SAND2022-4241C

SIAM Conference on Uncertainty Quantification (UQ22)

Sandia National Laboratories

Quantifying Uncertainty in Machine Learning Models for Time Series Classification

April 13, 2022

Ahmad Rushdi*, Erin Acquesta†, Gabriel Huerta†, Kyle Neal†, India Dytzel†, Bill Rider†

* Stanford University. Work was done while at Sandia National Laboratories

† Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

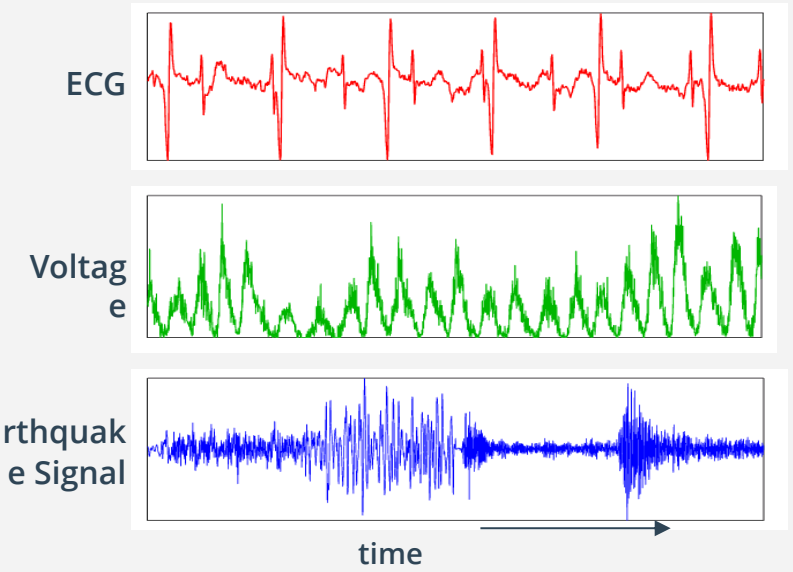
owned subsidiary of H
subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.A0003525.

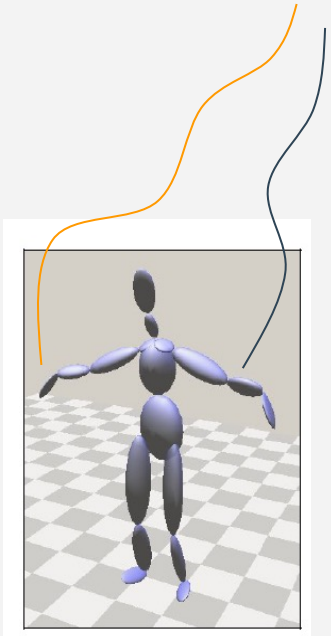
Agenda

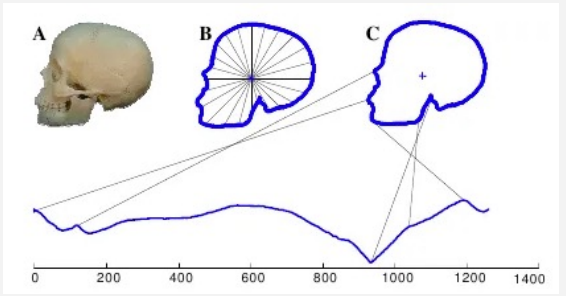
1. Problem Statement
2. Discriminative Features
3. Benchmark: Distance to *k* Nearest Neighbors
4. Convolutional NNs for Time Series Classification
5. Recurrent NNs for Time Series Classification
6. Conclusions

What is a Time Series?

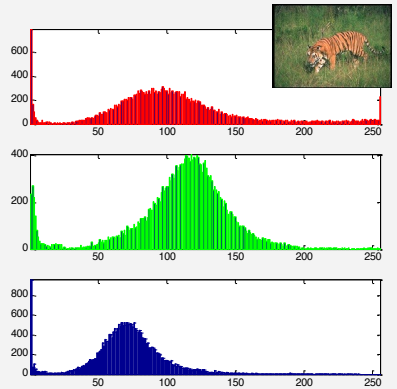
Observations $X = [x_1 \quad x_2, \ldots, \quad x_n] \in \mathbb{R}^n$: an ordered set of generally real valued measurements, with a natural temporal ordering. Given a labeled dataset $D := \{(\mathbf{x}_i, y_i)\}_{i=1}^N \subset \mathcal{X} \times \mathcal{Y}$, TSC aims to train a classifier model C, by learning from D in order to classify a query TS or generally map the space of possible inputs to a probability distribution over possible output classes.







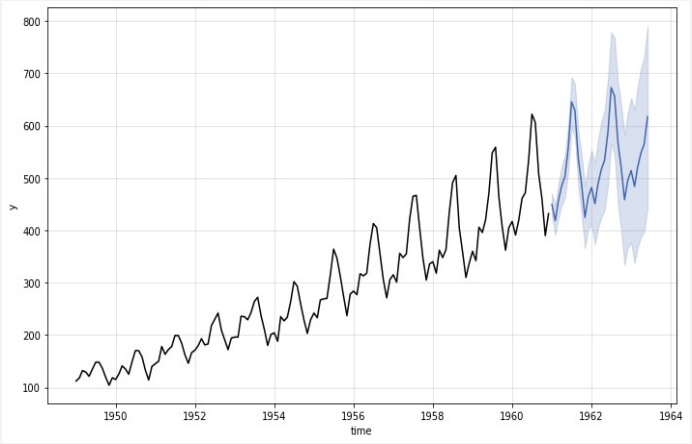
Keogh, Eamonn, et al. "LB_Keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures." 2006.

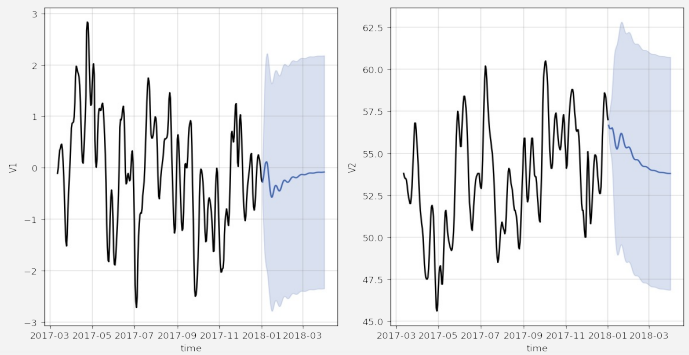


Time Series Classification

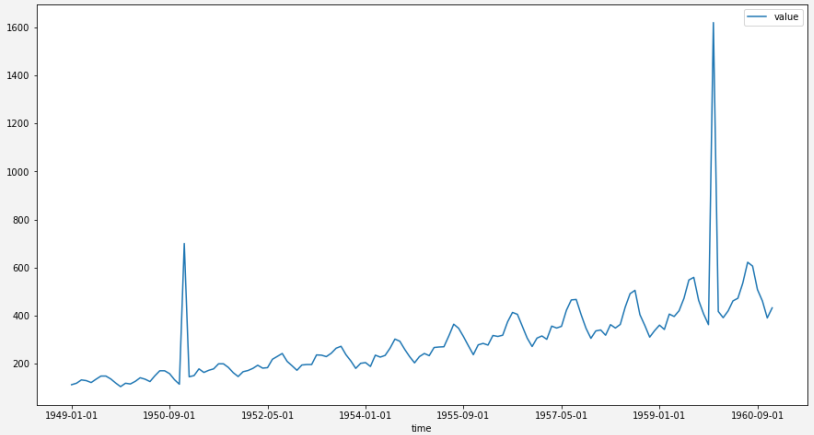
Uncertainty Quantification in Time Series Analysis: essential for many scientific and engineering applications.

