

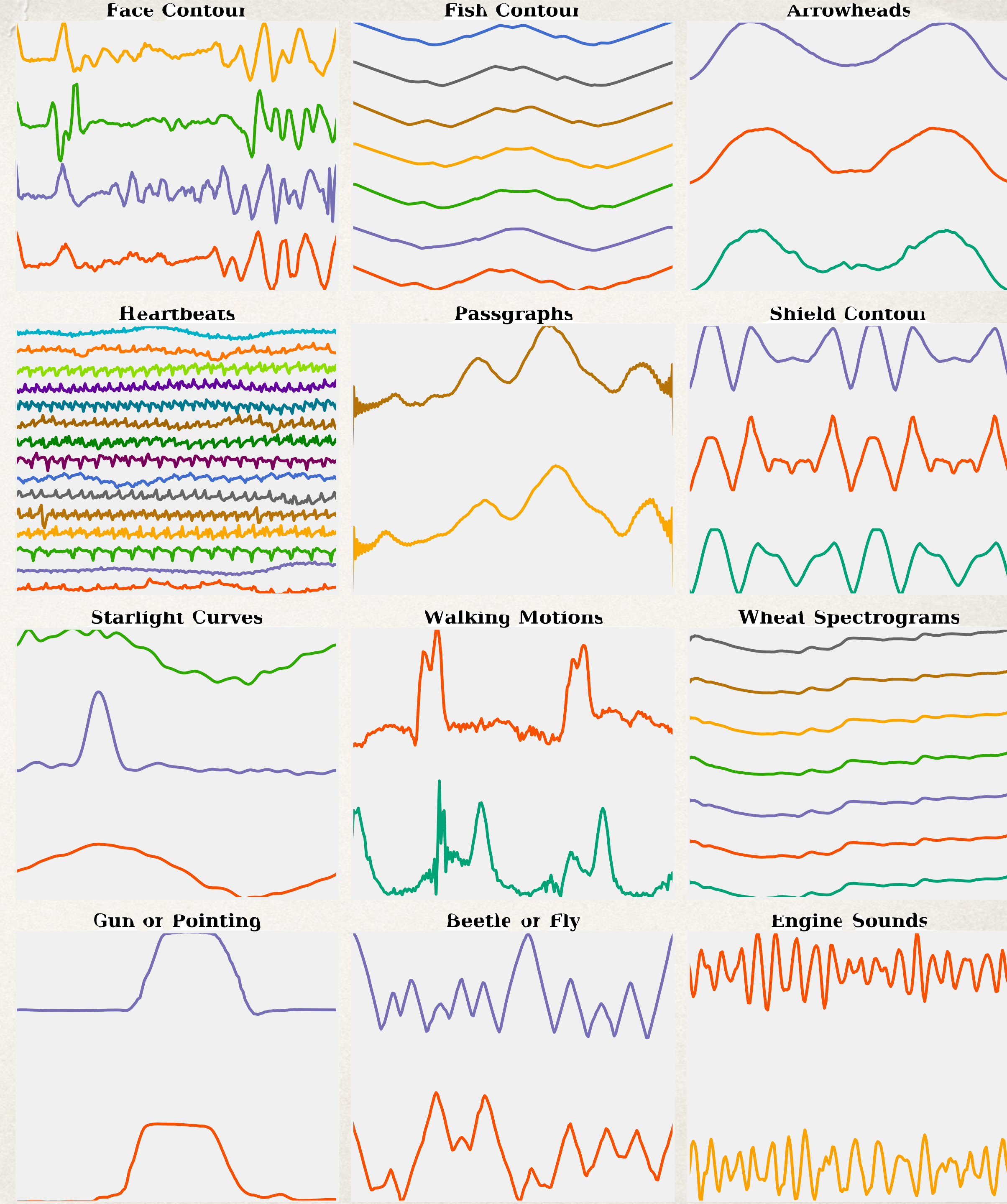
Benchmarking (State-of-the-Art) Univariate Time Series Classifiers

Patrick Schäfer and Ulf Leser

Humboldt-Universität zu Berlin, Wissensmanagement in der Bioinformatik



- ❖ Time series (TS) result from recording data over time.
- ❖ Increasingly popular due to the growing importance of automatic sensors producing an increasing flood of large, high-resolution TS.
- ❖ Application areas: motion sensors, personalized medicine (ECG/EEG signals), machine surveillance, spectrograms, astronomy (starlight-curves), and image outlines / contour of objects.



- ❖ UCR time series archive contains 85 benchmark datasets used in TS research.
- ❖ Datasets from a whole range of application, grouped by: synthetic, motion sensors, sensor readings and image outlines.
- ❖ Overall, there are 50.000 train and 100.000 test TS or 55 million values.
- ❖ At most thousands of TS with thousands of measured values for a single dataset.



UCR Time Series Classification Archive

Last major update, Summer 2015: [0237918](#), and it continues to be fur

We suggest you begin by reading the briefing document in [PDF](#) or [PowerPoint](#), which also contains the pass zipped format).

