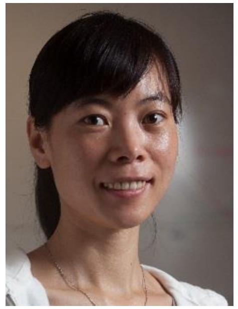
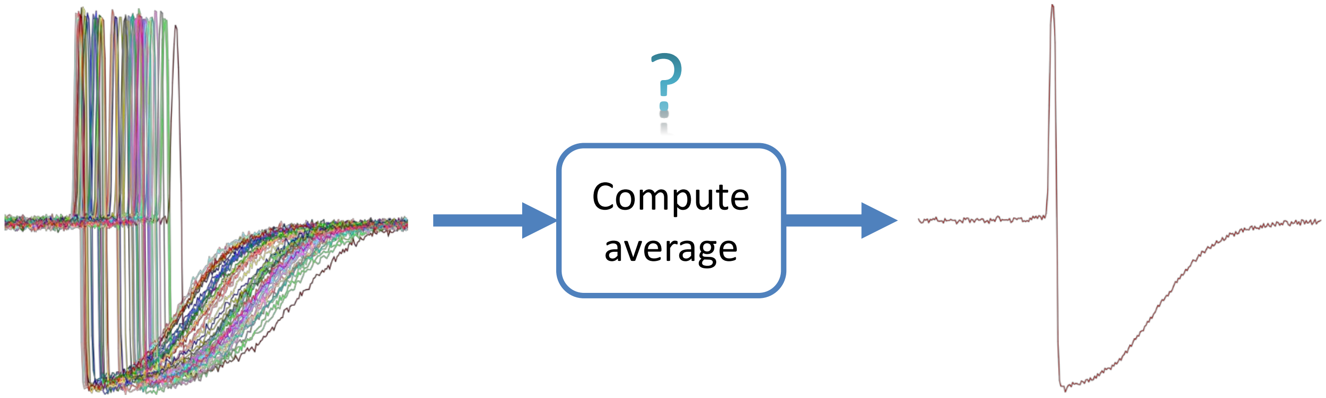
Dynamic Time Warping Averaging of Time Series allows Faster and more Accurate Classification








F. Petitjean G. Forestier G.I. Webb A.E. Nicholson Y. Chen E. Keogh



Dec 14-17, 2014
ICDM
SHENZHEN, CHINA

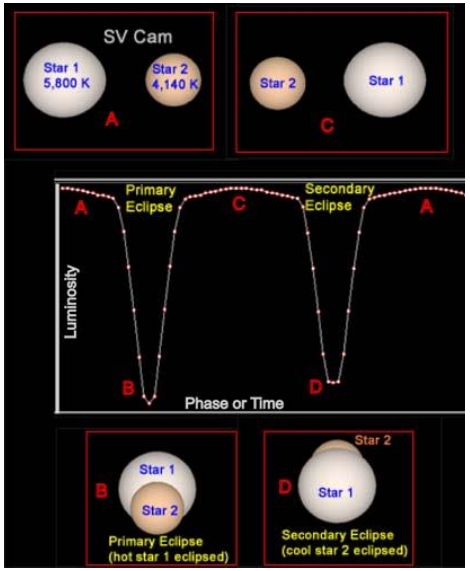
The Ubiquity of Time Series

Sensors on machines

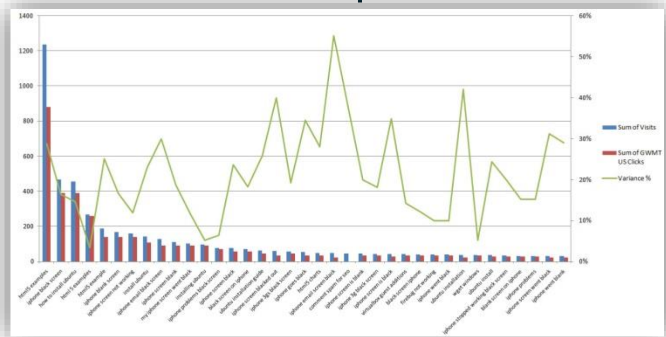
Stock prices

Wearables

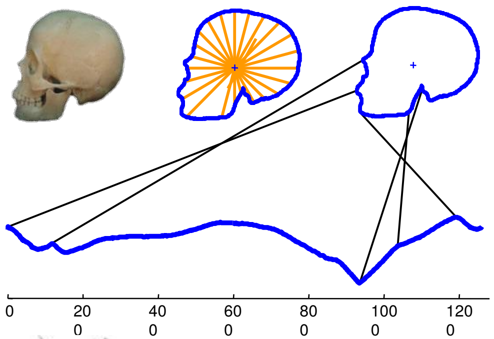
Astronomy: star light curves



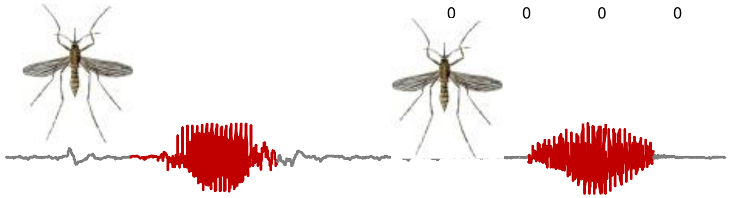
Web clicks

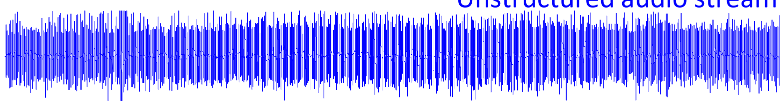


Shapes



Sound





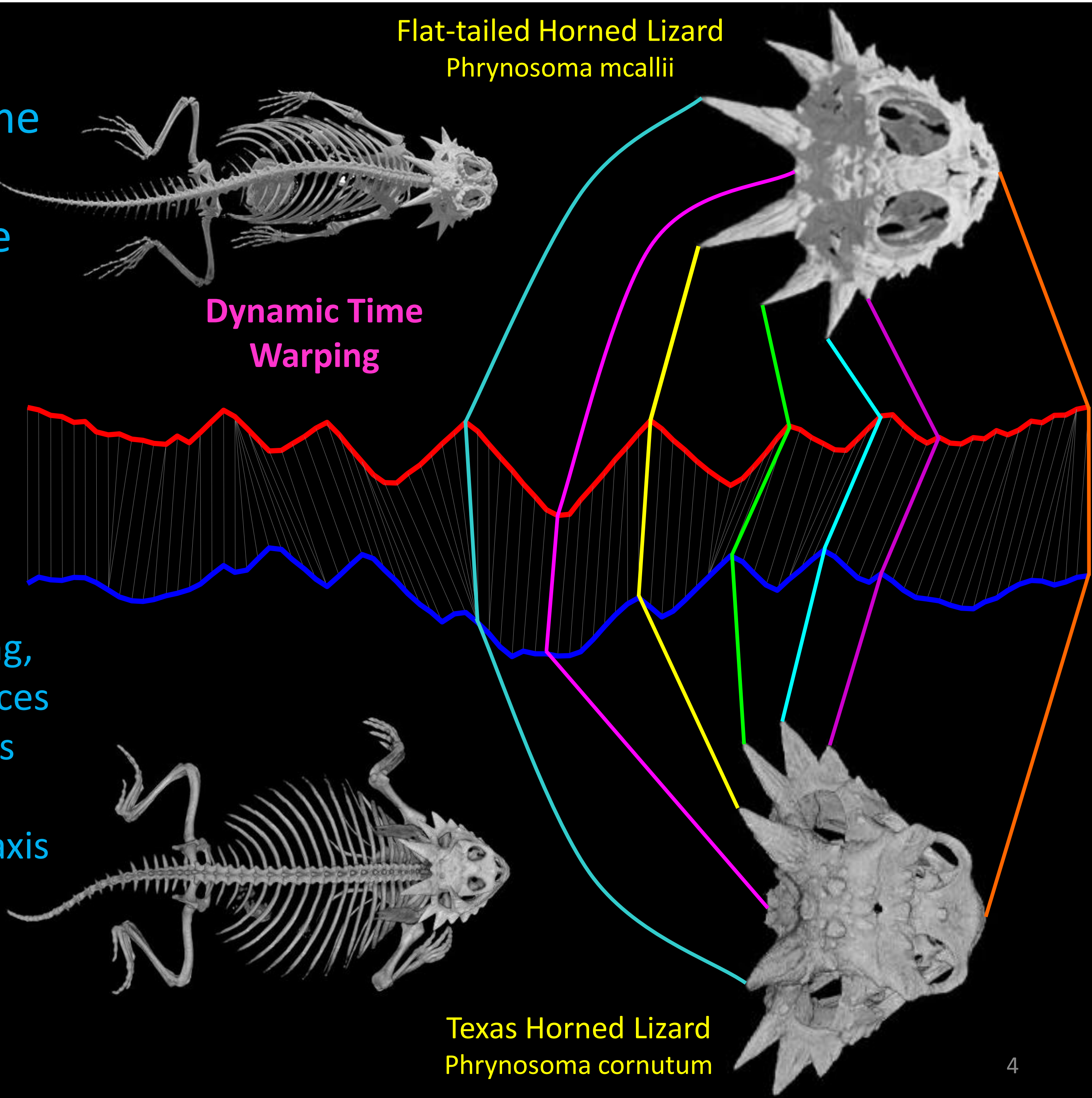
Slightly Surprising Facts

1. The *Nearest Neighbor algorithm* is virtually always most accurate for time series classification.
2. *Dynamic Time Warping* (DTW) is the most accurate measure for time series across a huge variety of domains.

This is not a place to discuss *why* this is true (see [a,b,c]), but this is the strong consensus of the community, supported by large-scale reproducible experiments.