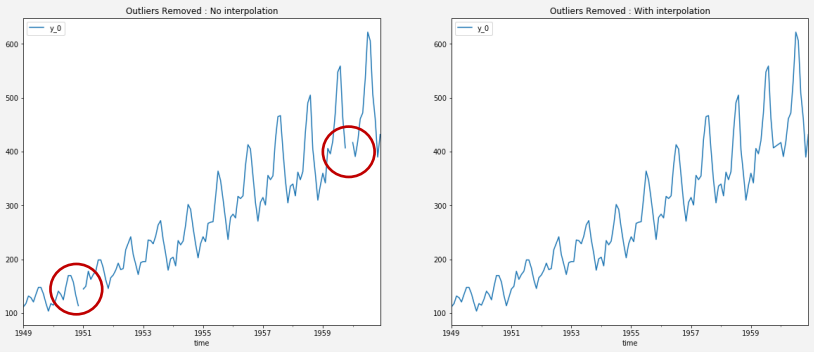
Forecasting, extrapolation



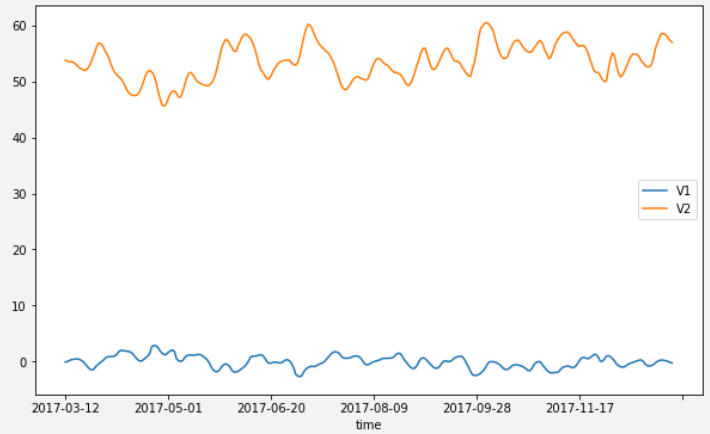


Outlier detection, interpolation





TS Classification

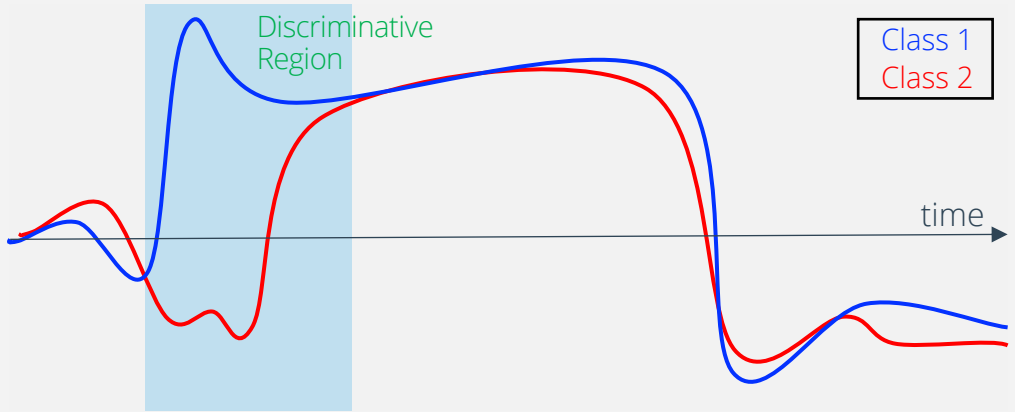


Time Series Classification

Problems

1. Uncertainty
   - Measurements are often noisy, low SNR, scaled, include missing/outlier points, not well aligned, not of the same length, or way too long to maintain long term memory
2. Feature Engineering
   - Discriminatory features are not always obvious, require SME expertise for labeling.
   - TS -> image (e.g., Gramian fields, recurrence plots, Markov transition fields).
3. Data availability:
   - Only a few samples/class
   - Classes are mislabeled or imbalanced.

In contrast to feature engineering, **end-to-end deep learning** aims to incorporate the feature learning process while fine-tuning the discriminative classifier.



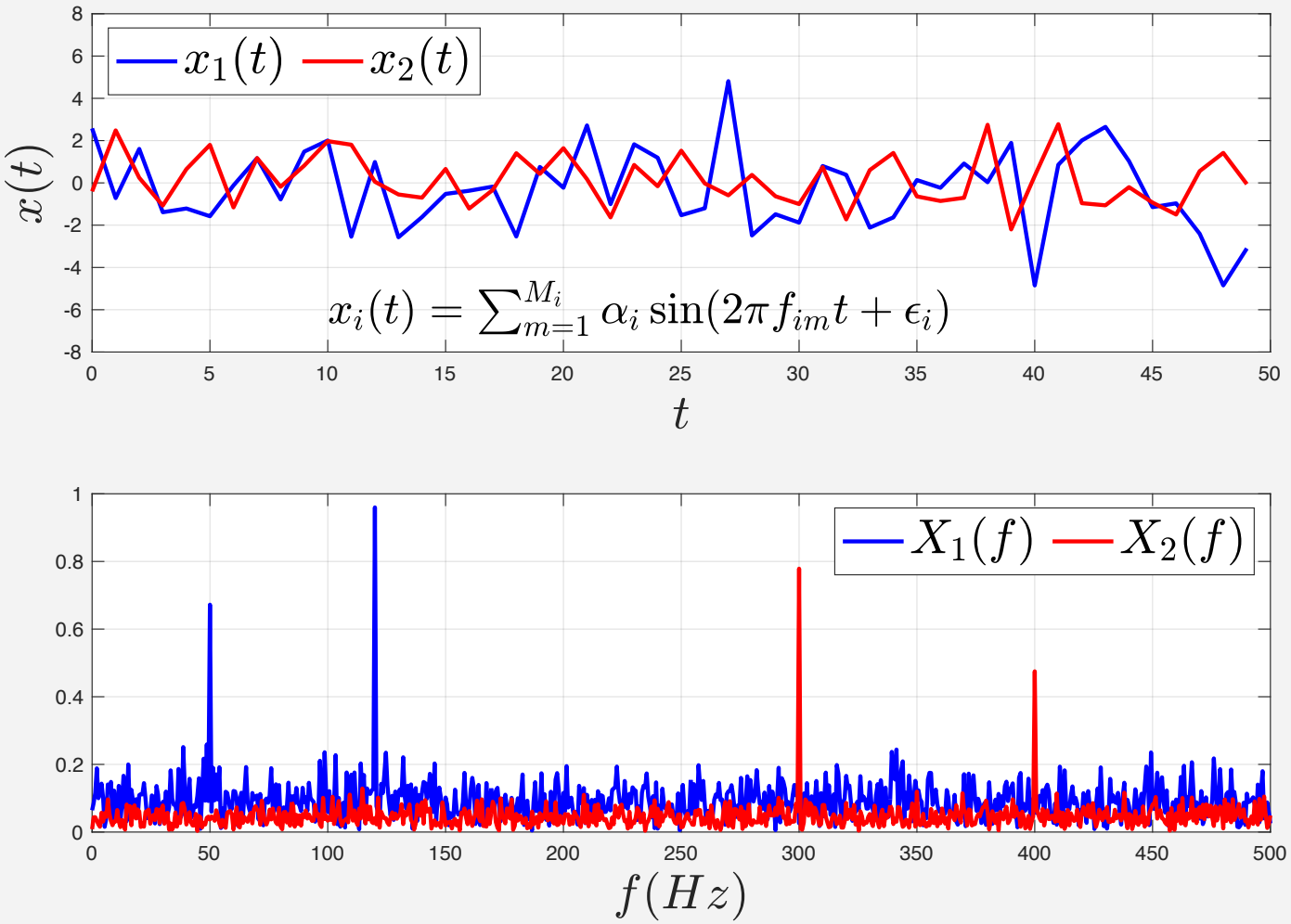
Domain Representation for Time Series Classification

Hard to identify discriminative features in a lengthy time series.

Move to another feature space?

- Fourier transform, DFT or STFT
- Cosine transform (DCT)
- Shapelets transform
- Manually-selected features

Global-local tradeoffs.



Time Series Classification Methods

Distance-based	Interval-based	Ensembles	Deep Learning
• *k* Nearest Neighbors with some distance measure • Support Vector Machines	• Short time FT • Wavelets • Time Series Forest (TSF) • Shapelet Transform	• Collective Of Transformation-based Ensembles (COTE, 35) • Hierarchical Vote Collective of Transformation-Based Ensembles (HIVE-COTE) • Bag of SFA Symbols (BOSS)	• Convolutional Neural Networks • Recurrent Neural Networks

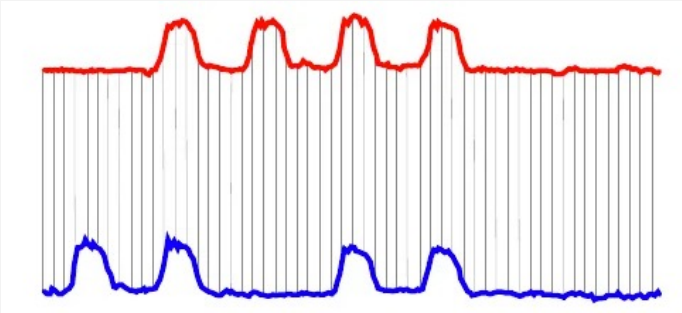
- An ensemble of nearest neighbor classifiers with different distance measures outperforms in accuracy all of the ensemble's individual members.
  - **Scalability:** what about the **computational cost**?
  - e.g., Shapelet (a member of HIVE-COTE): $O(n^2 l^4)$

Bagnall, Anthony, et al. "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances." *Data mining and knowledge discovery* 31.3 (2017): 606-660.