Please reference as: Yanping Chen, Eamonn [Keogh](#), Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah *Classification Archive*. URL www.cs.ucr.edu/~eamonn/time_series_data/

```
@misc{UCRArchive,
title={The UCR Time Series Classification Archive},
author={ Chen, Yanping and Keogh, Eamonn and Hu, Bing and Begum, Nurjahan and Bagnall, Anthony and Mueen, Abdullah and Batista, G
year={2015},
month={July},
note = {\url{www.cs.ucr.edu/~eamonn/time_series_data/}}
}
```

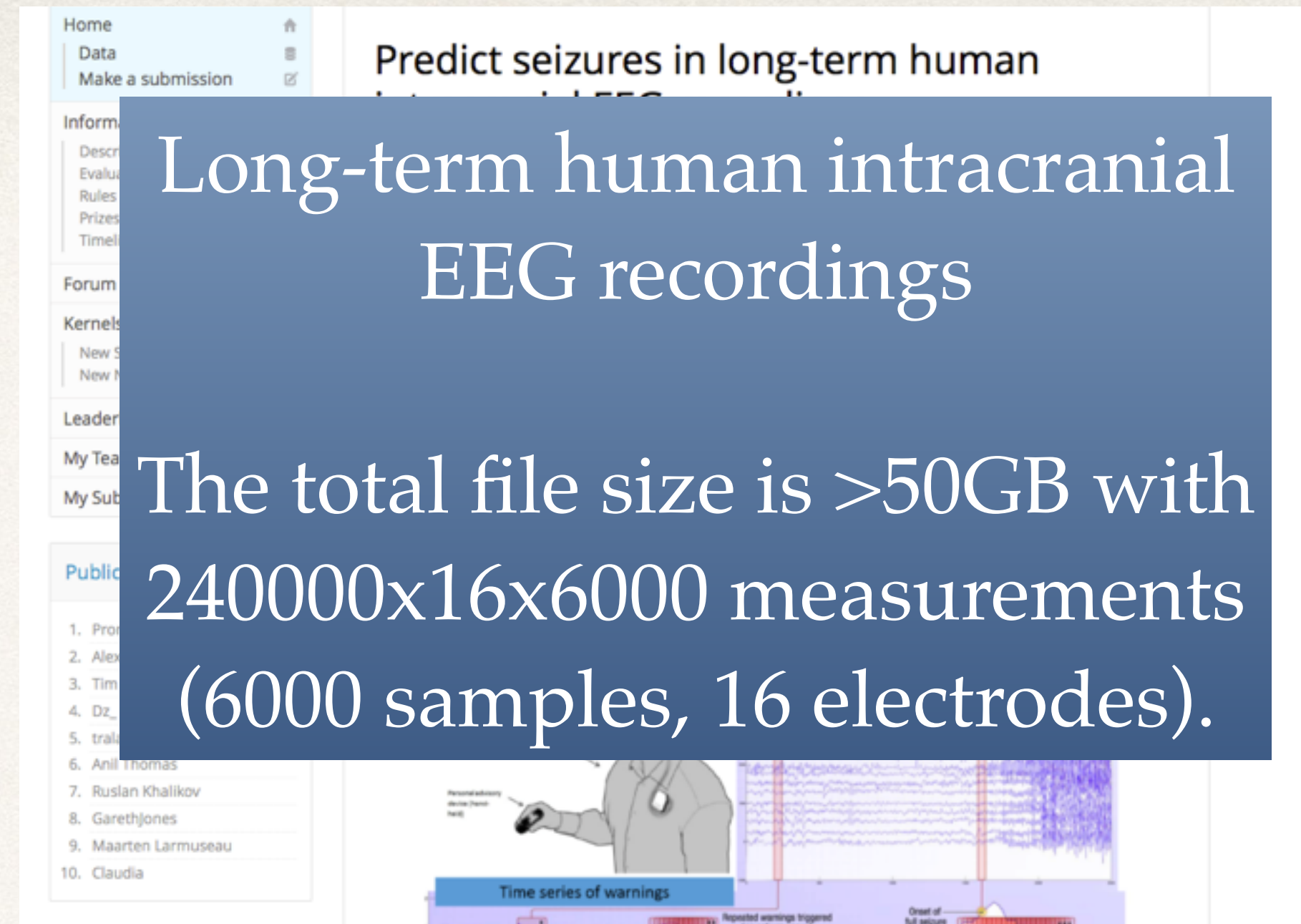
Name	First paper or data creator	Number of classes	Size of training set	Size of testing set	Time series Length	1-NN Euclidean Distance
Synthetic Control	Pham	6	300	300	60	0.120
Gun-Point	Ratanamahatana	2	50	150	150	0.087
CBF		3	30	900	128	0.148
Face (all)	Xi	14	560	1690	131	0.286
OSU Leaf	Gandhi	6	200	242	427	0.479
Swedish Leaf	Soderkvist	15	500	625	128	0.211
50Wanda	Bath	50	150	155	270	0.260

- ❖ At the same time real-time systems emerge: Billions of measurements for thousands of sensors.

Predict seizures in long-term human

Long-term human intracranial EEG recordings

The total file size is >50GB with 240000x16x6000 measurements (6000 samples, 16 electrodes).