[a] A. Bagnall and J. Lines, "An experimental evaluation of nearest neighbour time series classification. technical report #CMP-C14-01," Department of Computing Sciences, University of East Anglia, Tech. Rep., 2014.
[b] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Int. Conf. on Machine Learning*, 2006, pp. 1033–1040.
[c] X. Wang, A. Mueen, H. Ding, G.Trajcevski, P. Scheuermann, E. Keogh: Experimental comparison of representation methods and distance measures for time series data. Data Min. Knowl. Discov. 26(2): 275-309 (2013)

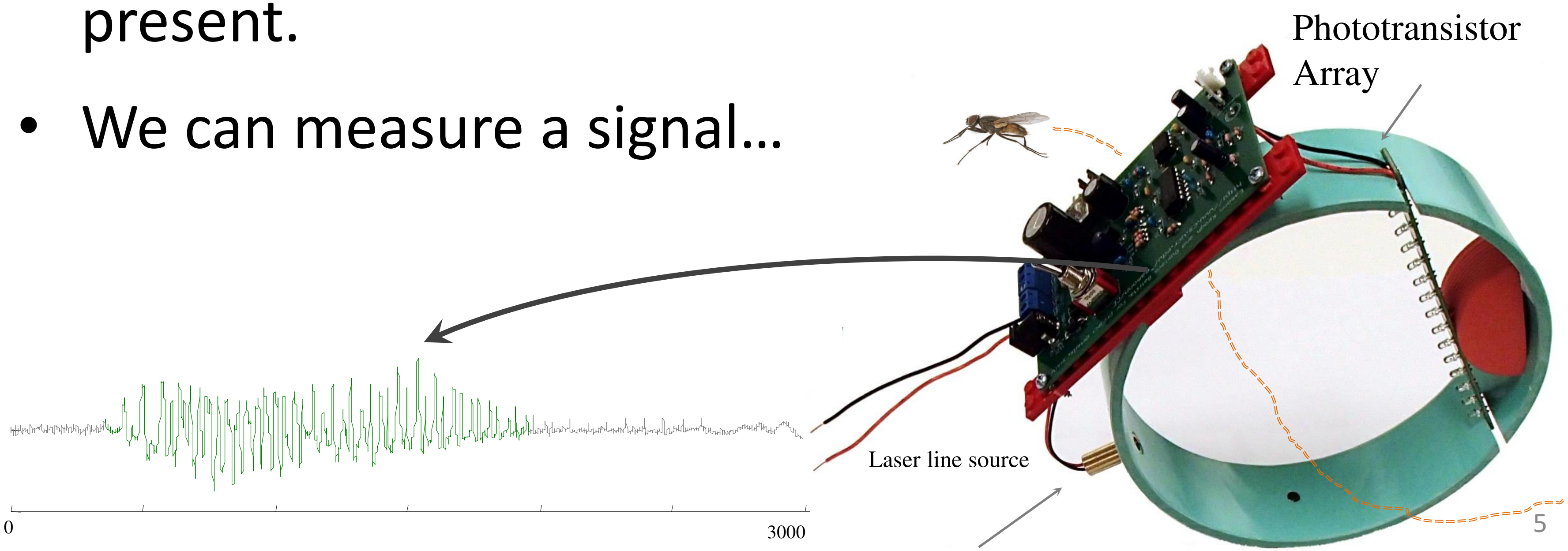
DTW works well even if the two time series are *not* well aligned in the time axis.



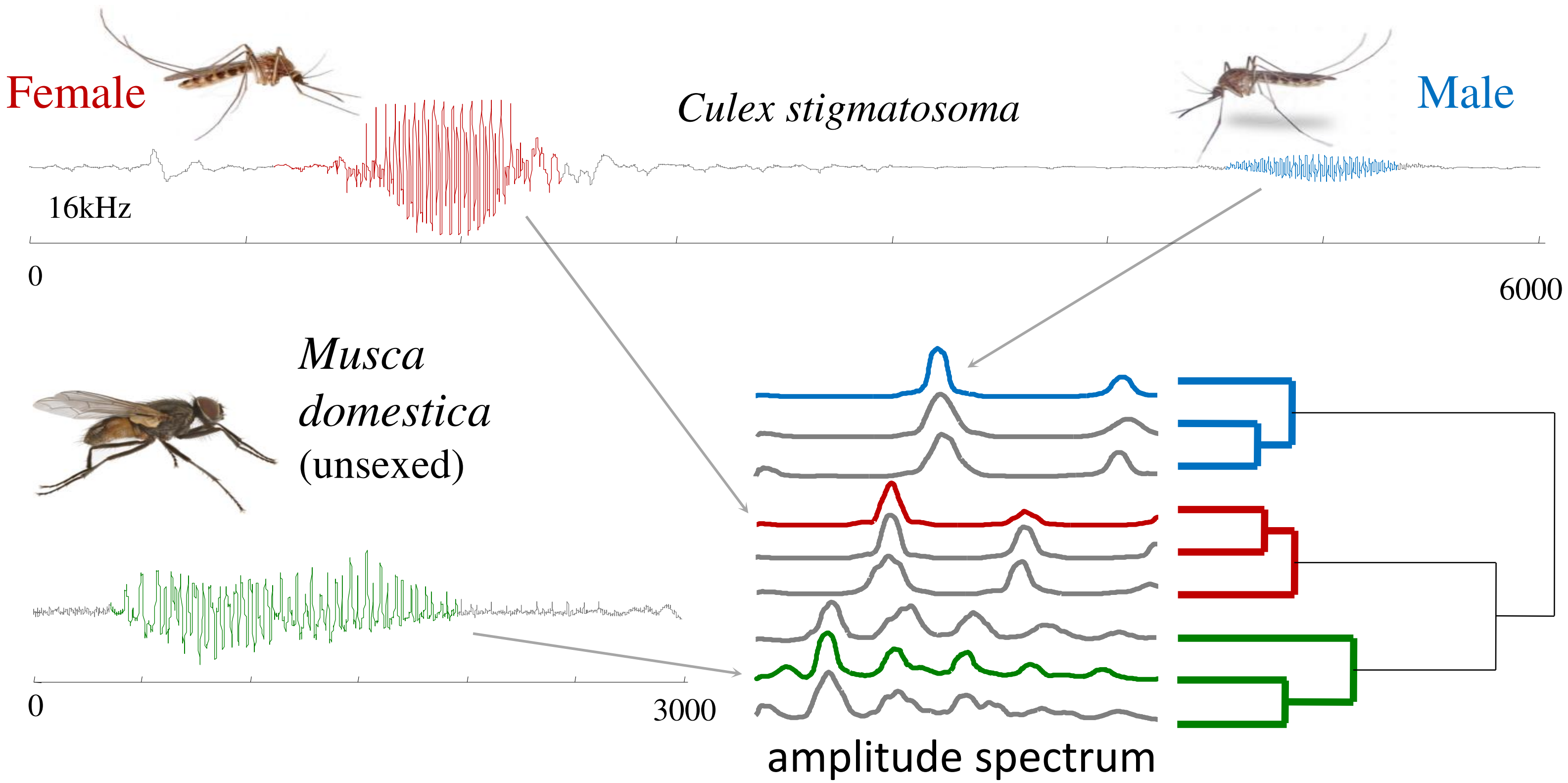
Without time warping, insignificant differences in time axis appear as very significant differences in the Y-axis

Case Study: Classifying Flying Insects

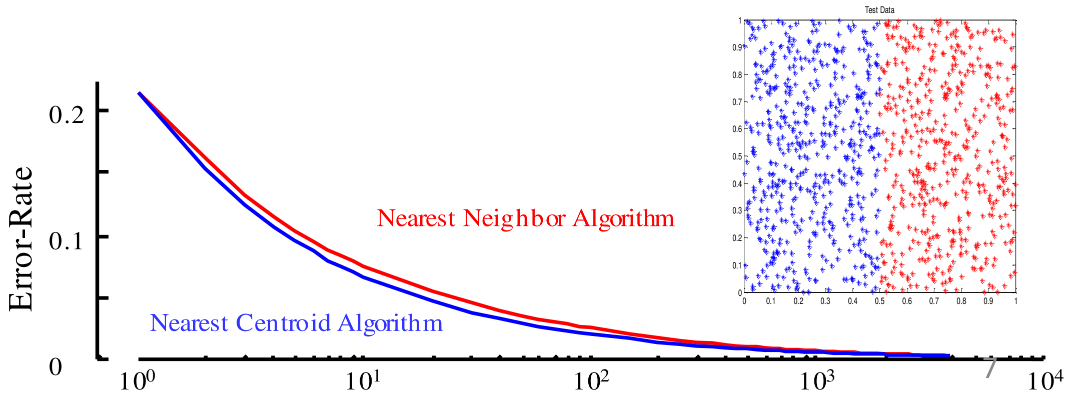
- Insects kill about a million people each year
- Insects destroy tens of billions of dollars' worth of food each year
- To mitigate insect damage we must determine which sex/species are present.
- We can measure a signal…



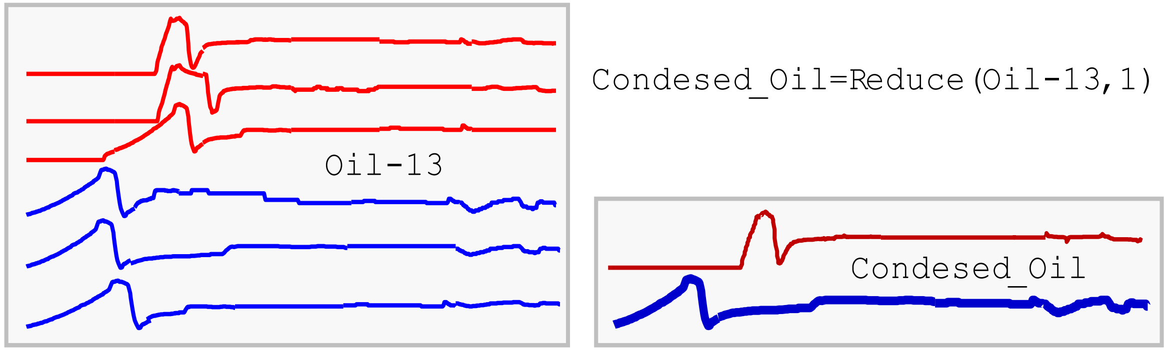
- The "audio" of insect flight can be converted to an amplitude spectrum, which is essentially a time series.
- As the dendrogram hints at, this does seem to capture some class specific information…



- If we are going to put devices into the field, there are going to be resource constraints.
- One solution is to average our large training dataset into a small number of prototypes.
- This:
    - Will speed up NN classification
    - May be more accurate, since averaging can produce prototypes that capture the essence of the set

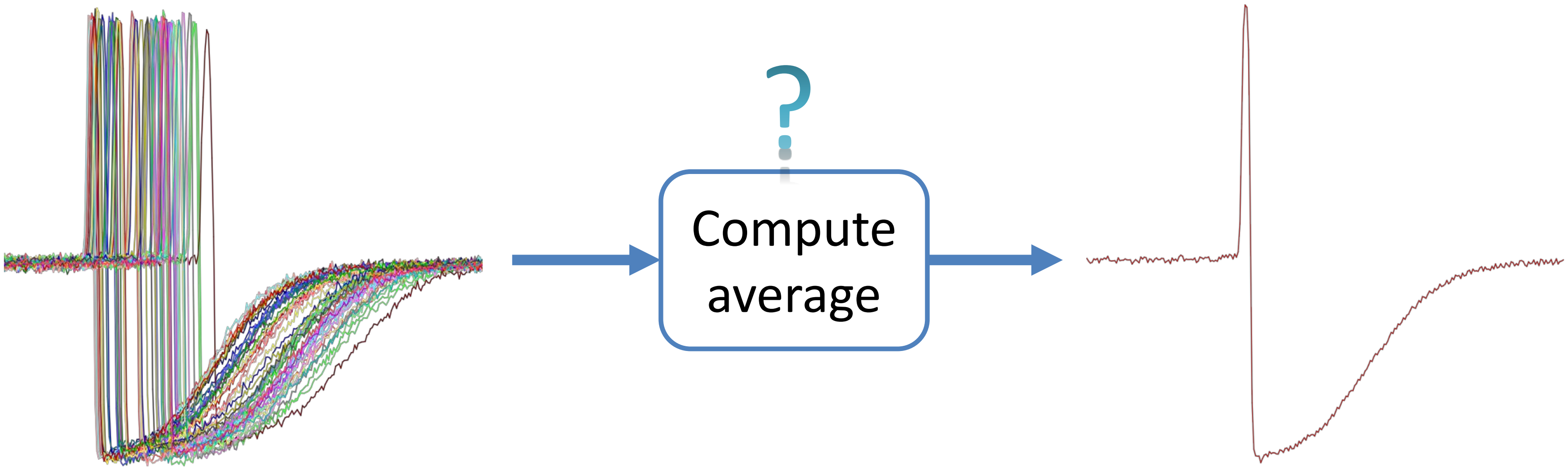


Our idea for a fast and accurate classification system:



The issue is then:

- ➢ *How to average time series consistently with DTW?*



What is the mean of a set?