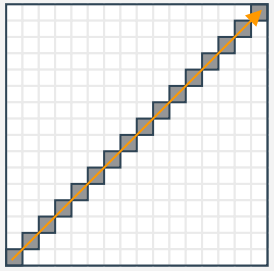
Distance Metrics: Euclidean vs. Dynamic Time Warping

Euclidean Distance

$$d(X_1, X_2) = \left(\sum_{i=1}^{n} |x_{1,i} - x_{2,i}|^p\right)^{\frac{1}{p}}$$



$O(n)$



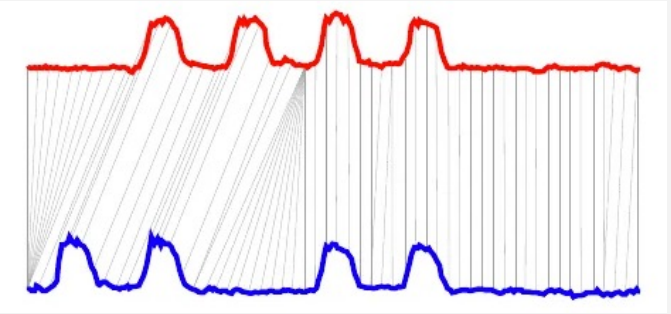
The Euclidean distance assumes 1-1 mapping, which implies the same length. Can we allow for sequence compression & decompression in time?

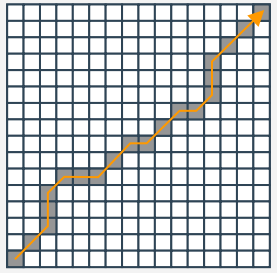
Dynamic Time Warping (DTW) Distance

$$d(X_1, X_2)$$

$$= \arg \min_{W=w_1,\ldots,w_K} \sqrt{\sum_{k=1,w_k=(i,j)}^{K} (x_{1,i} - x_{2,j})^2}$$

With standard dynamic programming, complexity of DTW: $O(n^2)$



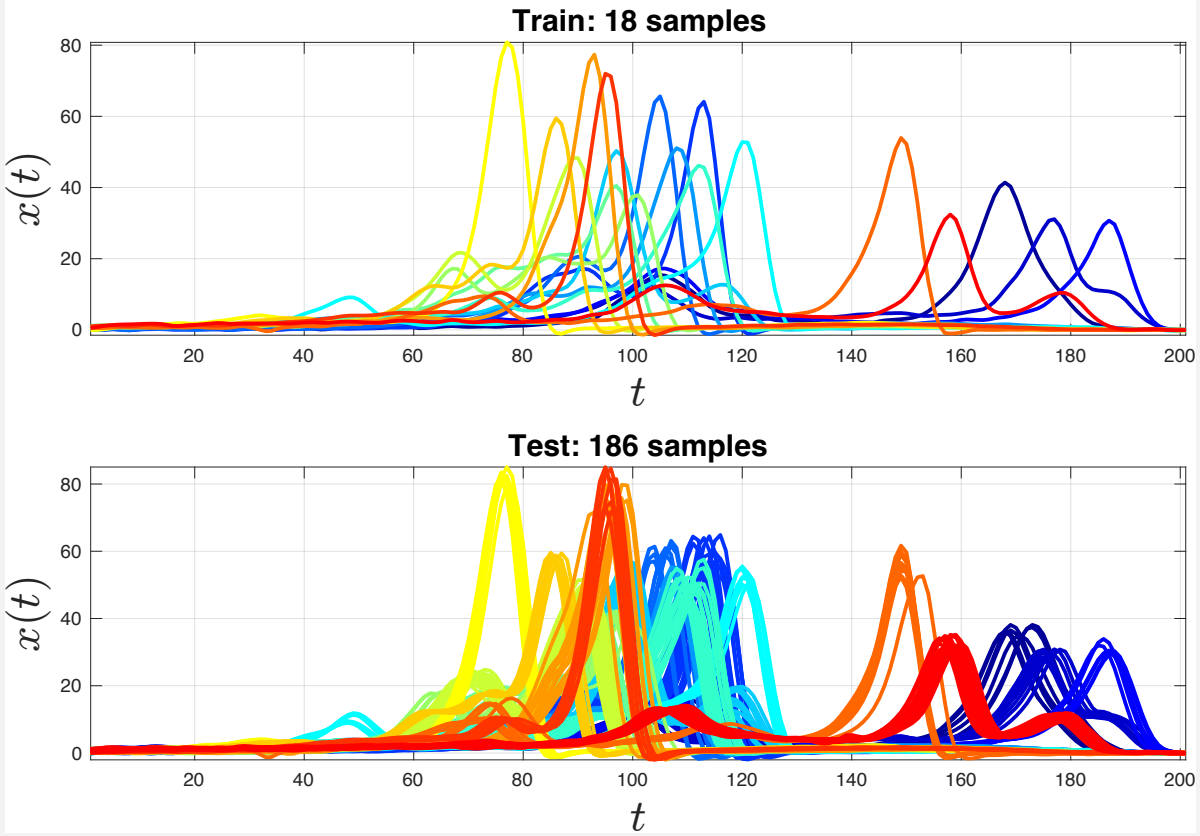


Accuracy-efficiency tradeoff: We want our warping path to stay relatively close to a diagonal line.

- DTW with w= 100% (no warping window) => unconstrained DTW.
- DTW with w= 0% degenerates to the ED distance.

Illustrations: 1NN-ED & 1NN-DTW

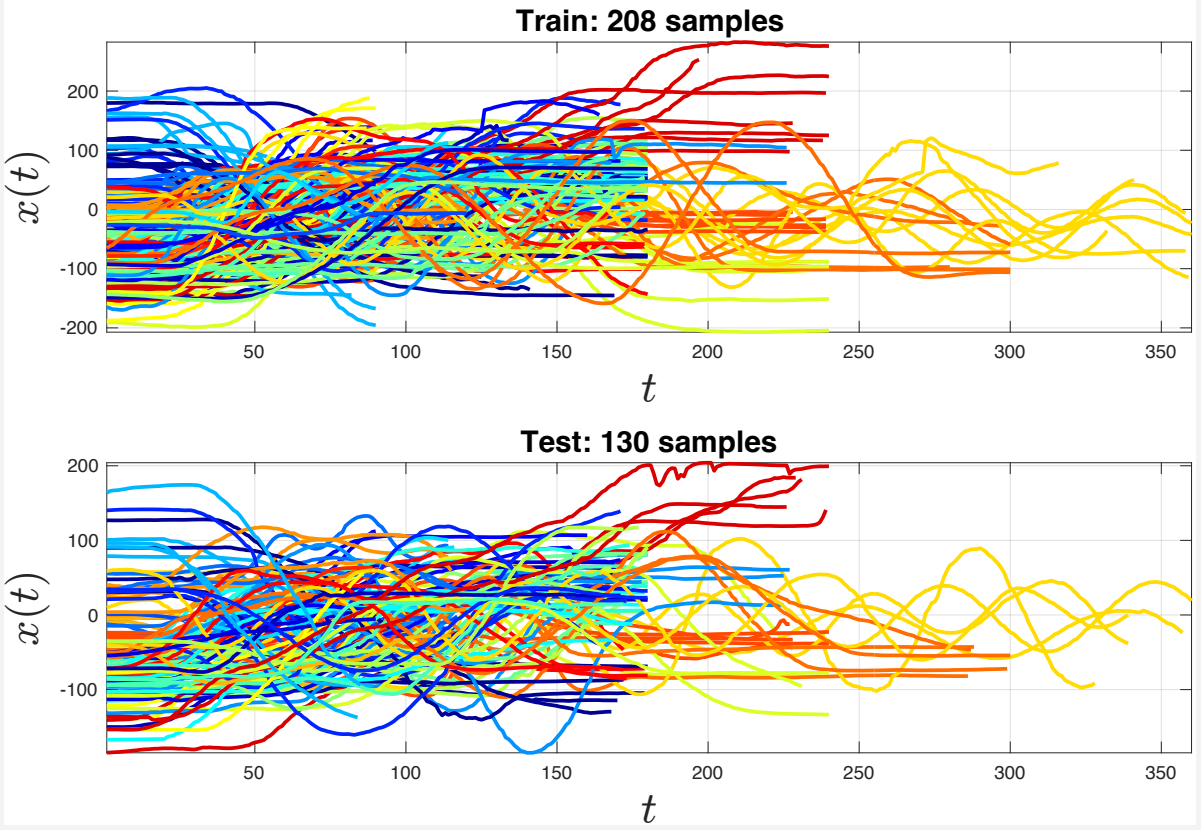
Fungi dataset: 18 classes



1 TS/class for training < 10% test samples

1NN-ED (w=0), error rate = 0.1774
1NN-DTW (learn w), error rate = 0.1774 (0)
1NN-DTW (w=100), error rate = 0.1613