The DEBS 2014 Grand Challenge

Zbigniew Jerzak
SAP AG
Chemnitzer Straße 48
01187 Dresden, Germany

Holger Ziekow
AGT International
Hilpertstr. 35
64295 Darmstadt, Germany

Smart Plugs

„4055 Millions of measurements for 2125 plugs distributed across 40 houses.“

Keywords
event processing, streaming, utilities

1. INTRODUCTION

real-time analytics over high volume sensor data. The underlying scenario stems from the smart grid domain and targets the analysis of energy consumption measurements. Specifically, the DEBS 2014 Grand Challenge focuses on

The DEBS 2013 Grand Challenge

Christopher Mutschler
University of
Erlangen-Nuremberg
and
Fraunhofer Institute for

Holger Ziekow
AGT International
Hilpertstr. 35
64295 Darmstadt, Germany
HZiekow@agtinternational.com

Zbigniew Jerzak
SAP AG
Chemnitzer Straße 48
01187 Dresden, Germany
Zbigniew.Jerzak@sap.com

Real-Time Location System

„The total filesize is 2.6 GB and it contains a total of 49,576,080 position events.“

General Terms
Performance, Experimentation

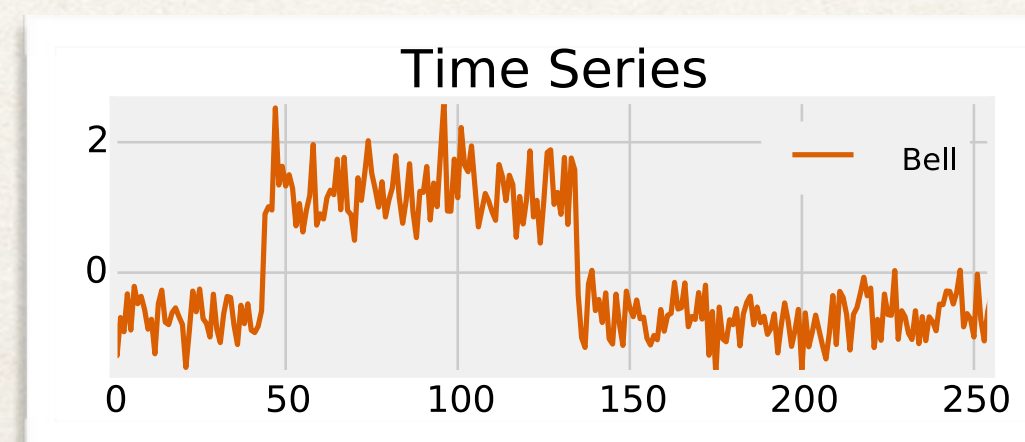
Keywords
event processing, streaming, utilities

2013 Grand Challenge.
The remainder of this paper is structured as follows: in Section 2 we present the technical details of the real-time location system RedFIR which was used for collecting the raw data. In Section 3 we provide a detailed description of the recorded raw data. In Section 4 we provide a description