Averaging is the tool that makes it possible to define a prototype informing about the central tendency of a set in its space.

Mathematically, the mean $\bar{o}$ of a set of objects O embedded in a space induced by a distance d is:

$$\arg\min_{\bar{o}} \sum_{o \in O} d^2(\bar{o}, o)$$

The mean of a set minimizes the sum of the squared distances.

Optimization problem

$$\arg\min_{\bar{o}} \sum_{o \in O} d^2(\bar{o}, o)$$

If d is the Euclidean distance

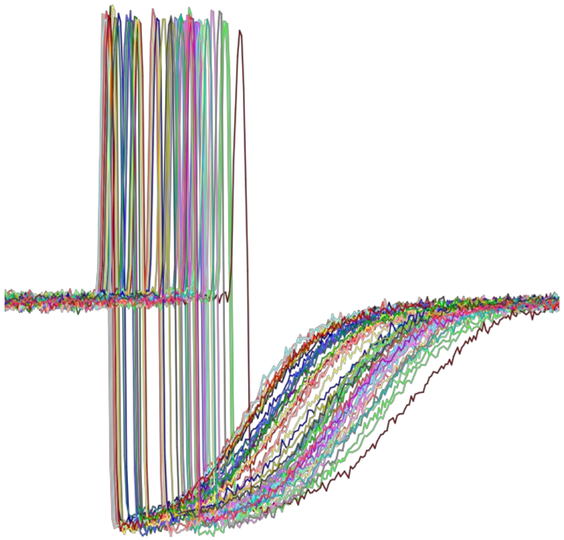
The *arithmetic mean* solves the problem exactly

$$\bar{o} = \frac{1}{N} \sum_{o \in O} o$$

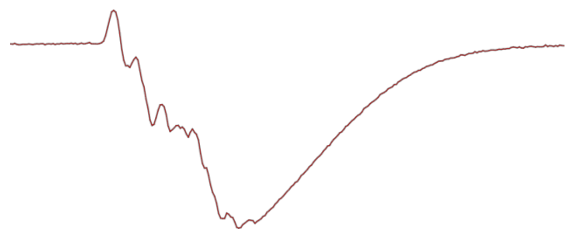
If d is **DTW**

The *arithmetic mean* does **not** solve the problem

This is **not surprising**, because the arithmetic mean does not take **warping** into account!



Arithmetic mean

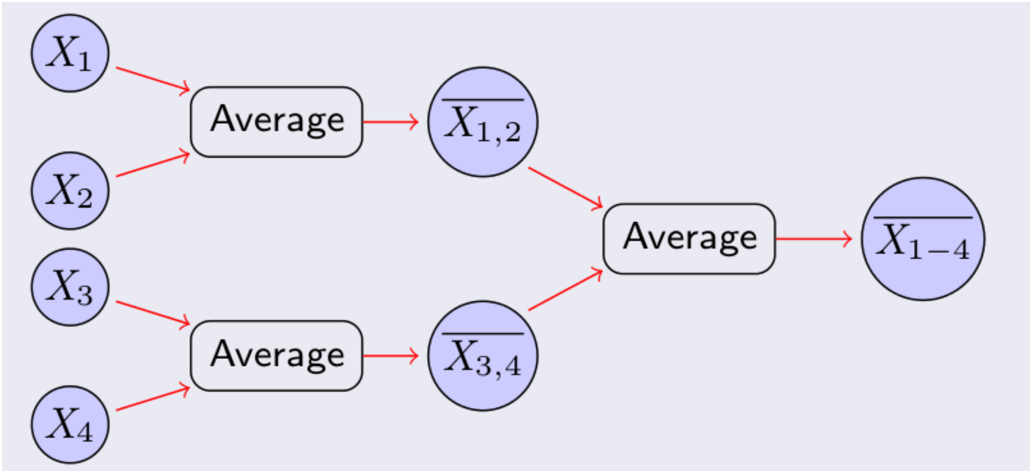


State of the art in averaging for DTW

Main idea exploited [a][b][c][d] and more:

We know how to exactly compute the average of 2 sequences…

…so we can build the average pairwise.



But, this only works if the operator is associative…

…which is *not* the case for DTW pairwise average.

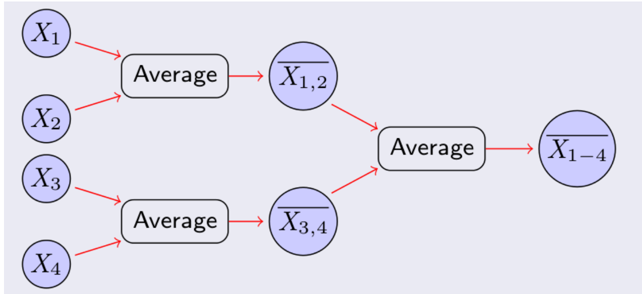
[a] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, "Nonlinear alignment and averaging for estimating the evoked potential," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 4, pp. 348–356, 1996.

[b] V. Niennattrakul and C. A. Ratanamahatana, "On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping," IEEE International Conference on Multimedia and Ubiquitous Engineering, pp.733-738, 2007.

[c] S. Ongwattanakul and D. Srisai, "Contrast enhanced dynamic time warping distance for time series shape averaging classification," in *Int. Conf. on Interaction Sciences: Information Technology, Culture and Human,* ACM, 2009, pp. 976–981.

[d] V. Niennattrakul and C. A. Ratanamahatana, "Shape averaging under time warping," in *Int. Conf. on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, IEEE, vol. 2, 2009, pp. 626–629.

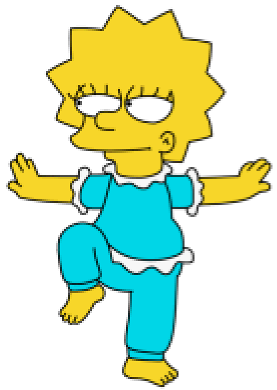
Pairwise averaging is not good enough:



1. Even the medoid sequence often provides a better solution than state-of-the-art methods [a]
2. Using k-means, centers often "drift out" of the cluster [b]

We are seeking a solution that would not rely on associativity

- ➢ **No pairwise** methods



[a] F. Petitjean and P. Gançarski, "Summarizing a set of time series by averaging: From Steiner sequence to compact multiple alignment," *Theoretical Computer Science, 2012.*

[b] V. Niennattrakul and C. A. Ratanamahatana, "Inaccuracies of Shape Averaging Method Using Dynamic Time Warping for Time Series Data," *International Conference on Computational Science*, 2007.

Back to the source

- DTW is the extension of the edit distance to sequences of numerical values (time series).
- Finding a “consensus” sequence is a very close problem to the one of defining an average sequence for DTW (same objective function).
- Having the multiple alignment (≈ simultaneous alignment) of a set of sequences.

  ⇒ consensus sequence computable “column by column”

```
------D-PGDF--DRNVPRICGVCGDRATGFHFNAMTCEGCKGFFRRSMKRKA--LFTCP-FNGDCRITKDNRRHCQACRLKRCVDIGMMKEFILTD
IRPQKRK-KGPAP-KMLGNELCSVCGDKASGFHYNVLSCEGCKGFFRRSVIKGA--HYICH-SGGHCPMDTYMRRKCQECRLRKCRQAGMREECVLSE
SVPGKPS-VNADE-EVGGPQICRVCGDKATGYHFNVMTCEGCKGFFRRAMKRNA--RLRCPFRKGACEITRKTRRQCQACRLRKCLESGMKKEMIMSD
EPERKRK-KGPAP-KMLGHELCRVCGDKASGFHYNVLSCEGCKGFFRRSVVRGGARRYACR-GGGTCQMDAFMRRKCQQCRLRKCKEAGMREQCVLSE
PVTKKPRMGASAG-RIKGDELCVVCGDRASGYHYNALTCEGCKGFFRRSITKNA--VYKCK-NGGNCVMDMYMRRKCQECRLRKCKEMGMLAECMYTG
QTEEKKC-KGYIPSYLDKDELCVVCGDKATGYHYRCITCEGCKGFFRRTIQKNLHPSYSCK-YEGKCVIDKVTRNQCQECRFKKCIYVGMATDLVLDD
----SPS-PPPPP---RVYKPCFVCNDKSSGYHYGVSSCEGCKGFFRRSIQKNM--VYTCH-RDKNCIINKVTRNRCQYCRLQKCFEVGMSKEAVRND
----PPS-PLPPP---RVYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNM--IYTCH-RDKNCVINKVTRNRCQYCRLQKCFEVGMSKESVRND
----PPS-PPPLP---RIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNM--VYTCH-RDKNCIINKVTRNRCQYCRLQKCFEVGMSKESVRND
```

Multiple alignment, consensus sequence and average time series

Multiple alignment example

$A = \langle a, c, a, a, b \rangle$
$B = \langle a, a, c, a, a \rangle$
$C = \langle a, a, a, c, a \rangle$

A	a	a	a	c	a	a	b
B	a	a	a	c	a	a	a
C	a	a	a	c	a	a	a
M	a	a	a	c	a	a	a

Same result for time series

$A = \langle 1, 10, 0, 0, 4 \rangle$
$B = \langle 0, 2, 10, 0, 0 \rangle$
$C = \langle 0, 0, 0, 10, 0 \rangle$

A	1	1	1	10	0	0	4
B	0	0	2	10	0	0	0
C	0	0	0	10	0	0	0
M	$\frac{1}{3}$	$\frac{1}{3}$	1	10	0	0	$\frac{4}{3}$

But, finding the optimal multiple alignment:

1. Is **NP-complete** [a]
2. Requires $\boldsymbol{O(L^N)}$ operations
    - L is the length of the sequences (≈ 100)
    - N is the number of sequences ($\approx$ 1,000)