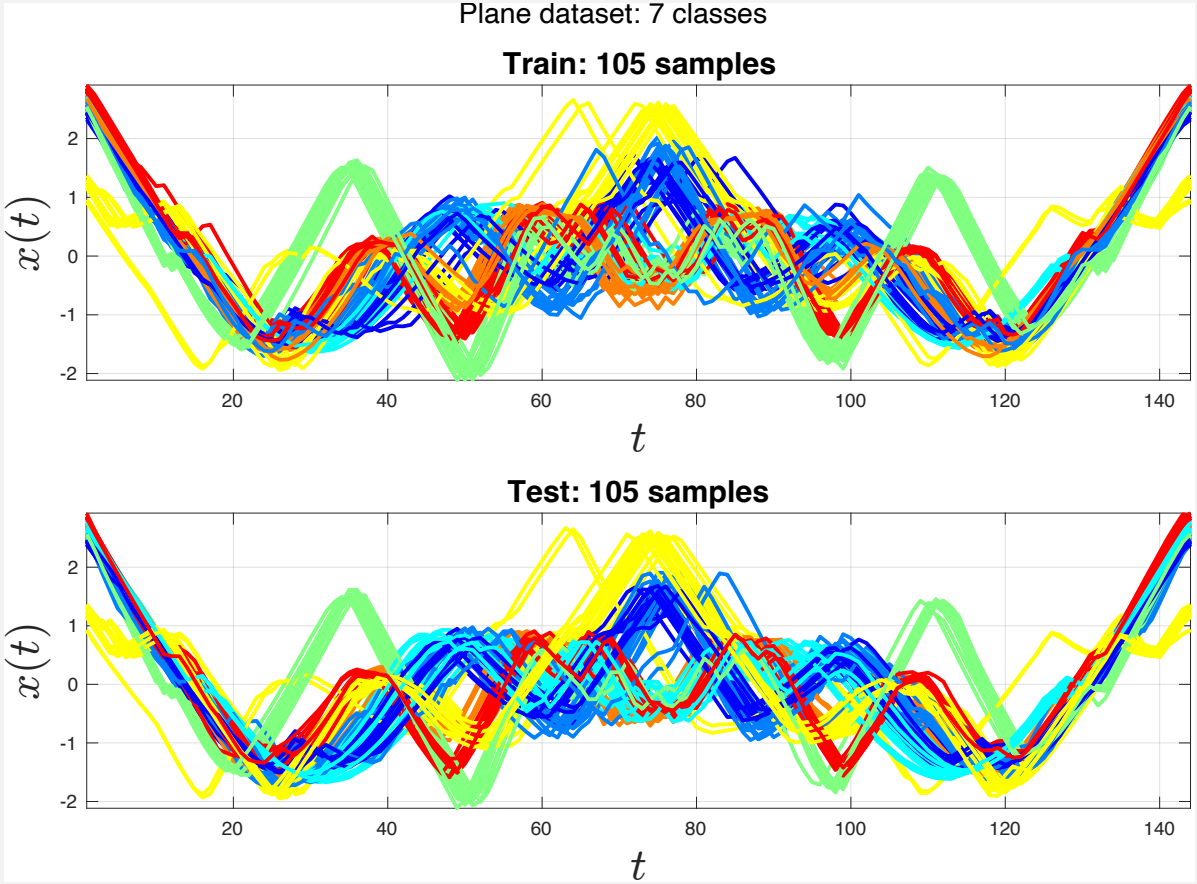
GestureMidAirD1 dataset: 26 classes



Variable-length TS

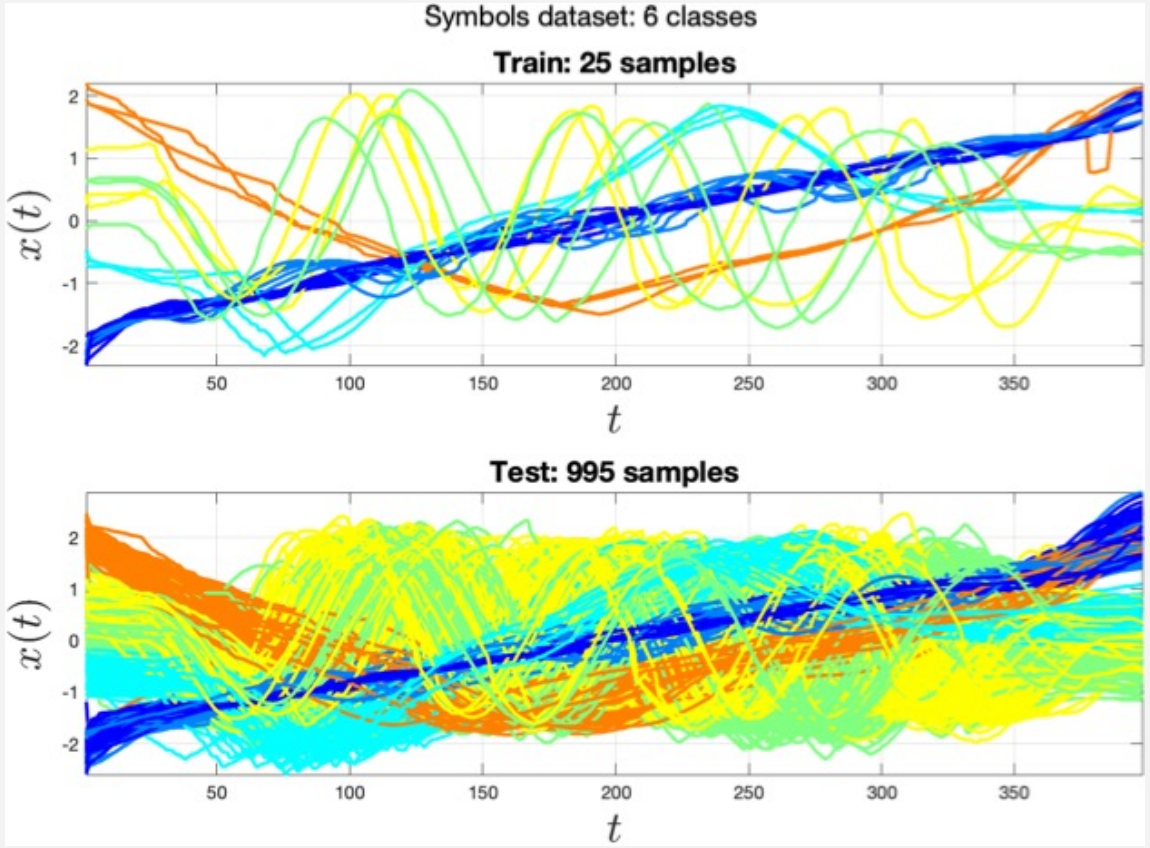
1NN-ED (w=0), error rate = 0.4231
1NN-DTW (learn w), error rate = 0.3615 (5)
1NN-DTW (w=100), error rate = 0.4308

Illustrations: 1NN-ED & 1NN-DTW



Discriminative regions for all 7 classes

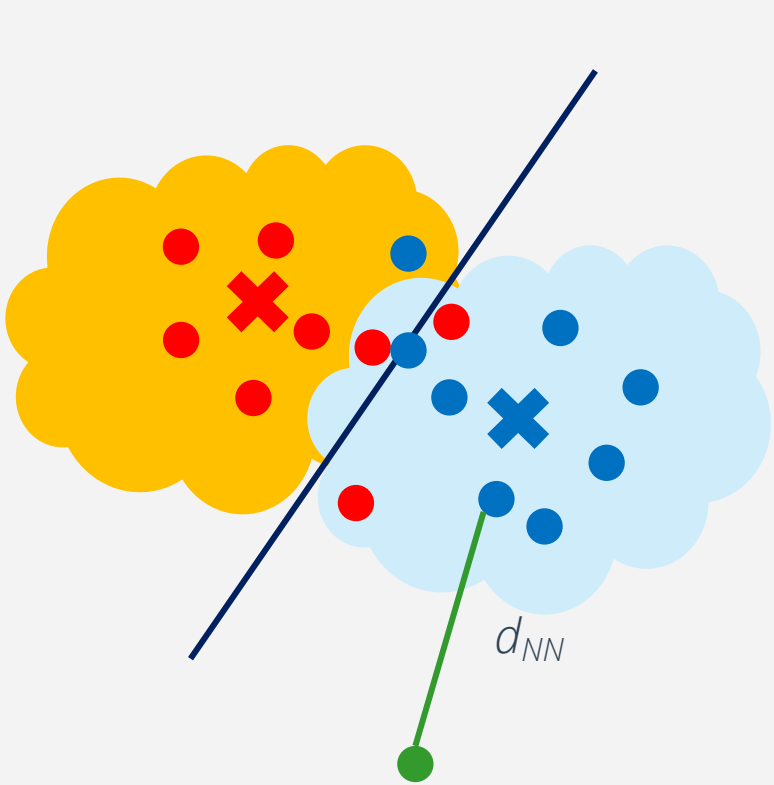
1NN-ED (w=0), error rate = 0.0381
1NN-DTW (learn w), error rate = 0.0000 (5)
1NN-DTW (w=100), error rate = 0.0000