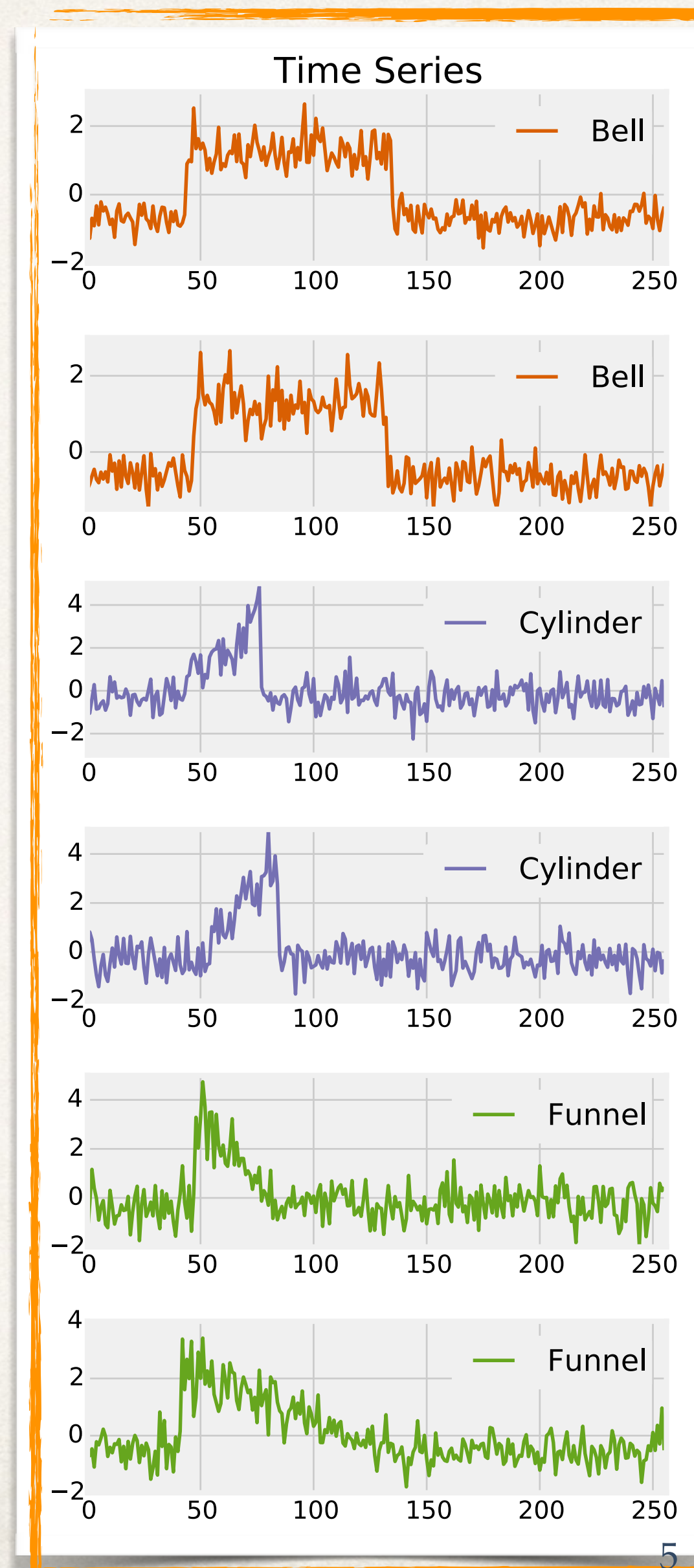
Model

- ❖ Time series classification (TSC) aims at assigning a class label to an unlabeled *query* TS based on a *model* trained from labeled samples.
- ❖ Most basic: 1-nearest neighbor classifiers.
- ❖ We look into the four groups of TS classifiers: whole series, shapelets, bag-of-patterns, and ensembles.