$\gg 10^{85}$

#particles in the observable universe

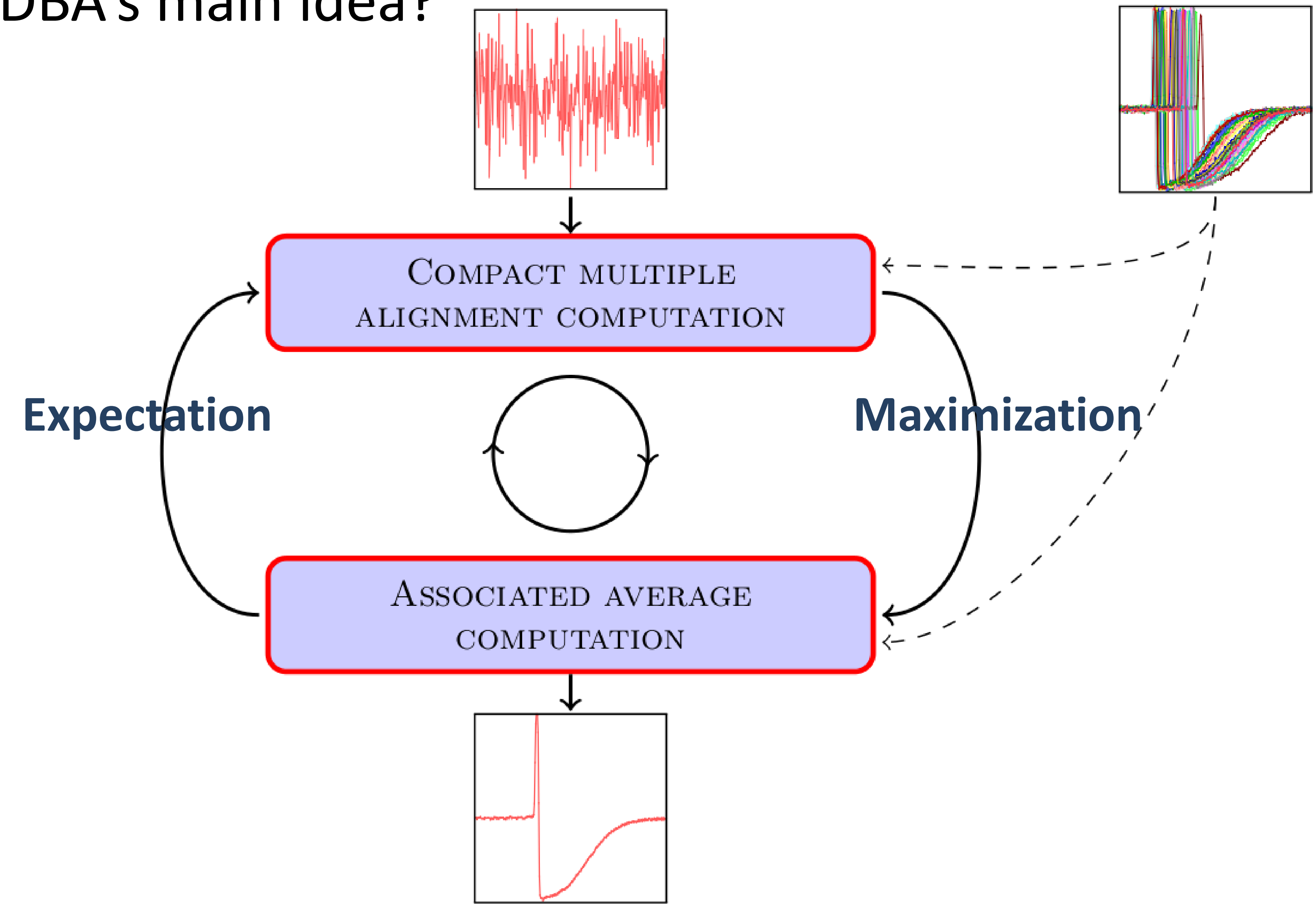
$\Rightarrow$ Efficient solutions will be heuristic

In 2011, we introduced DBA [a]:

- Takes inspiration from works in computational biology
- Is specifically designed for time series and DTW
- Does not function pairwise
- Does not use any order on the dataset it averages

[a] F. Petitjean, A. Ketterlin and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

DBA's main idea?



[a] F. Petitjean, A. Ketterlin and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

We have shown that (see the paper and [a]):

1. DBA outperforms **all** state-of-the-art methods

2. DBA improves on the optimization problem by **30%**

Optimization problem

$$\arg\min_{\bar{o}} \sum_{o \in O} d^2(\bar{o}, o)$$

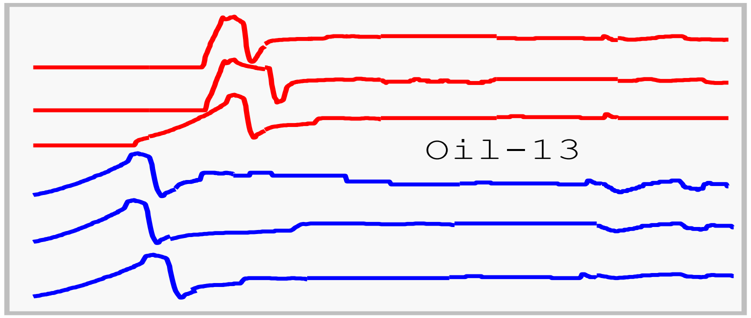
3. DBA converges between iterations
4. No centers "drifting out" of the cluster

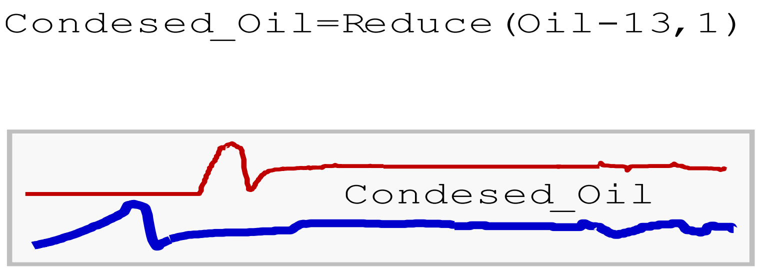
[a] F. Petitjean, A. Ketterlin and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