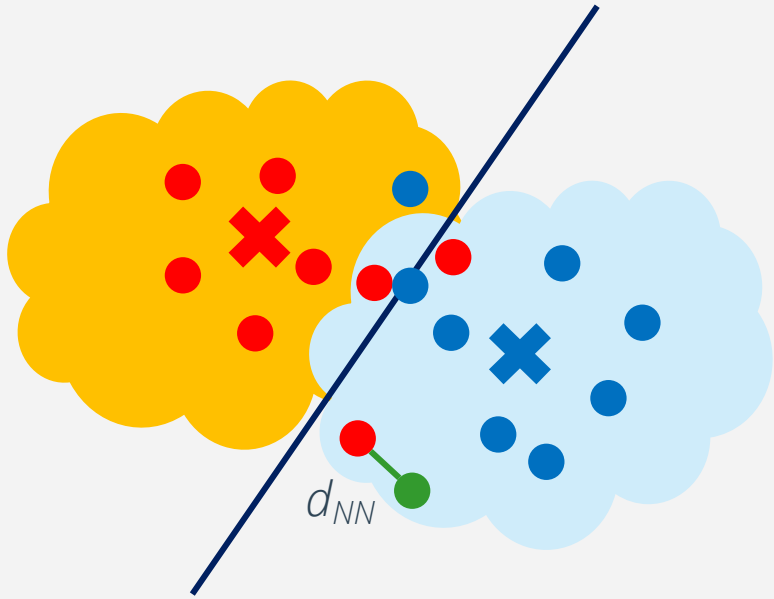
Good candidate for frequency domain

1NN-ED (w=0), error rate = 0.1005
1NN-DTW (learn w), error rate = 0.0623 (8)
1NN-DTW (w=100), error rate = 0.0503

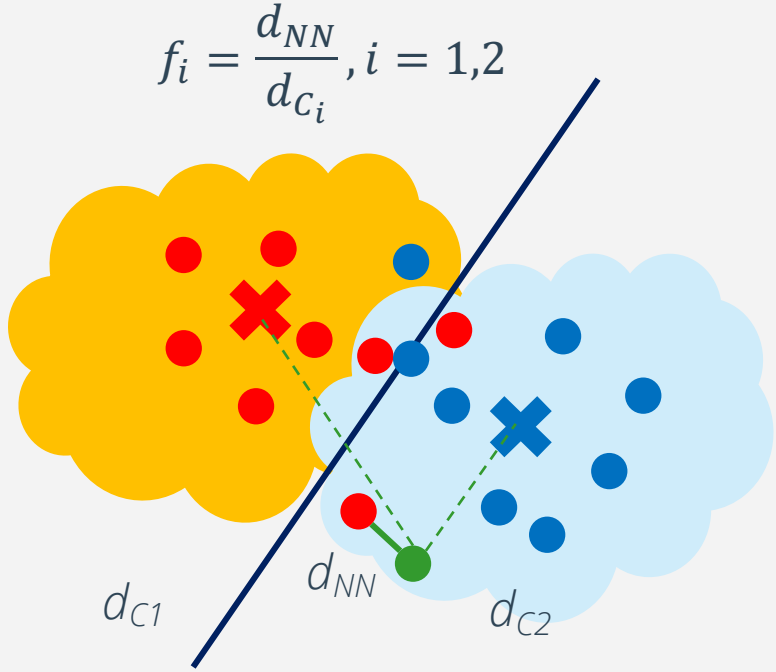
Relative Distance as a Confidence Metric



Correct prediction
despite long d_{NN}



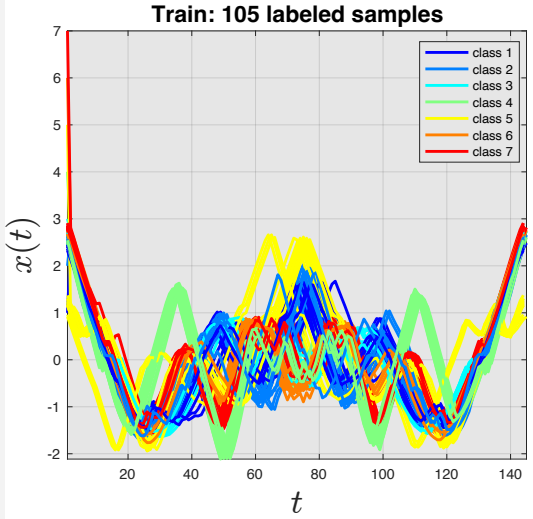
Incorrect prediction
despite short d_{NN}

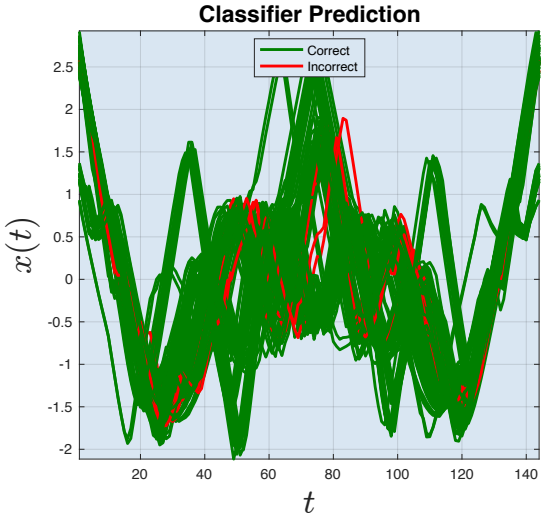


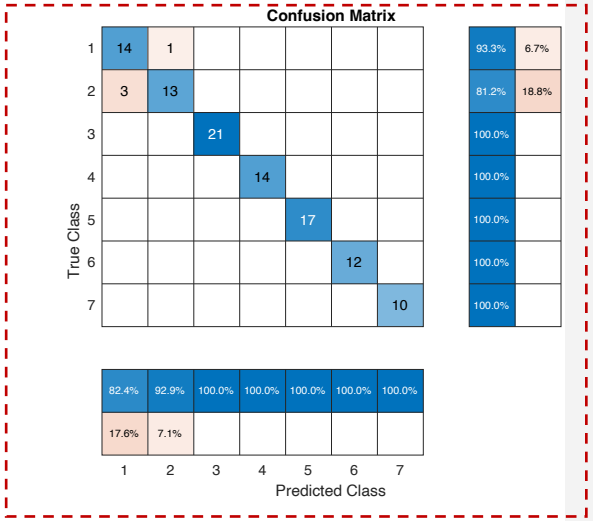
Incorrect prediction, but
low d_{NN}/d_{C1} confidence

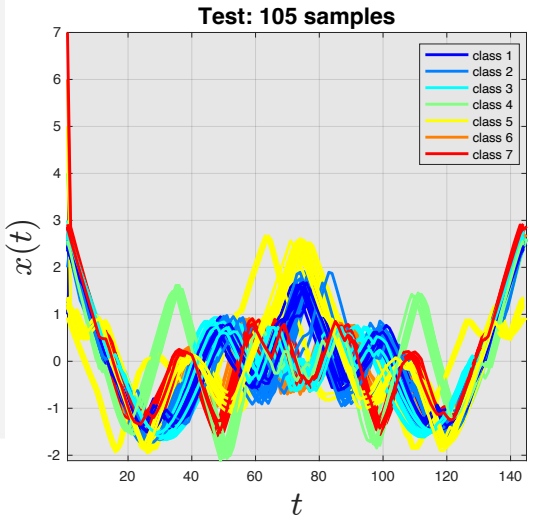
Illustration: Visualizing Classifier Confidence