Query

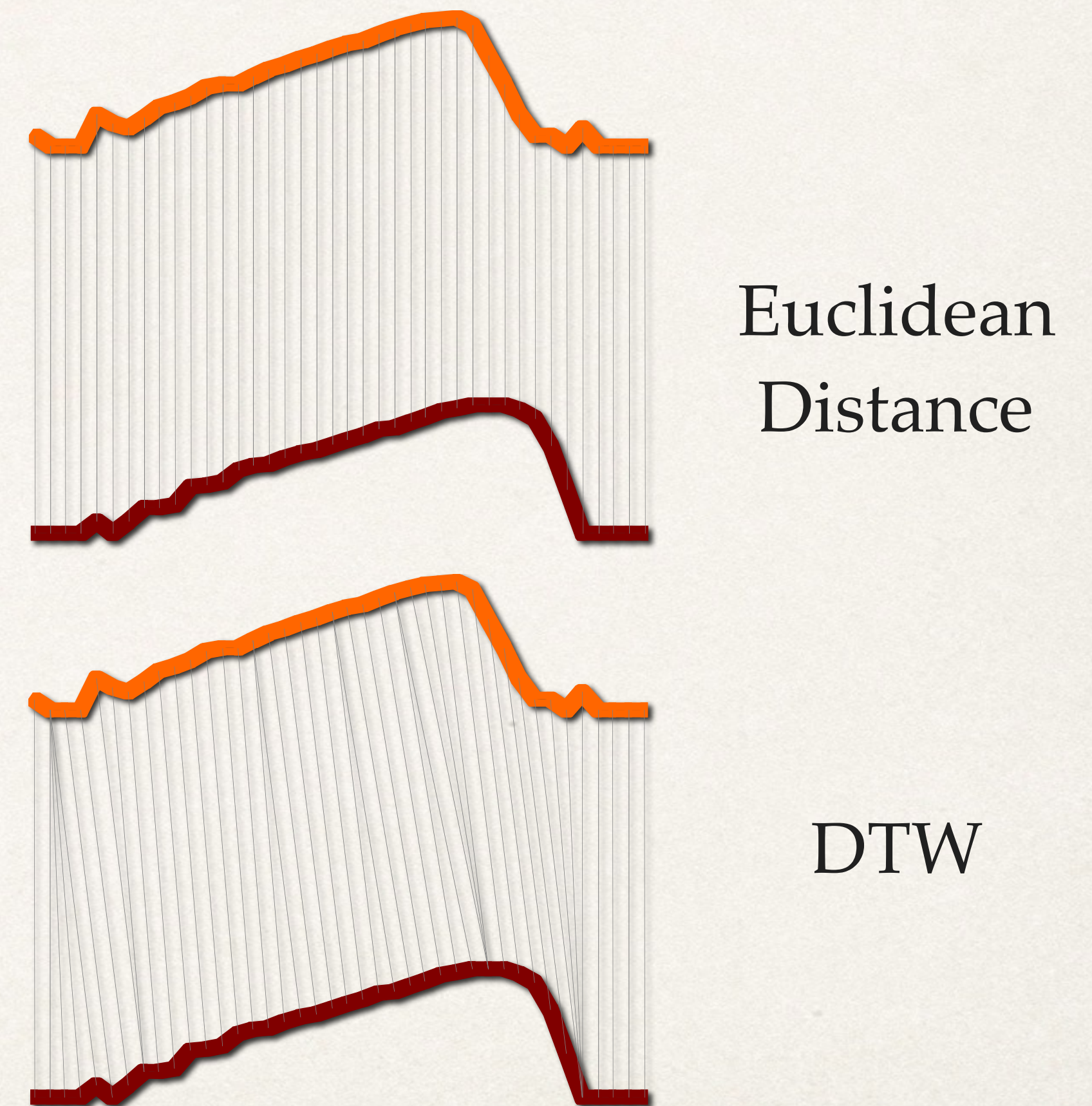


find
label



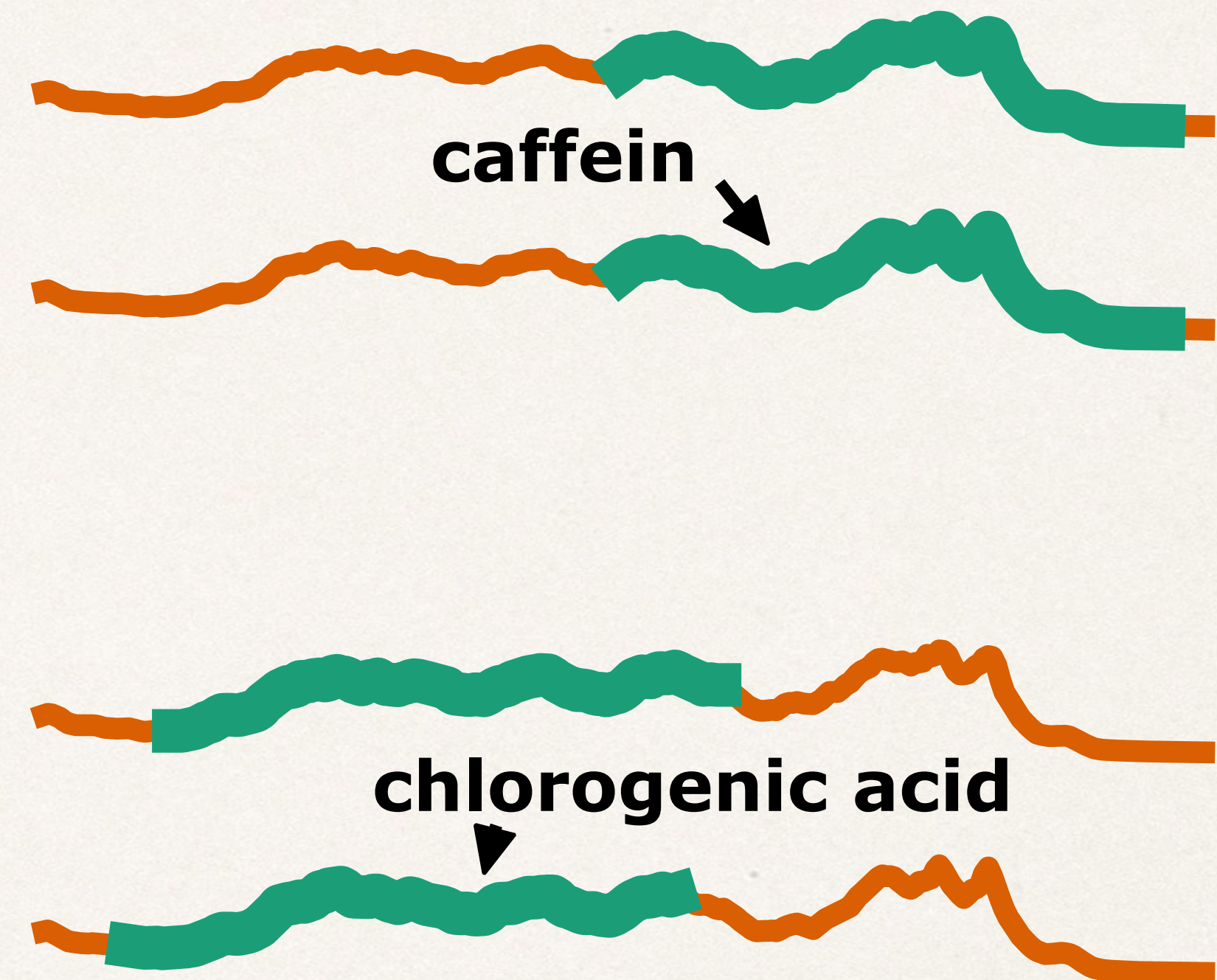
Whole Series

- ❖ Based on a distance measure defined on the whole TS data and 1-NN classification.
- ❖ *Elastic* distance measures compensate for small differences like warping in the time axis.
- ❖ Base-line, simple model, cannot skip irrelevant subsections, linear to quadratic complexity in TS length.
- ❖ Representatives: 1-NN Dynamic Time Warping (DTW) and 1-NN Euclidean distance (ED).