Experiments

Objective: Making 1NN with DTW faster

Mean: Condensing the "train" dataset with DBA





2 average-based techniques

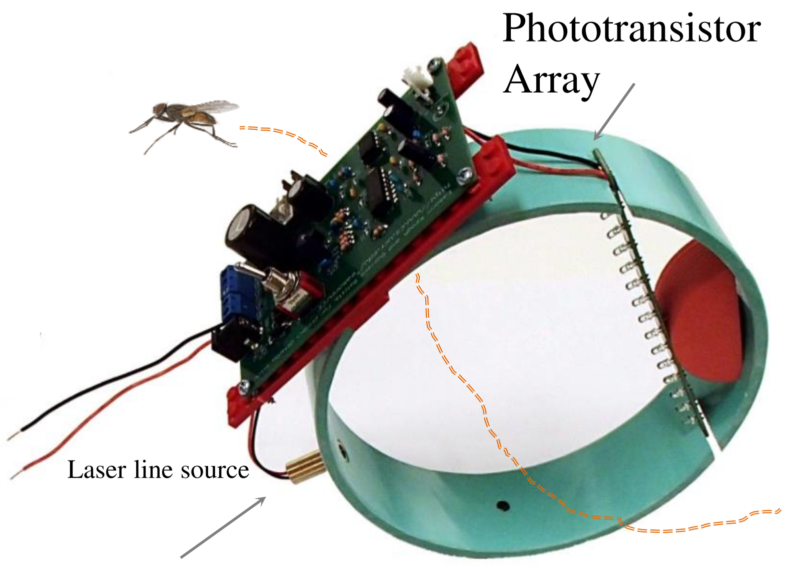
1. K-means
2. AHC

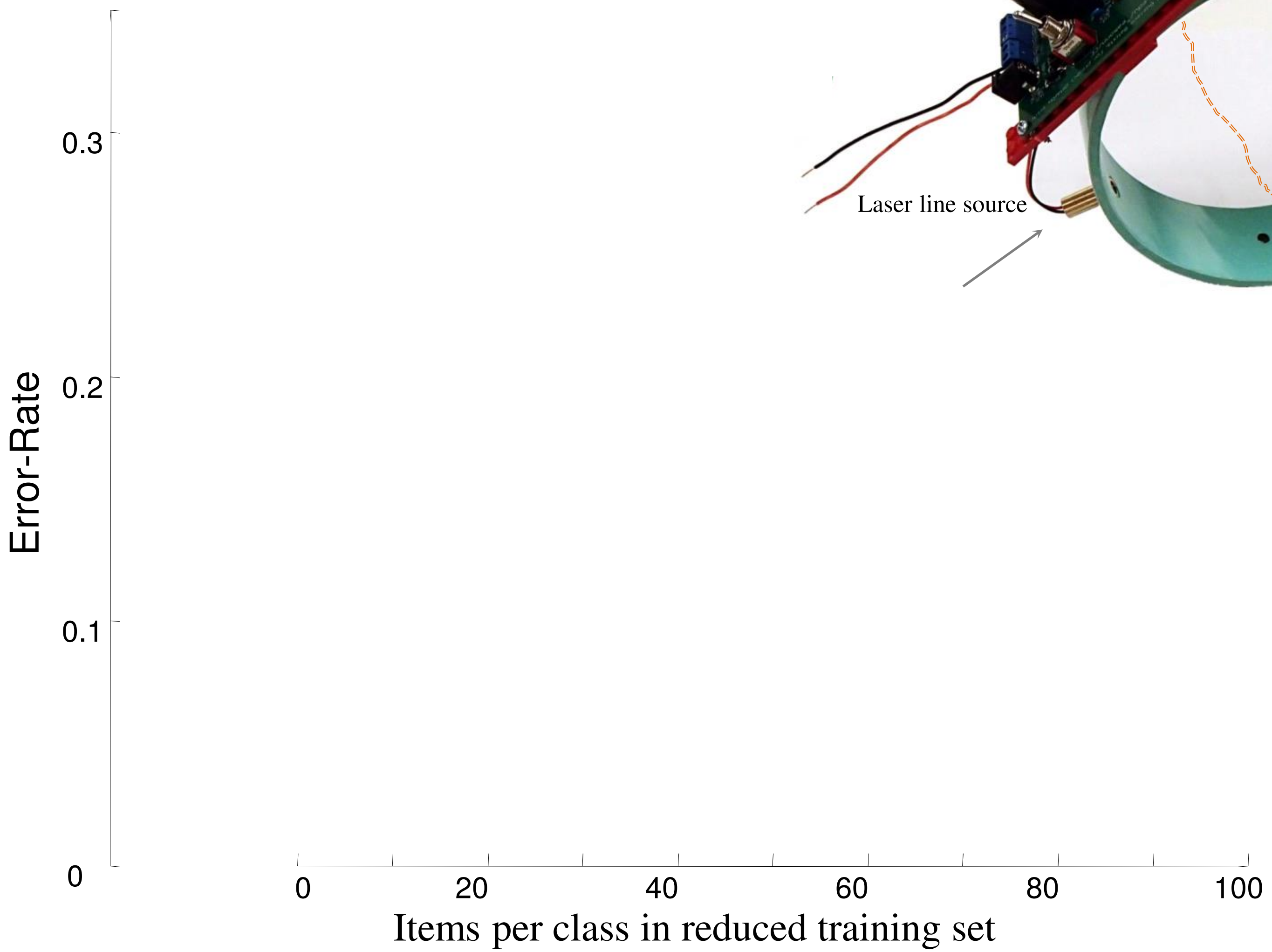
… both using DBA

6 competitors

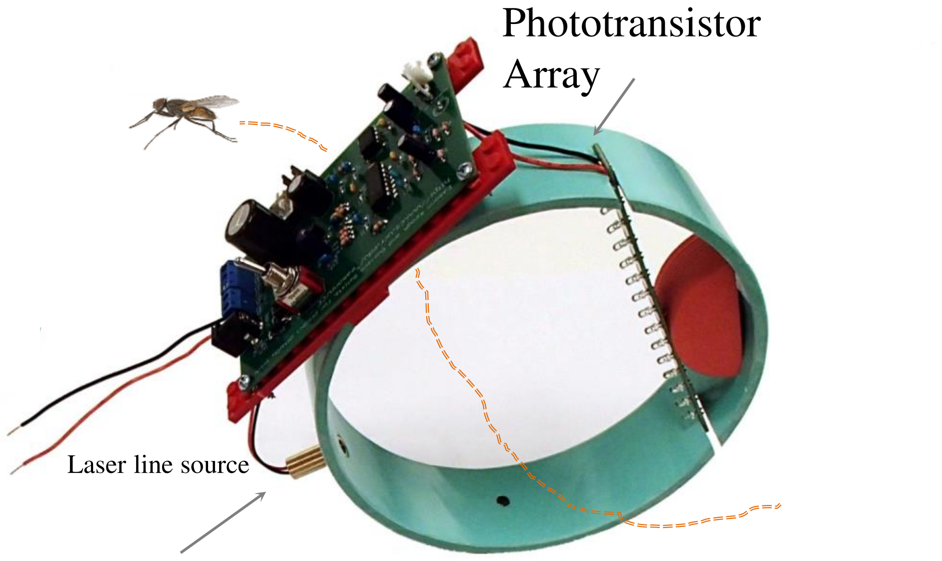
1. Random selection
2. Drop 1
3. Drop 2
4. Drop 3
5. Simple Rank
6. K-medoids