Plane dataset: 7 classes









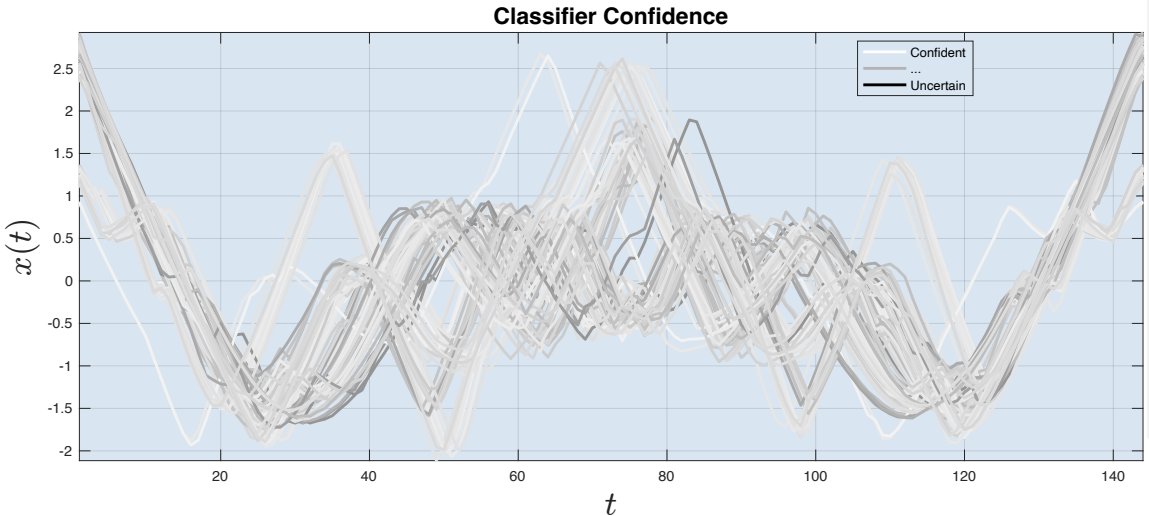
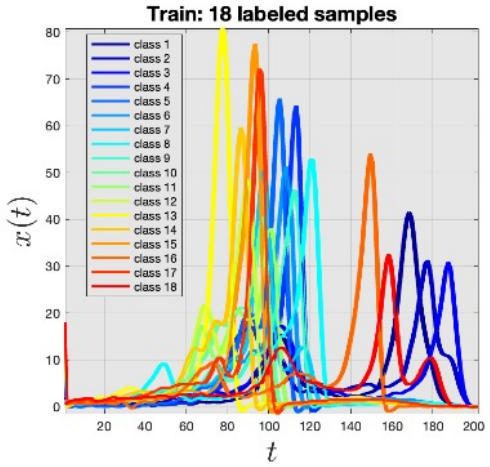
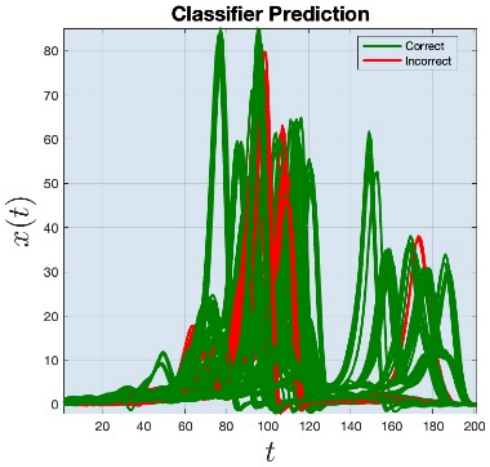
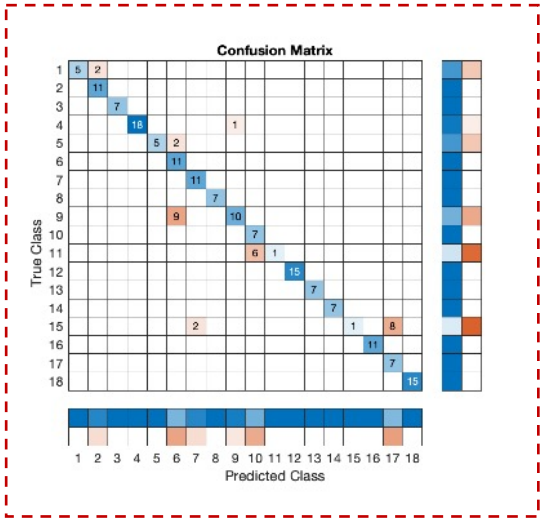


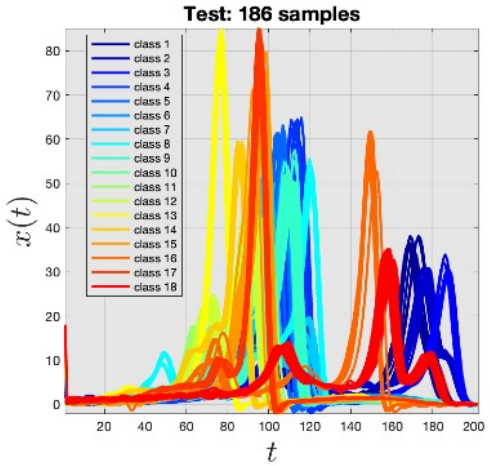
Illustration: Visualizing Classifier Confidence