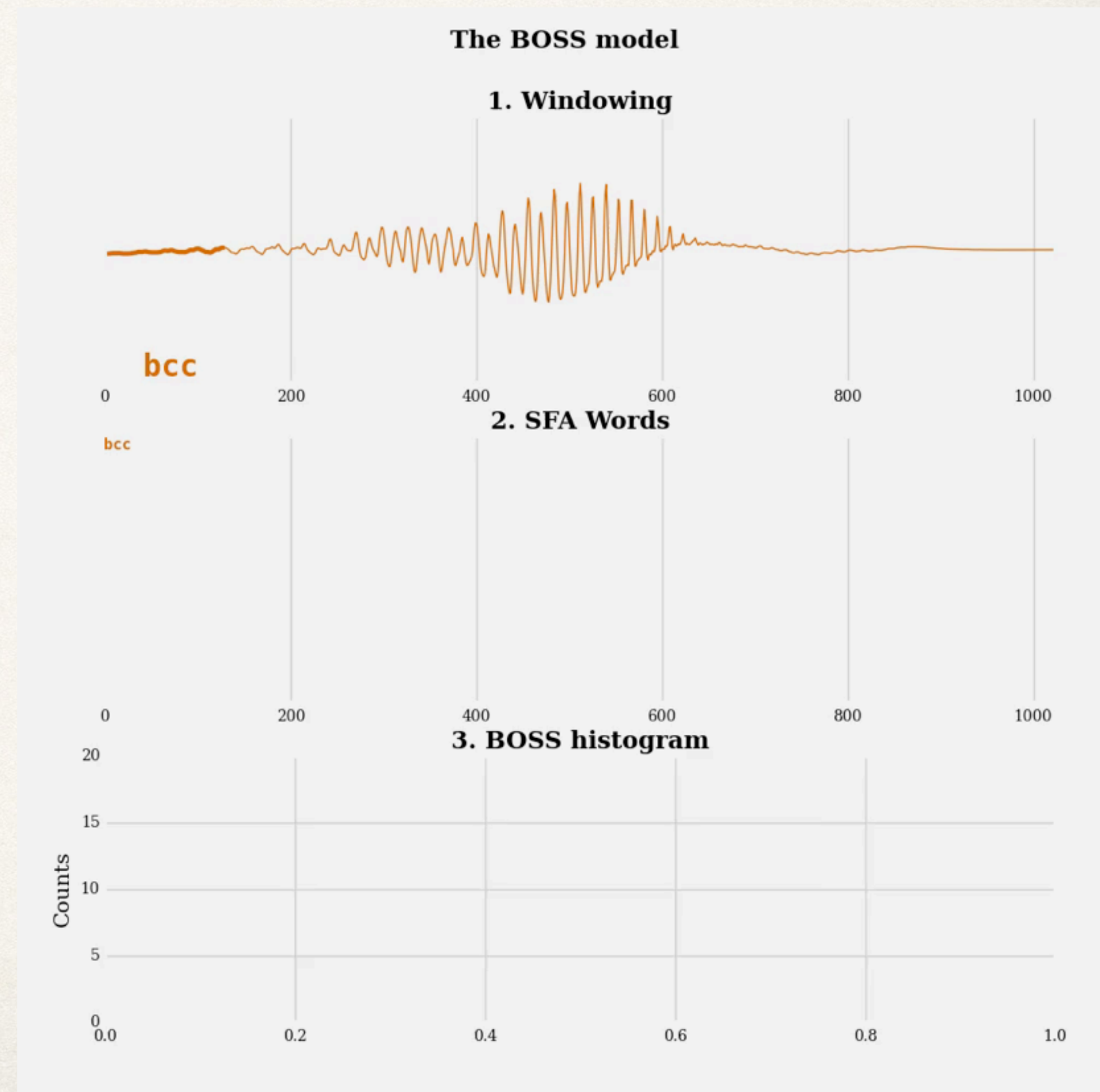
Shapelets

- ❖ Shapelets are TS subsequences that are maximally representative of a class label.
- ❖ A TS is labeled based on the similarity to a shapelet.
- ❖ Interpretable, high computational complexity (cubic to bi-quadratic in TS length).
- ❖ Representatives: Shapelet Transform (ST), Learning Shapelets (LS), Fast Shapelets (FS).