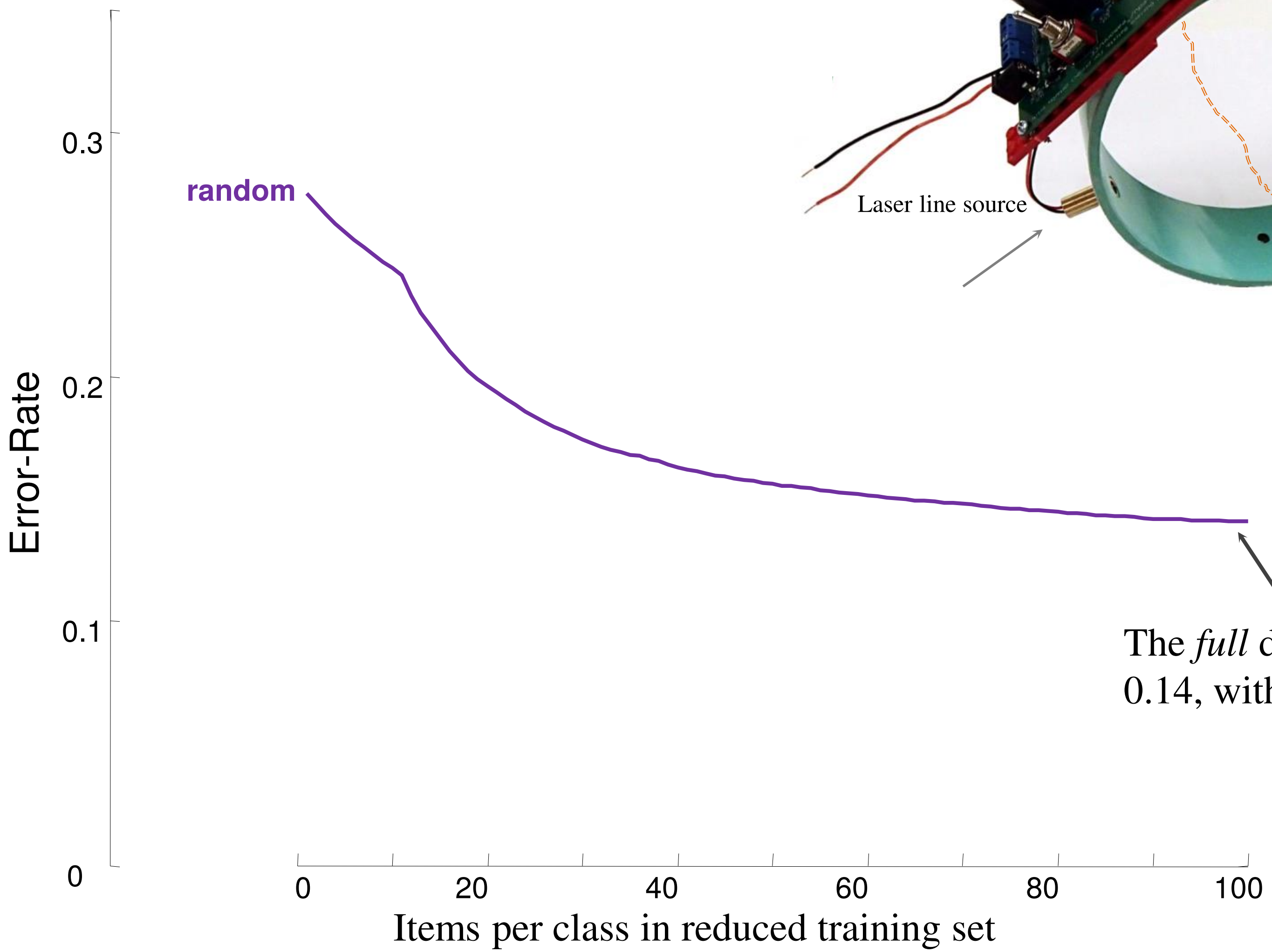
Back to insects





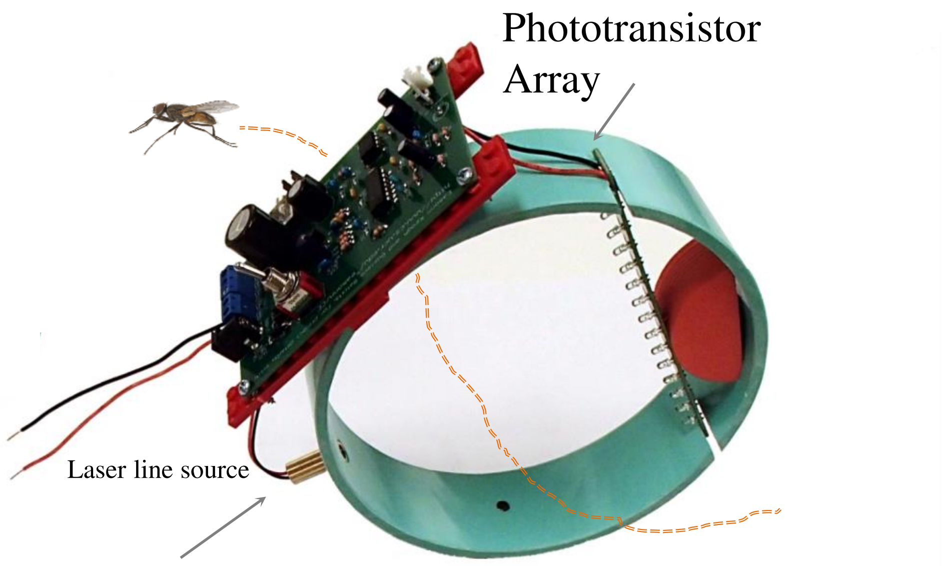
Back to insects

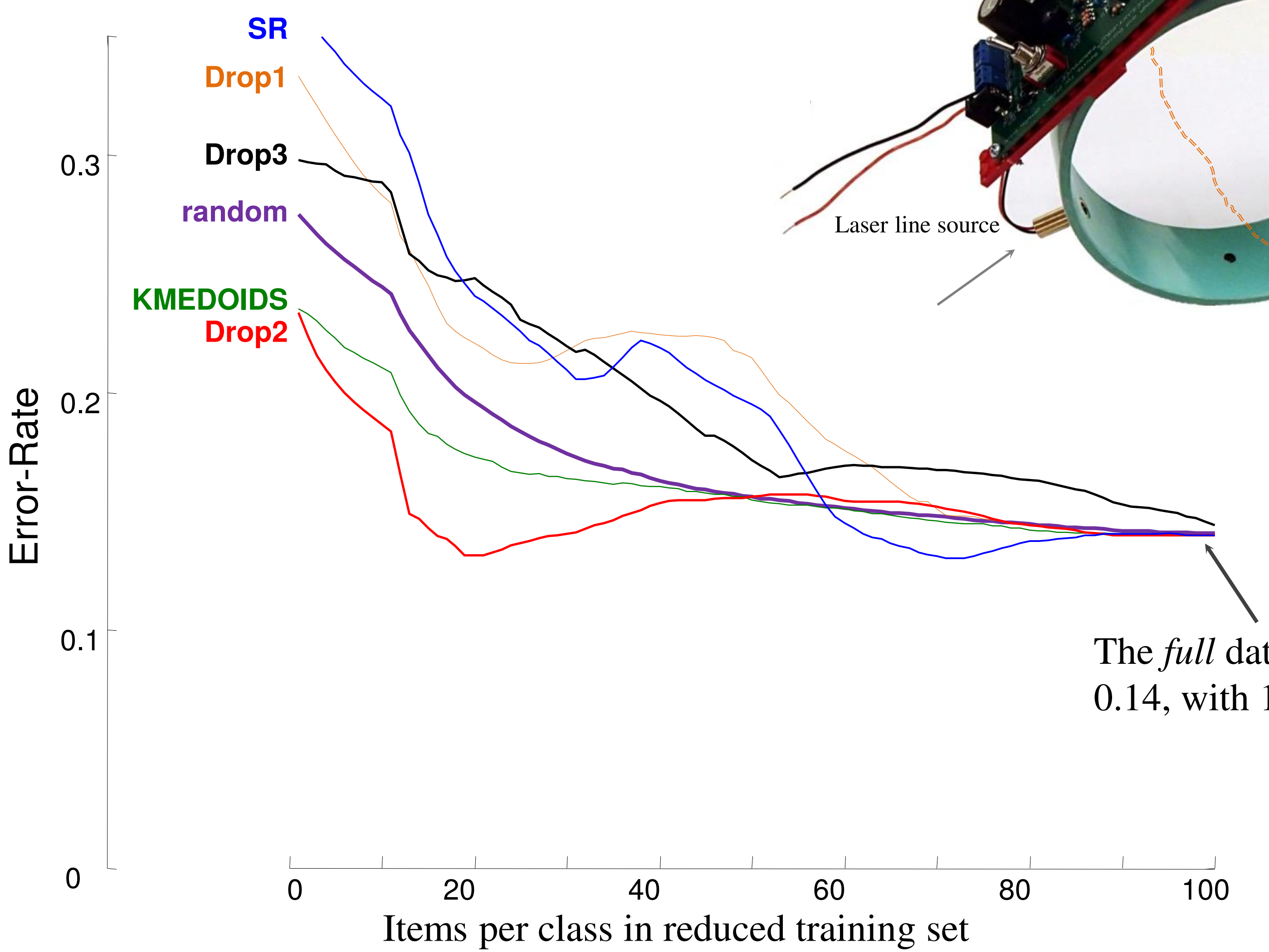




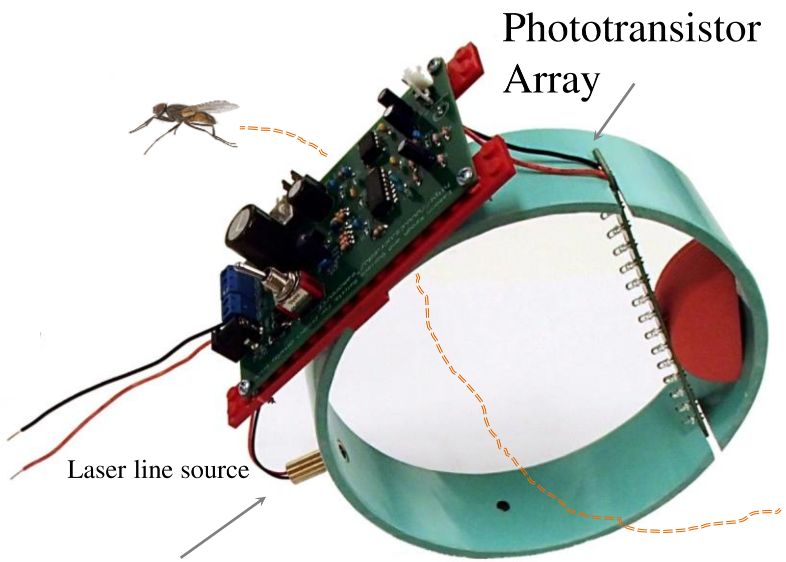
The *full* dataset error-rate is 0.14, with 100 pairs of objects