Fungi dataset: 18 classes









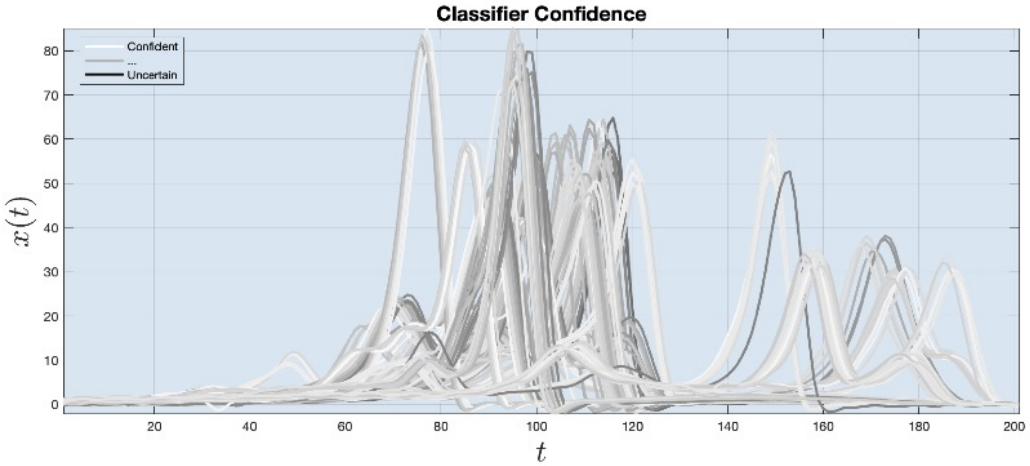


Illustration: Visualizing Classifier Confidence

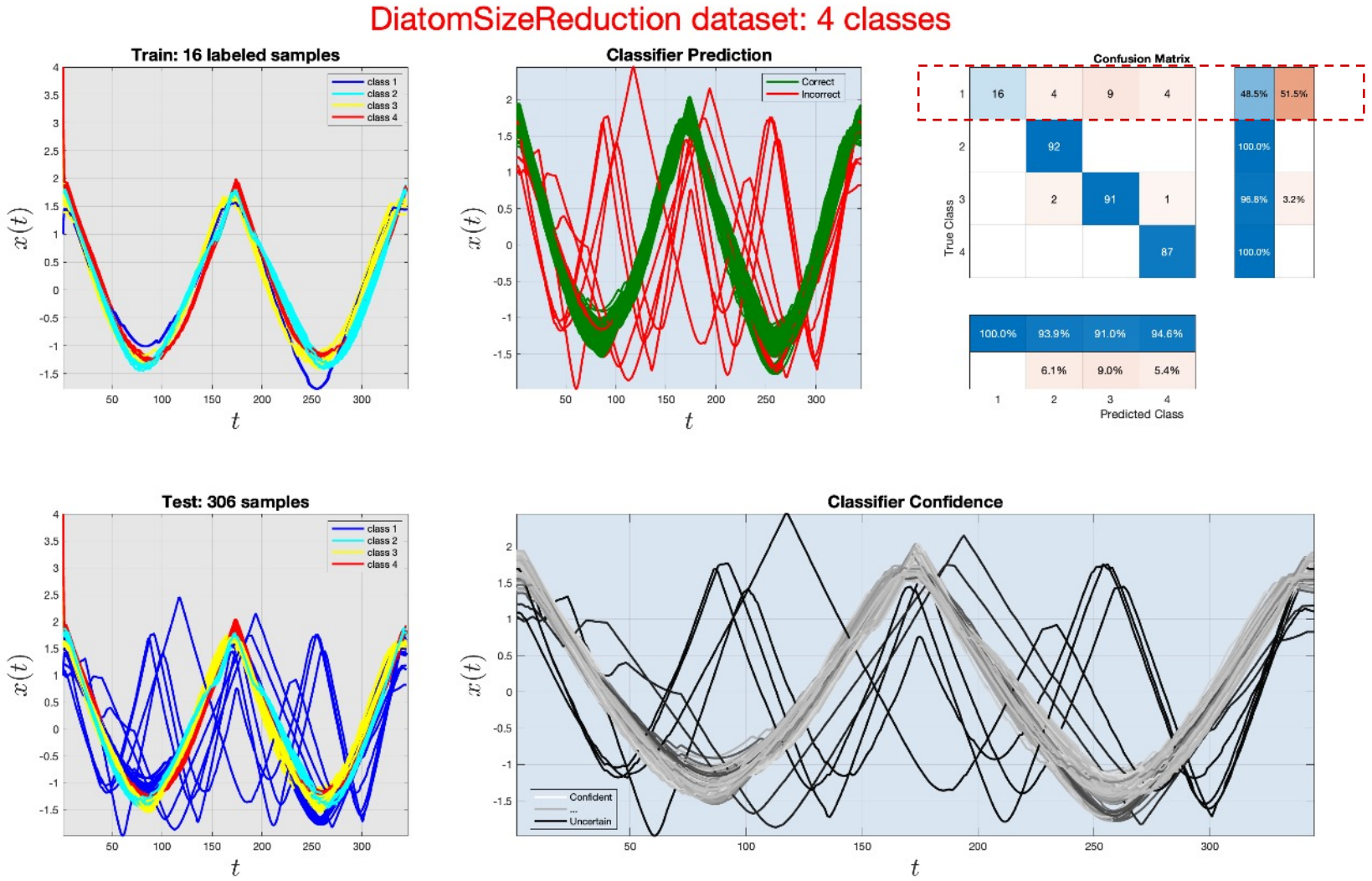
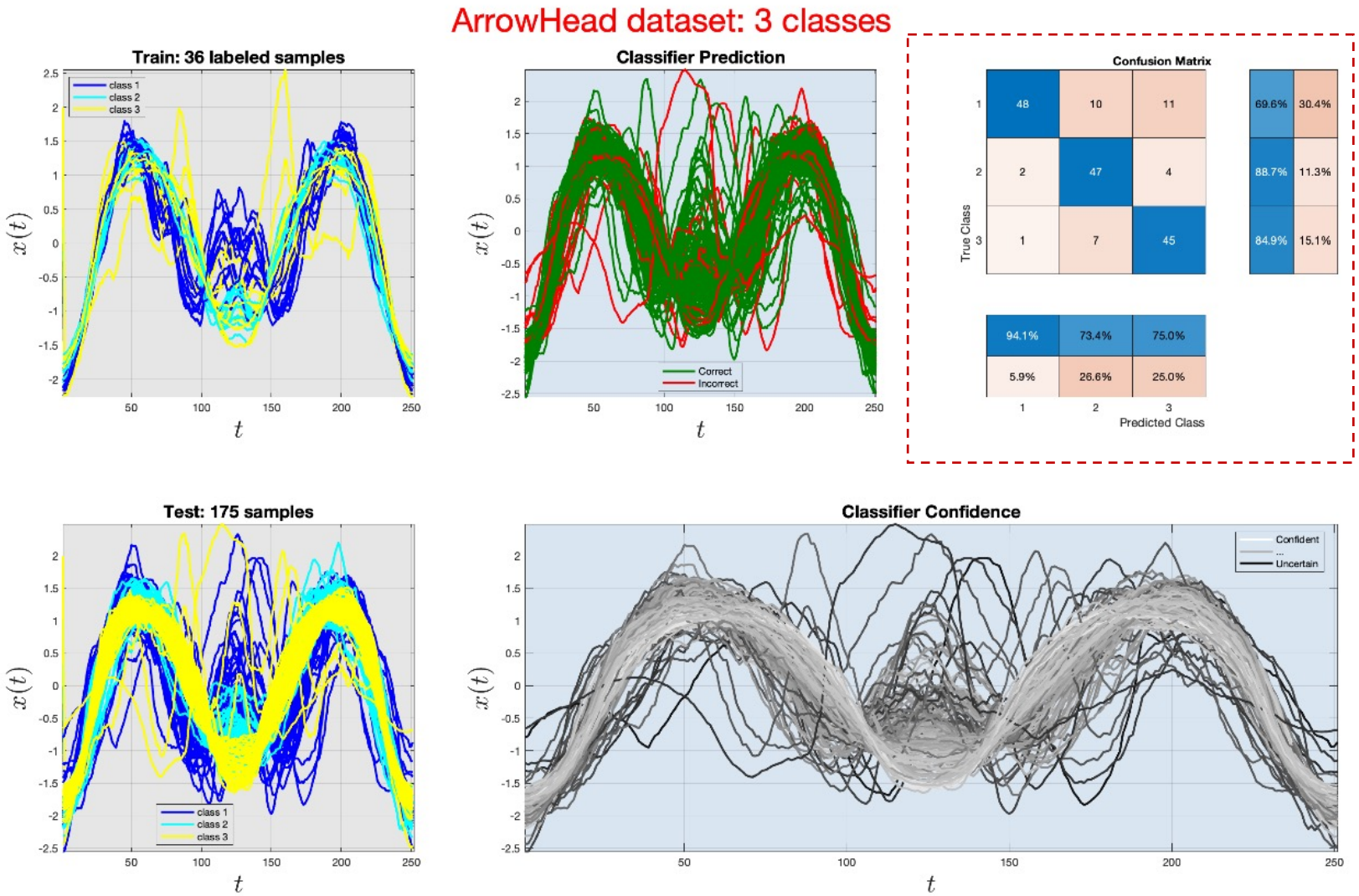


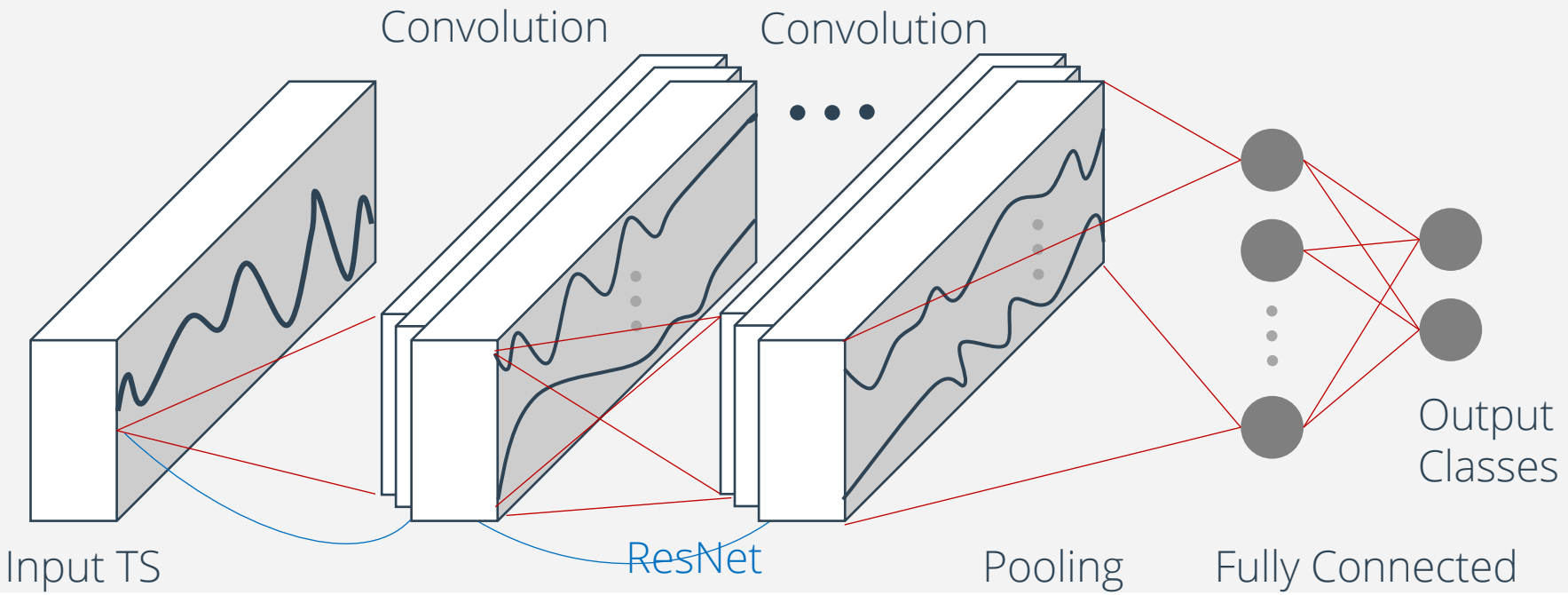
Illustration: Visualizing Classifier Confidence



CNNs for TSC

Our knowledge and intuition for CNNs on images carry over to time-series.

- A CNN model: $f_L(\theta_L, x) = f_{L-1}\left(\theta_{L-1}, f_{L-2}\left(\theta_{L-2}, \dots f_1(\theta_1, x)\right)\right)$ incorporates feature engineering internally. Hence are able to extract information in a faster and more direct way.
- Convolution: applying and sliding a 1d filter over time (images: 2d, width and height).
- Local/global pooling: avg/max over a sliding window or the whole time series.



CNNs for TSC: Experiments

Accuracy $\mu(\sigma)$ over 10 runs of each algorithm: *ResNet (deep flexible architecture)* outperforms other CNNs.

