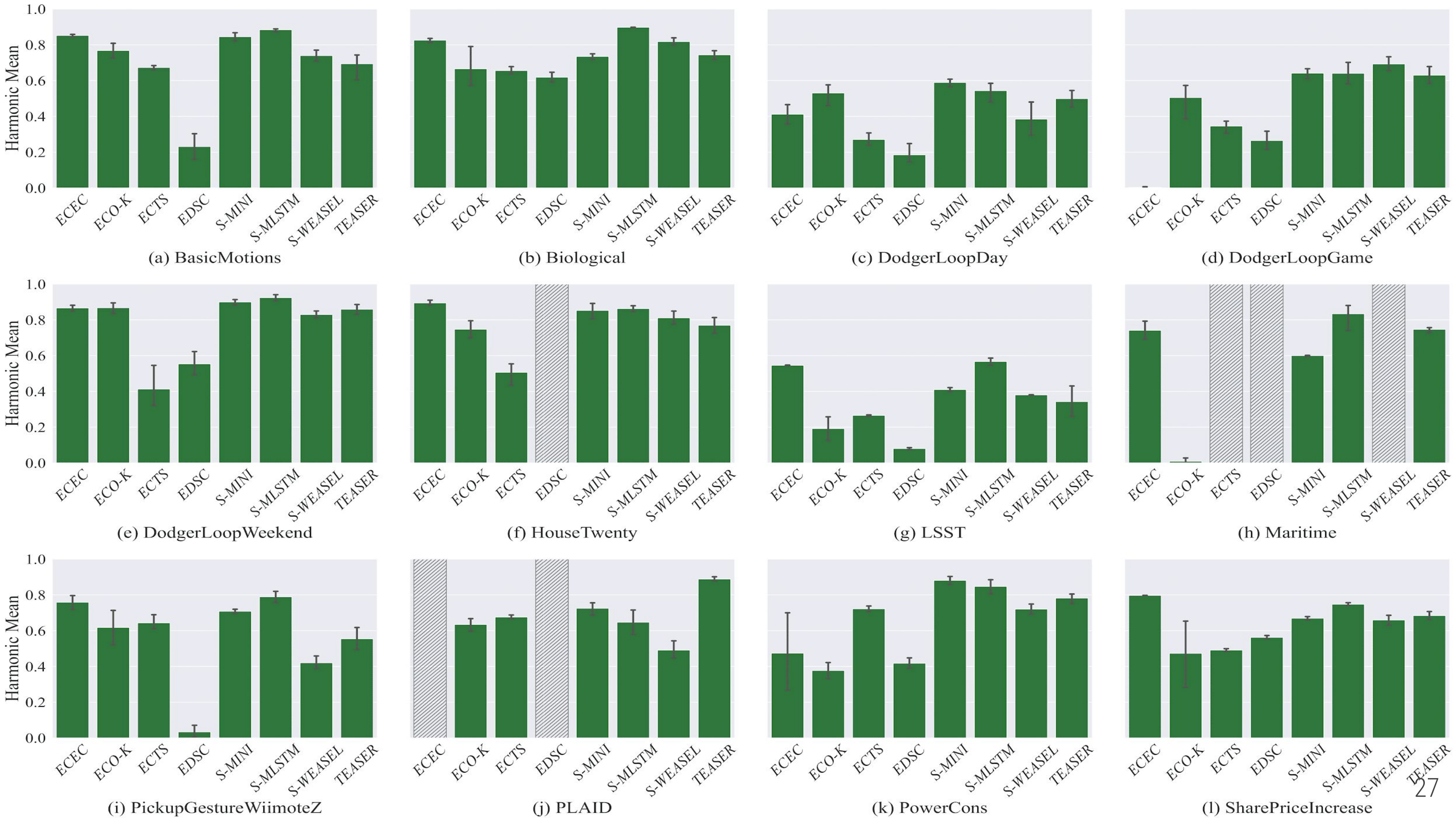


(a) BasicMotions (b) Biological (c) DodgerLoopDay (d) DodgerLoopGame

(e) DodgerLoopWeekend (f) HouseTwenty (g) LSST (h) Maritime

(i) PickupGestureWiimoteZ (j) PLAID (k) PowerCons (l) SharePriceIncrease

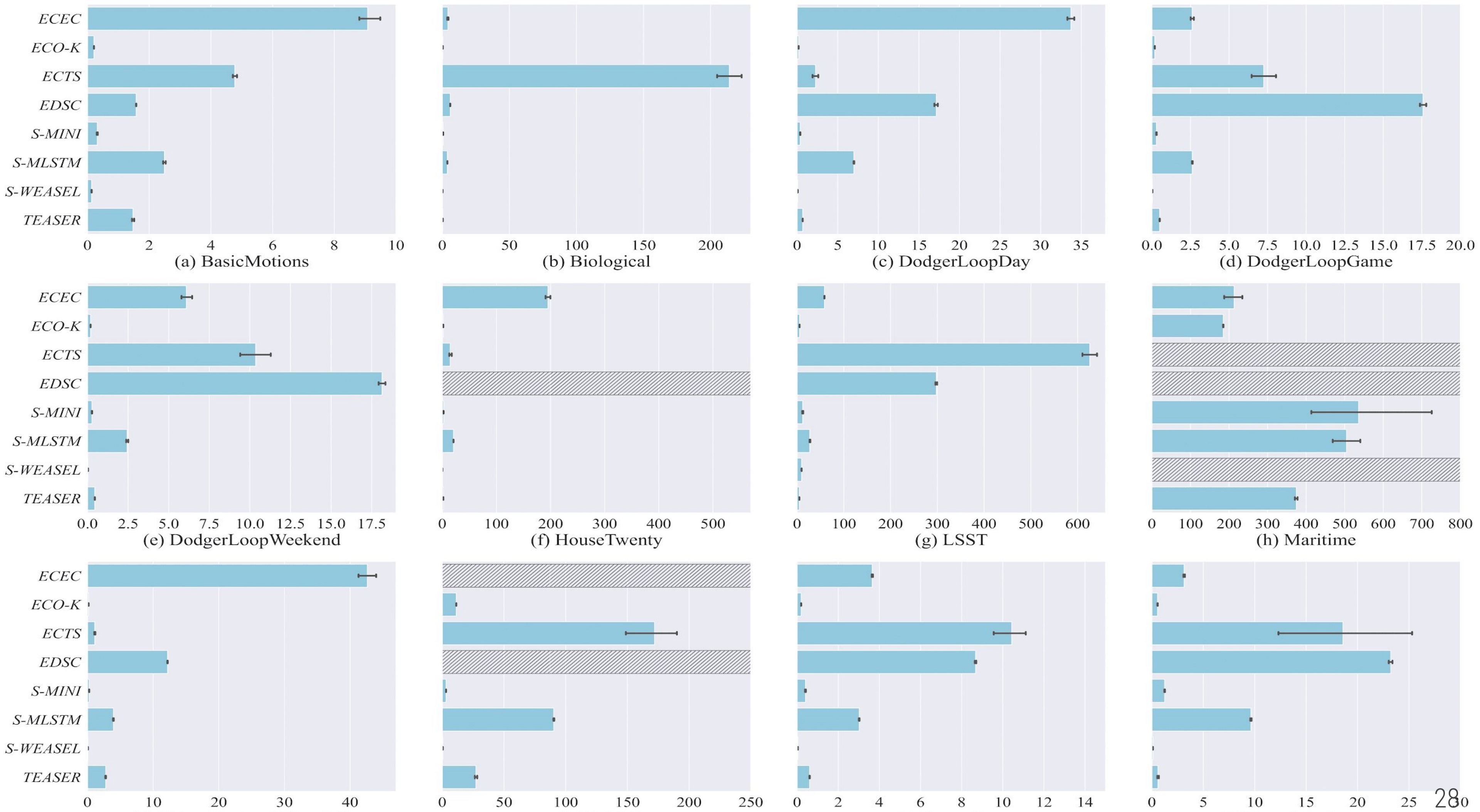

ECEC
ECO-K
ECTS
EDSC
S-MINI
S-MLSTM
S-WEASEL
TEASER
0.0
0.2
0.4
0.6
0.8
1.0
(a) BasicMotions
(b) Biological
(c) DodgerLoopDay
(d) DodgerLoopGame
(e) DodgerLoopWeekend
(f) HouseTwenty
(g) LSST
(h) Maritime
(i) PickupGestureWiimoteZ
(j) PLAID
(k) PowerCons
(l) SharePriceIncrease



(a) BasicMotions (b) Biological (c) DodgerLoopDay (d) DodgerLoopGame

(e) DodgerLoopWeekend (f) HouseTwenty (g) LSST (h) Maritime

(i) PickupGestureWiimoteZ (j) PLAID (k) PowerCons (l) SharePriceIncrease



(a) BasicMotions (b) Biological (c) DodgerLoopDay (d) DodgerLoopGame

(e) DodgerLoopWeekend (f) HouseTwenty (g) LSST (h) Maritime

(i) PickupGestureWiimoteZ (j) PLAID (k) PowerCons (l) SharePriceIncrease