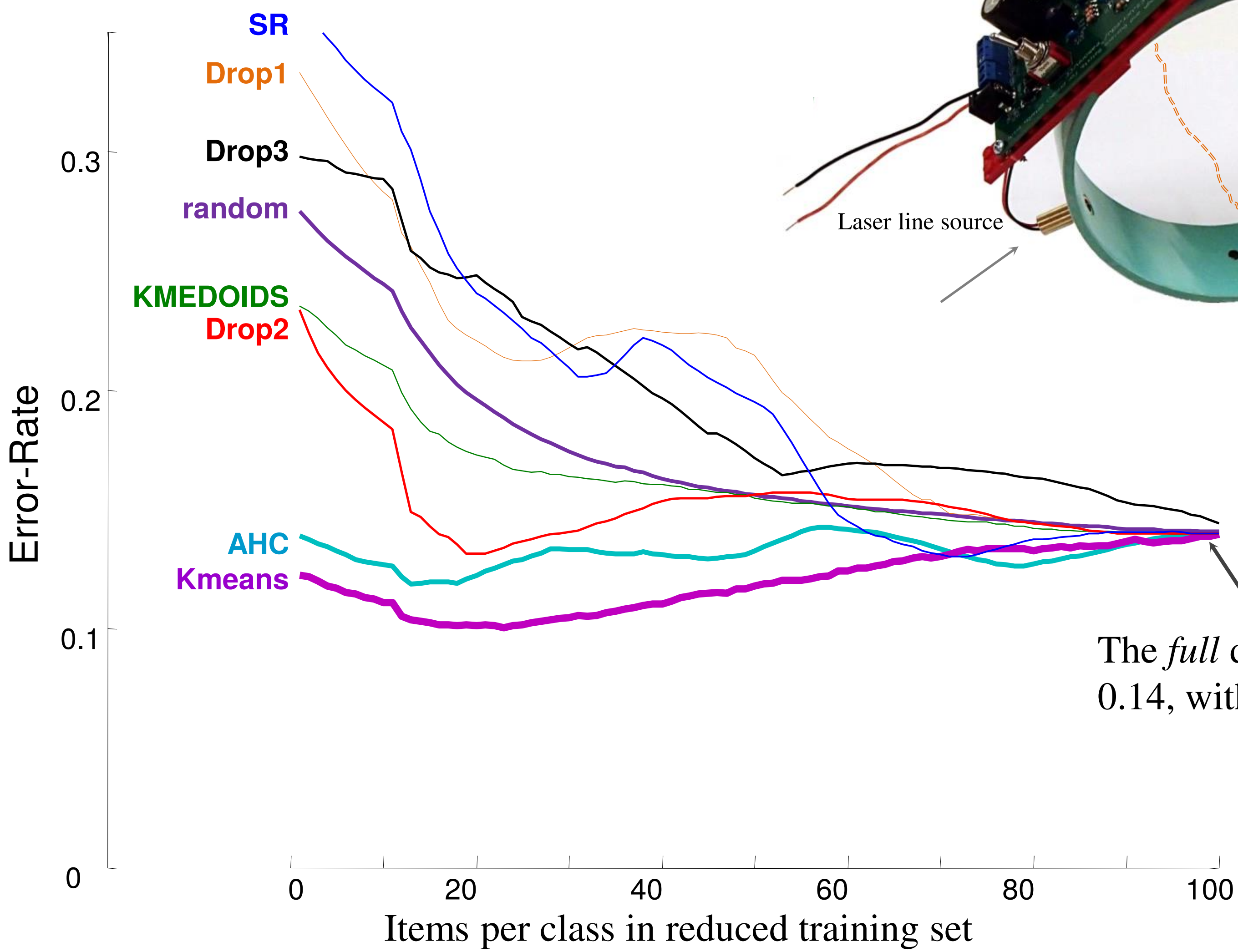
Back to insects





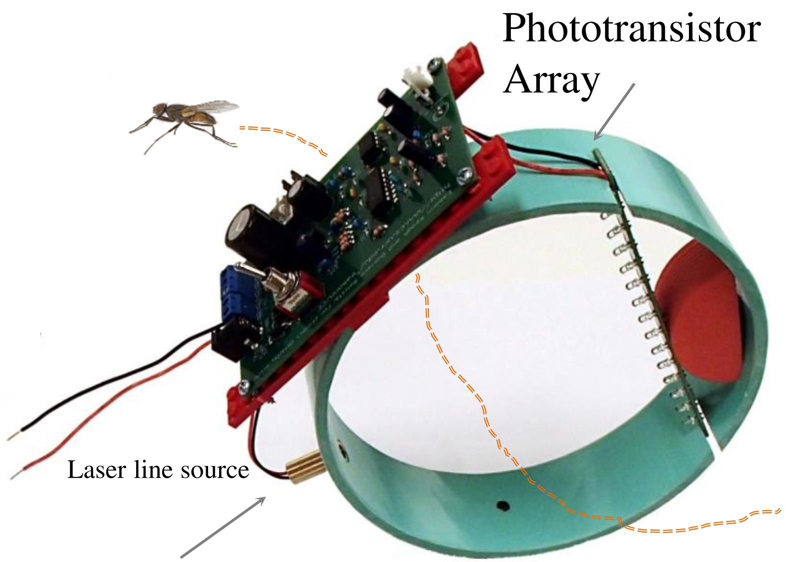
Back to insects

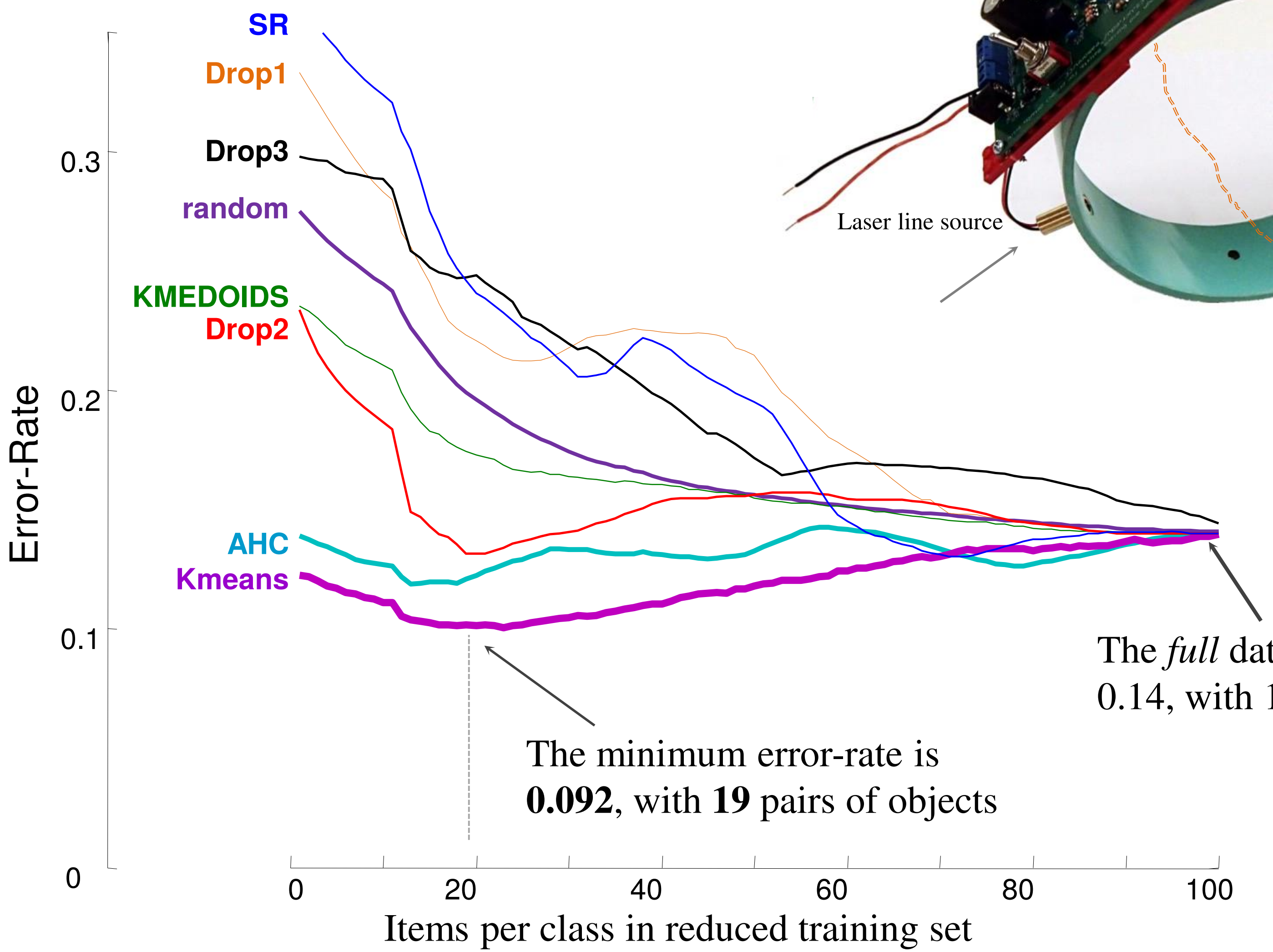




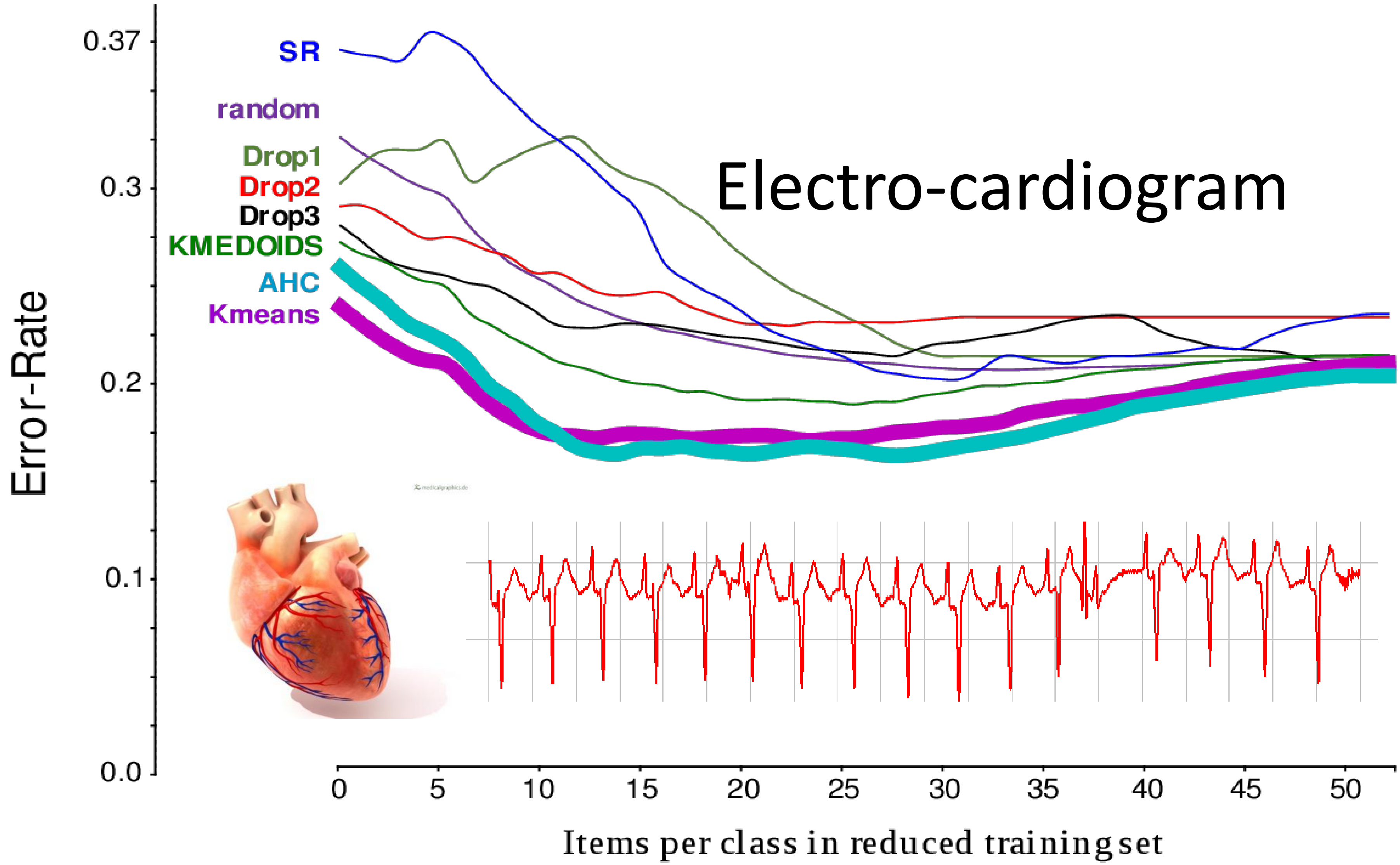
The *full* dataset error-rate is 0.14, with 100 pairs of objects

Back to insects

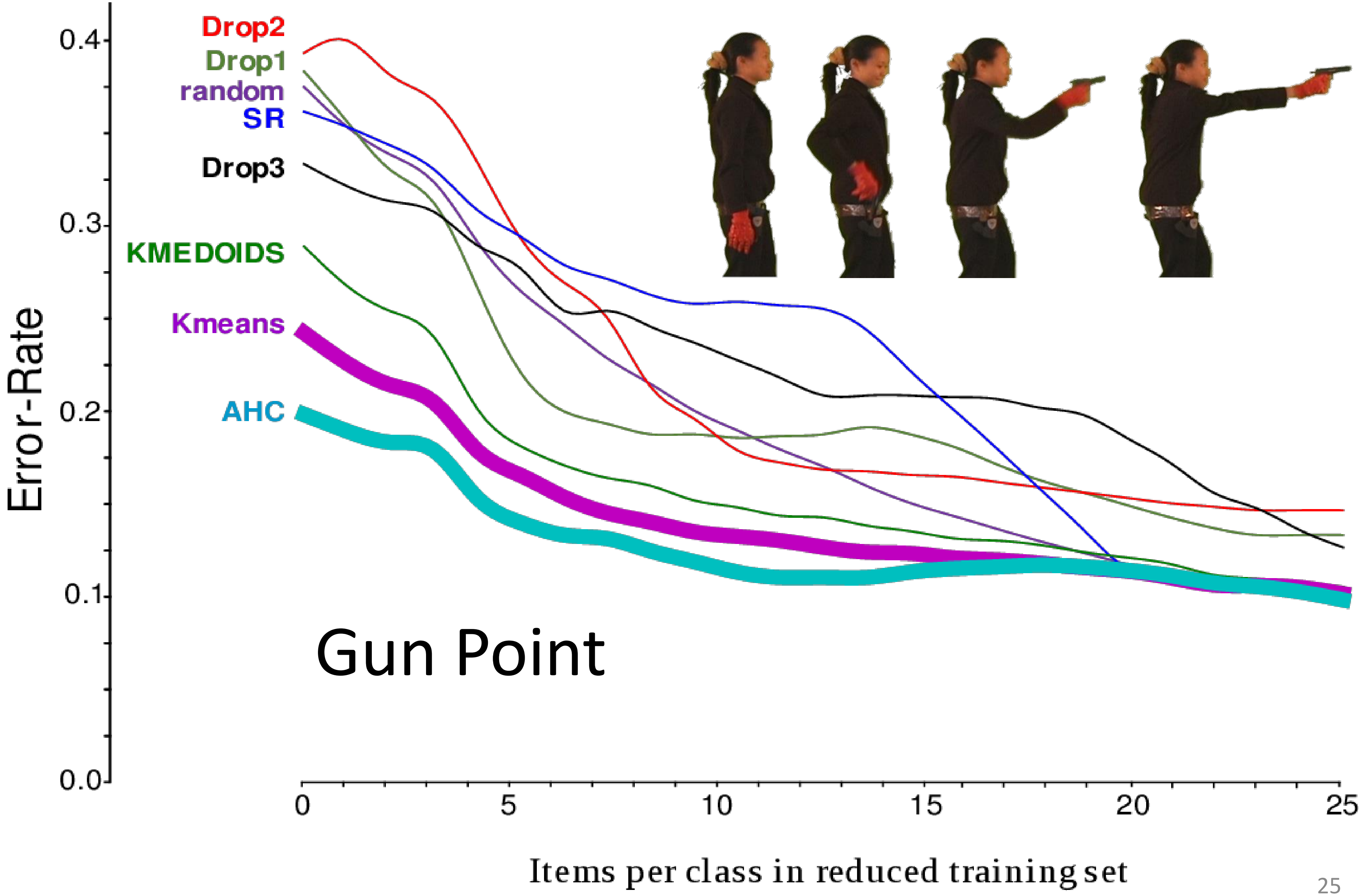




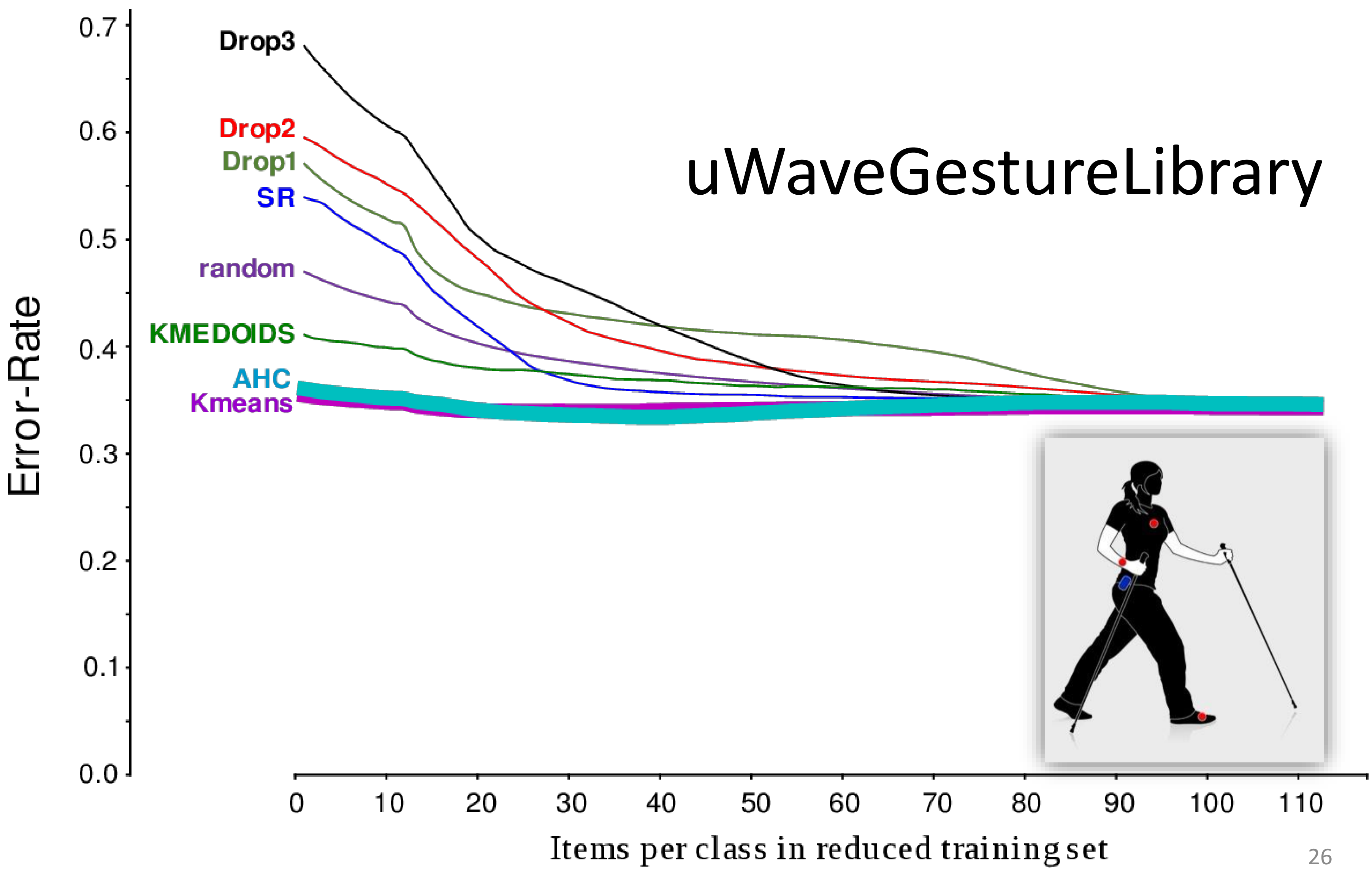
What about other datasets?