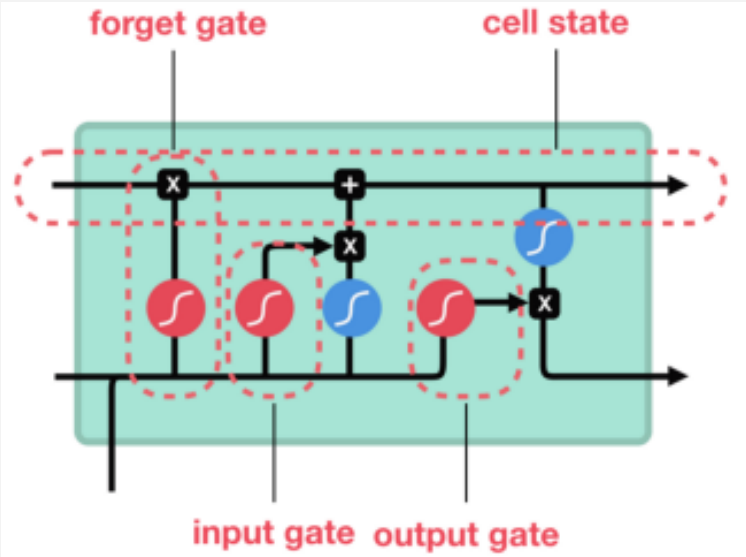
Datasets	Multi layer perceptron (MLP)	Fully Convolutional Networks (FCN)	Residual network (ResNet)	Encoder	Multi-scale CNN (MCNN)
Coffee	99.6(1.1)	100.0(0.0)	100.0(0.0)	97.9(1.8)	51.4(3.5)
Beef	72.0(2.8)	69.7(4.0)	75.3(4.2)	64.3(5.0)	20.0(0.0)
ECG200	91.6(0.7)	88.9(1.0)	87.4(1.9)	92.3(1.1)	64.0(0.0)
ECG5000	92.9(0.1)	94.0(0.1)	93.4(0.2)	94.0(0.2)	61.8(10.9)
50words	68.4(7.1)	62.7(6.1)	74.0(1.5)	72.3(1.0)	22.0(24.3)
TwoLeadECG	76.2(1.3)	100.0(0.0)	100.0(0.0)	86.3(2.6)	50.0(0.0)
Gun_Point	92.7(1.1)	100.0(0.0)	99.1(0.7)	93.6(3.2)	51.3(3.9)
Plane	97.8(0.5)	100.0(0.0)	100.0(0.0)	97.6(0.8)	13.0(4.5)
Symbols	83.2(1.0)	95.5(1.0)	90.6(2.3)	82.1(1.9)	22.6(16.9)
wafer	99.6(0.0)	99.7(0.0)	99.9(0.1)	99.6(0.0)	91.3(4.4)

RNNs for TSC

RNNs have remarkable performance on sequential learning problems.

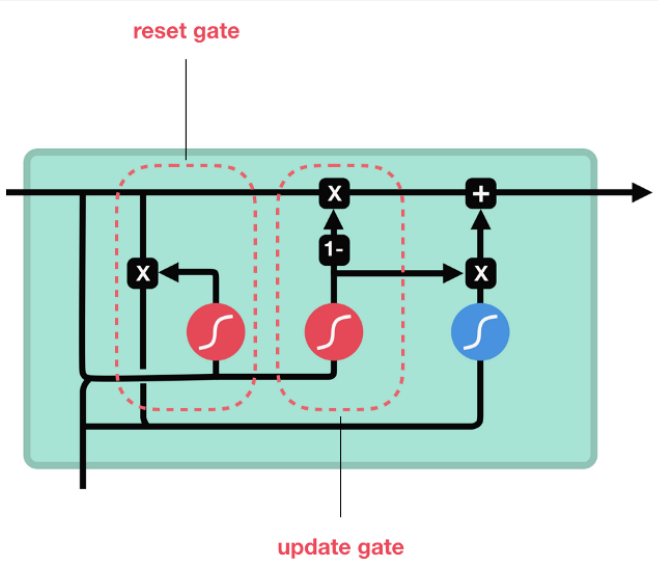
- However, long sequence learning with RNNs remains a challenging problem:
  1. **Short-term memory**, hard to memorize extremely long-term dependencies
  2. Training RNNs with back-propagation-through-time: **vanishing and exploding gradients**
  3. Forward and back-propagation are performed **sequentially**; time-consuming.

Long Short-Term Memory (LSTM) and **Gated Recurrent Units** (GRU) models powerfully model complex data dependencies. Gates are just neural networks that regulate the flow of information through the sequence chain.



LSTMs
Cell state: transport
Input, forget, output gates
Activations: keep/forget

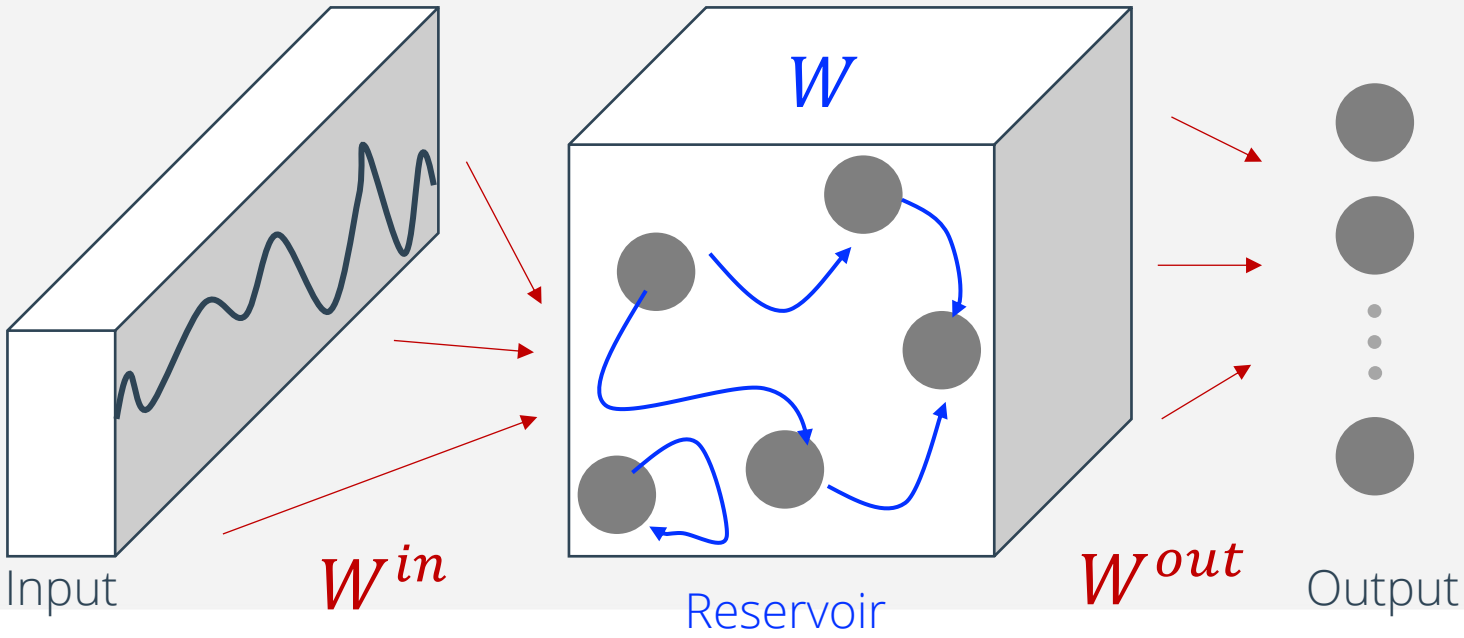
GRUs
Only two gates
Fewer tensor operations
Faster to train than LSTMs



RNNs for TSC: Architectures

Echo State Networks (ESNs), reservoir computing

- TSC in wireless communication channels
- The input signal is connected to a fixed (non-trainable) and random dynamical sparsly-connected system (the reservoir).
- Vanishing gradients: in BP, the gradient can get vanishingly small, effectively preventing a weight from getting updated. ESNs eliminate the need to compute the gradient for the hidden layers.



Jaeger, Herbert, and Harald Haas. "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication." *science* 304.5667 (2004): 78-80.

RNNs for TSC: Experiments

Accuracy over 1 run of each algorithm: *ESNs are competitive*. More experiments to follow.

Datasets	MLP	FCN	ResNet	LSTM	ESN
ECG200	0.920	0.900	0.870	0.890	0.920
ECG5000	0.935	0.941	0.931	0.943	0.944
50words	0.712	0.679	0.727	0.688	0.758
LrgKitApp	0.480	0.896	0.893	0.902	0.901
FordA	0.769	0.906	0.928	0.932	0.932
Plane	0.981	1.00	1.00	0.933	1.00
wafer	0.996	0.997	0.997	0.992	0.997

Conclusions

Work in progress …

- Choosing the right time-series classification algorithm depends heavily on the problem domain and discriminative features of the training data.
- *No one solution fits all.*
- New research for efficient time series classification methods is needed for emerging mission problems.

Next:

- Computational cost (runtime) comparisons
- Dropout for UQ measures in CNNs and RNNs

Thank you!

For questions or follow up discussions:

Ahmad Rushdi rushdi@stanford.edu
Erin Acquesta eacques@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.