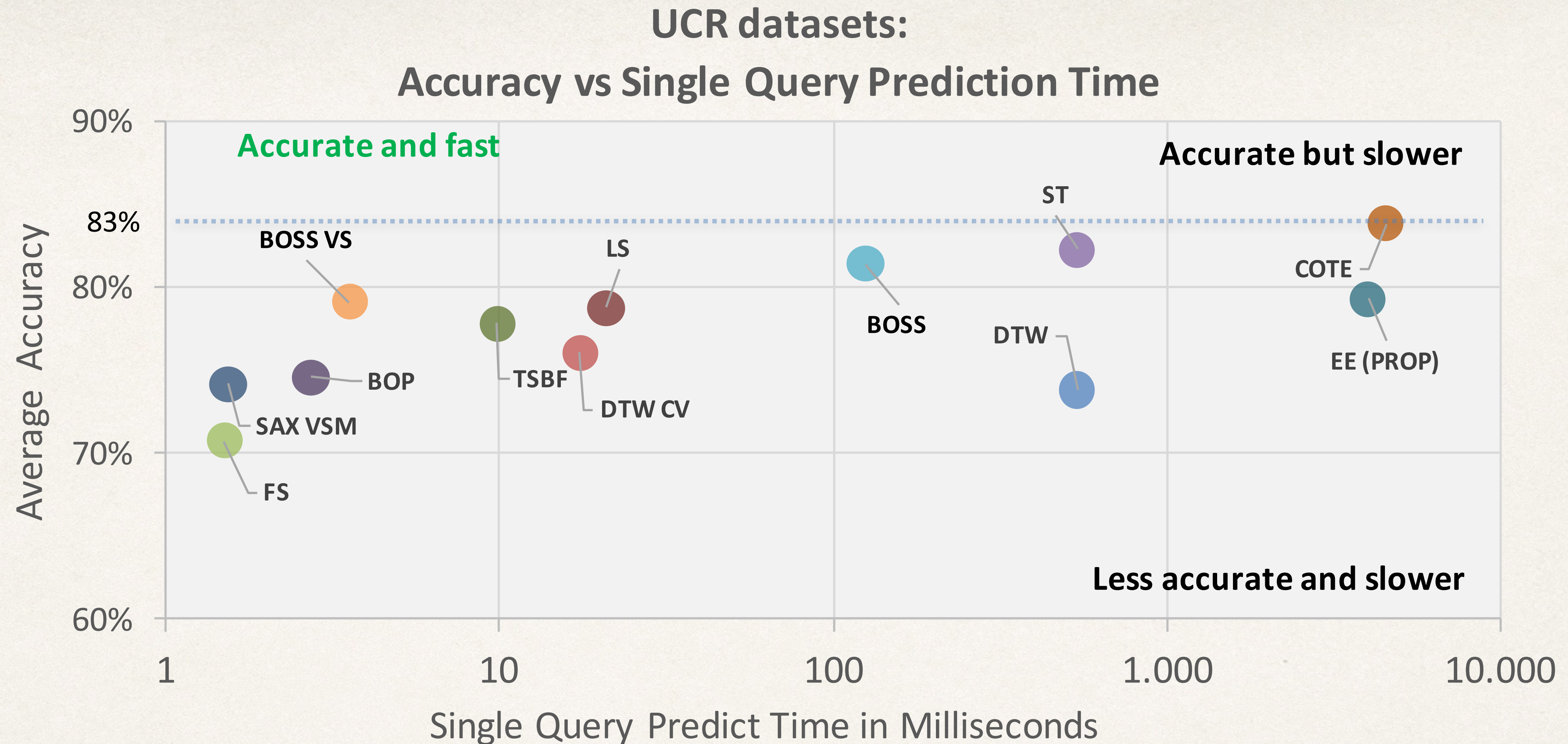
Bag-of-Patterns / Bag-of-Features

- ❖ TS are distinguished by *the frequency of occurrence* of features generated over substructures of the TS.
- ❖ A bag-of-patterns (histogram) of feature counts is used as input to classification.
- ❖ Fast (linear complexity), noise reducing, but order of substructures gets lost.
- ❖ Representatives: Bag-of-SFA-Symbols (BOSS), Bag-of-Patterns (BoP), Time Series Bag of Features (TSBF).



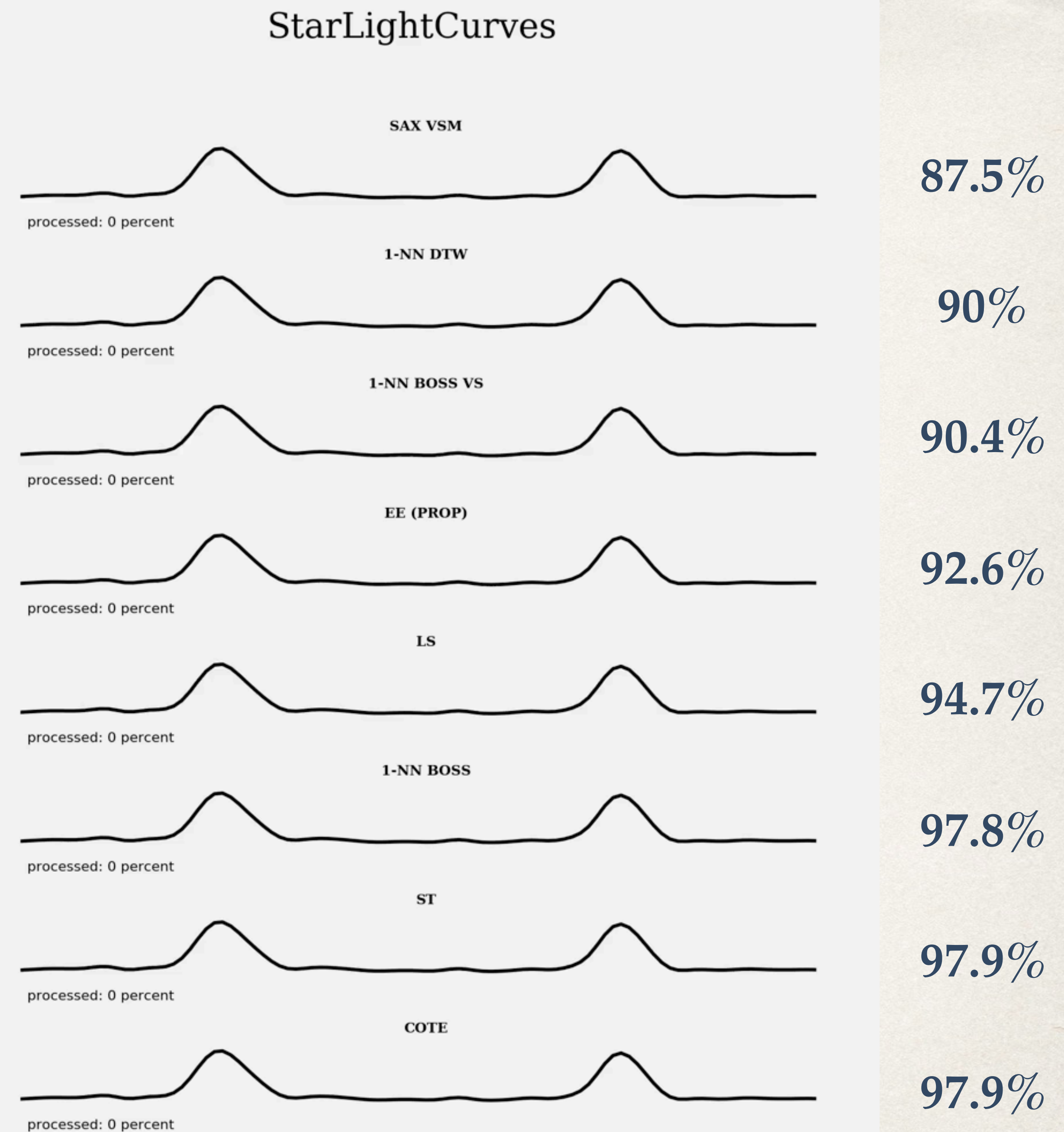
Ensembles

- ❖ Ensembles combine different core classifiers (i.e., shapelets, bag-of-patterns, whole series) into a single classifier using bagging or majority voting.
- ❖ High accuracy by combining different representations but high computational complexity (quadratic to bi-quadratic in TS length).
- ❖ Representatives: Elastic Ensemble (EE PROP), Collective of Transformation Ensembles (COTE).