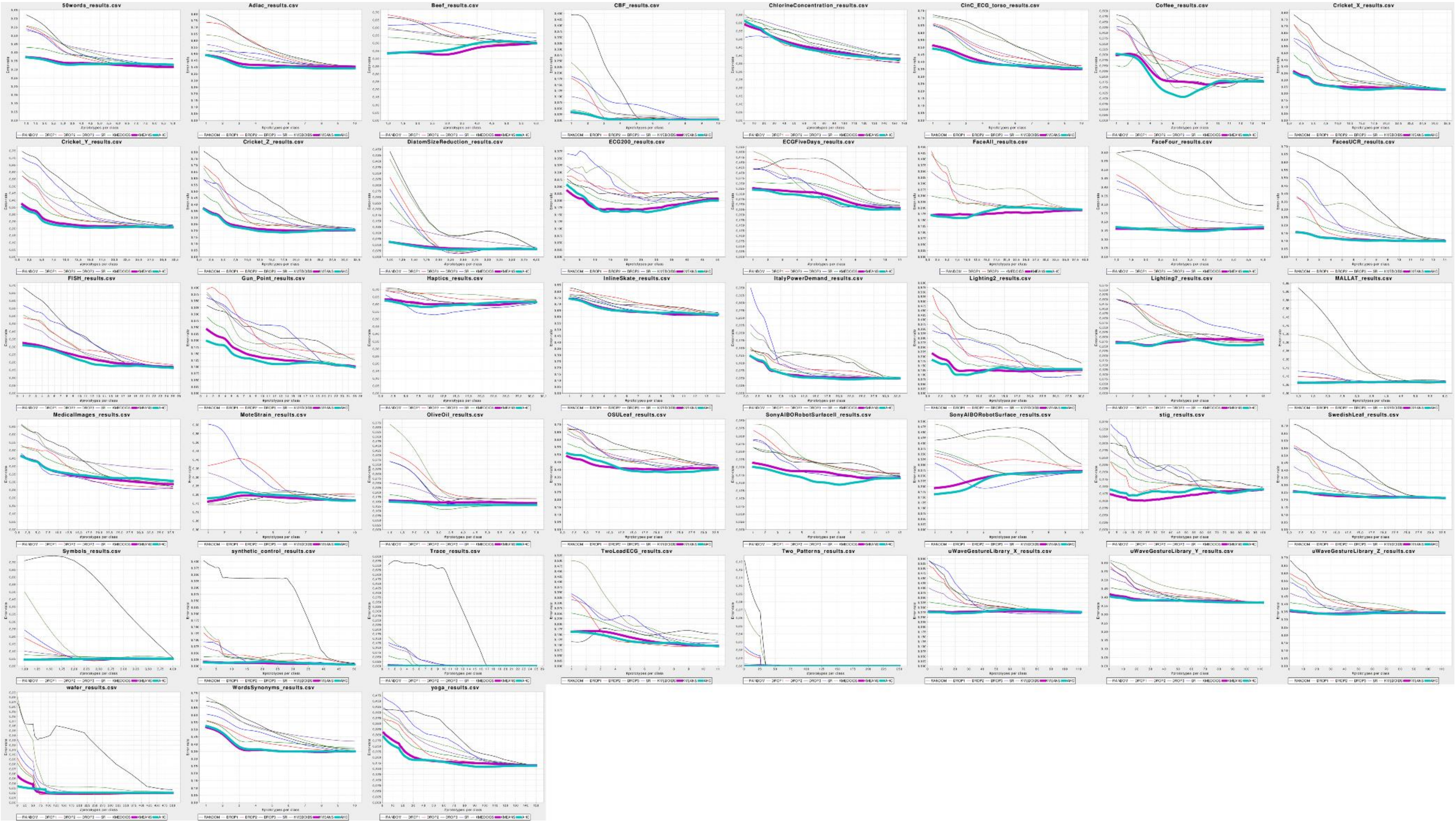
What about other datasets?



What about other datasets?

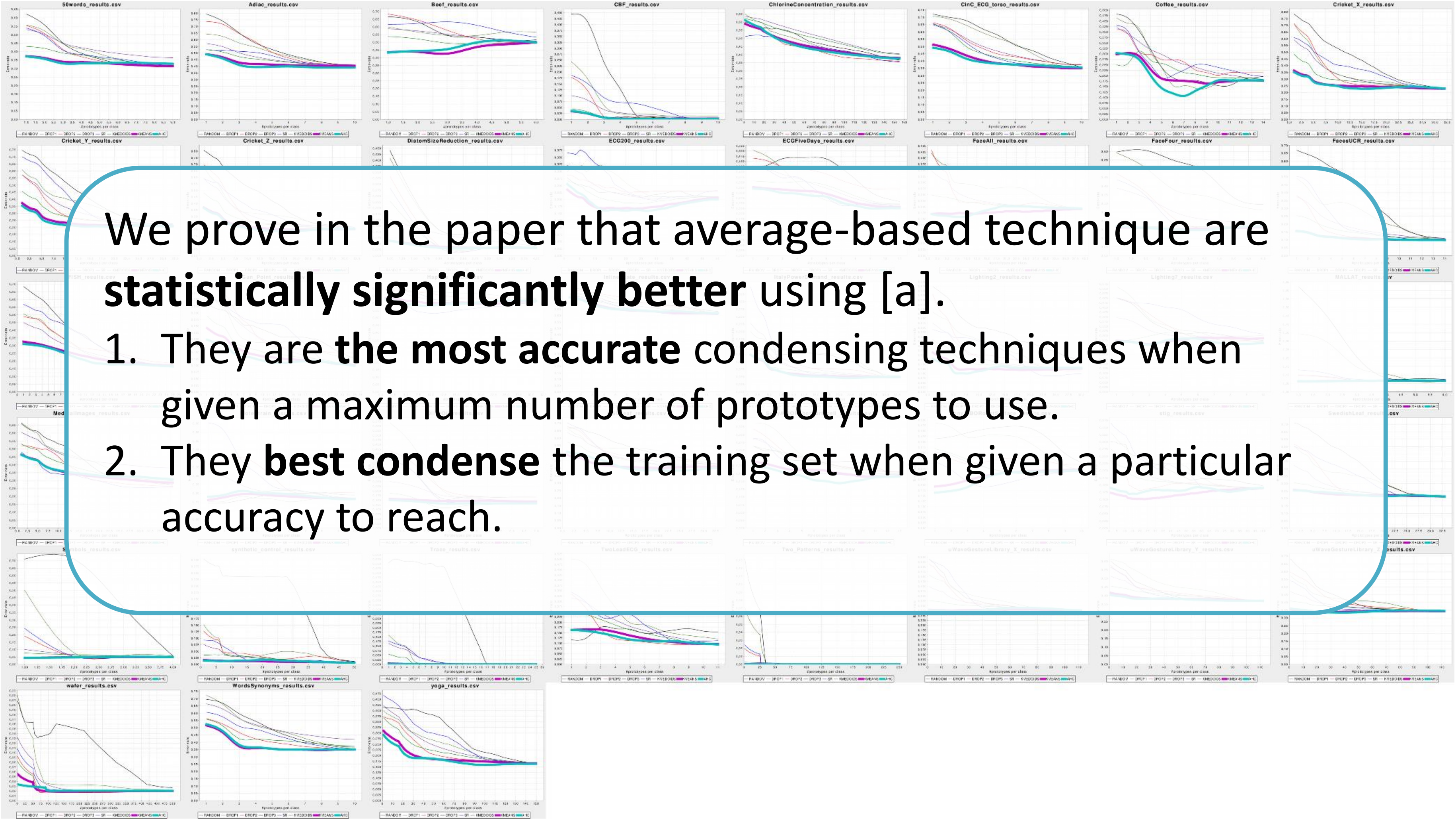


All results on 40+ datasets are online!



http://www.francois-petitjean.com/Research/ICDM2014-DTW

All results on 40+ datasets are online!



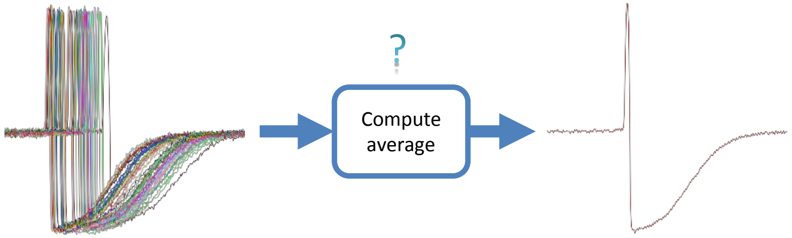
We prove in the paper that average-based technique are **statistically significantly better** using [a].

1. They are **the most accurate** condensing techniques when given a maximum number of prototypes to use.
2. They **best condense** the training set when given a particular accuracy to reach.

http://www.francois-petitjean.com/Research/ICDM2014-DTW

[a] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

Take-home message



Almost everything was in the title!

1. DBA computes the average time series for DTW
2. Averaging can make time series classification:
    1. Faster
    2. More accurate
3. We believe in reproducible research:
    1. We tested our approach on 40+ datasets from the UCR archive
    2. We computed the statistical significance of the results
    3. The source code is online

Web: http://www.francois-petitjean.com/Research/ICDM2014-DTW
E-mail: francois.petitjean@monash.edu
Twitter: @LeDataMiner

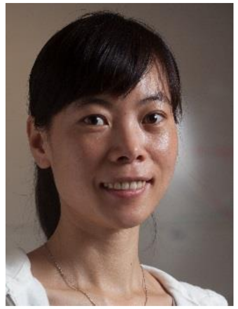
Thanks! Please come and have a chat!


F. Petitjean


G. Forestier


G.I. Webb


A.E. Nicholson


Y. Chen


E. Keogh

Support and funding