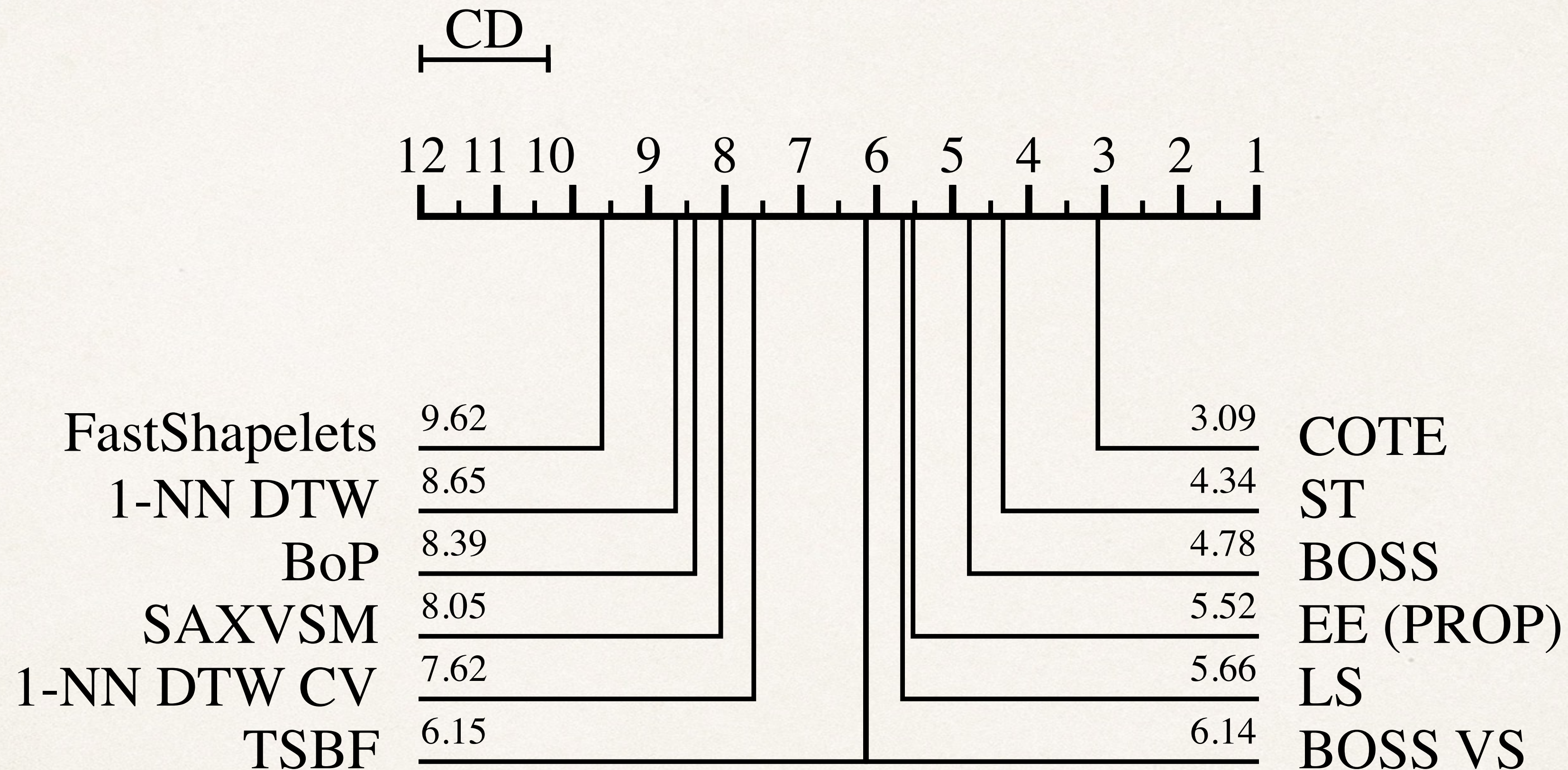


- ❖ Slowest (fastest) classifier took 4s (2ms).
- ❖ Methods are either scalable but offer only inferior accuracy, or they achieve state-of-the-art accuracy but do not scale to larger dataset sizes.

- ❖ Prediction times of state of the art.
- ❖ Using StarLightCurves dataset with 1000 train and 8236 test TS of length 1024.
- ❖ Video runs at 10x playback speed.
- ❖ Slowest classifier took 100 hours. Fastest took 20 ms.



Average Ranks on 85 UCR datasets



- ❖ Most accurate TSCs are Ensembles, Shapelets and Bag-of-Patterns: COTE, ST, BOSS and EE.

Conclusion

- ❖ Methods are either scalable but offer only inferior accuracy, or they achieve state-of-the-art accuracy but do not scale to larger dataset sizes.
- ❖ Bag-of-Patterns approaches are faster than Shapelets, Ensembles or Whole Series Measures.
- ❖ Overall, COTE, ST and BOSS show the highest classification accuracy at the cost of increased runtimes.
- ❖ FS, SAX VSM, BOP, BOSS VS show the lowest runtimes at the cost of limited accuracy.