EUROfusion

Open world learning: a new paradigm for disruption prediction

M.Gelfusa, A.Murari, M.Lungaroni, R.Rossi, L.Spolladore, J.Vega and JET Contributors

TOR VERGATA UNIVERSITY OF ROME

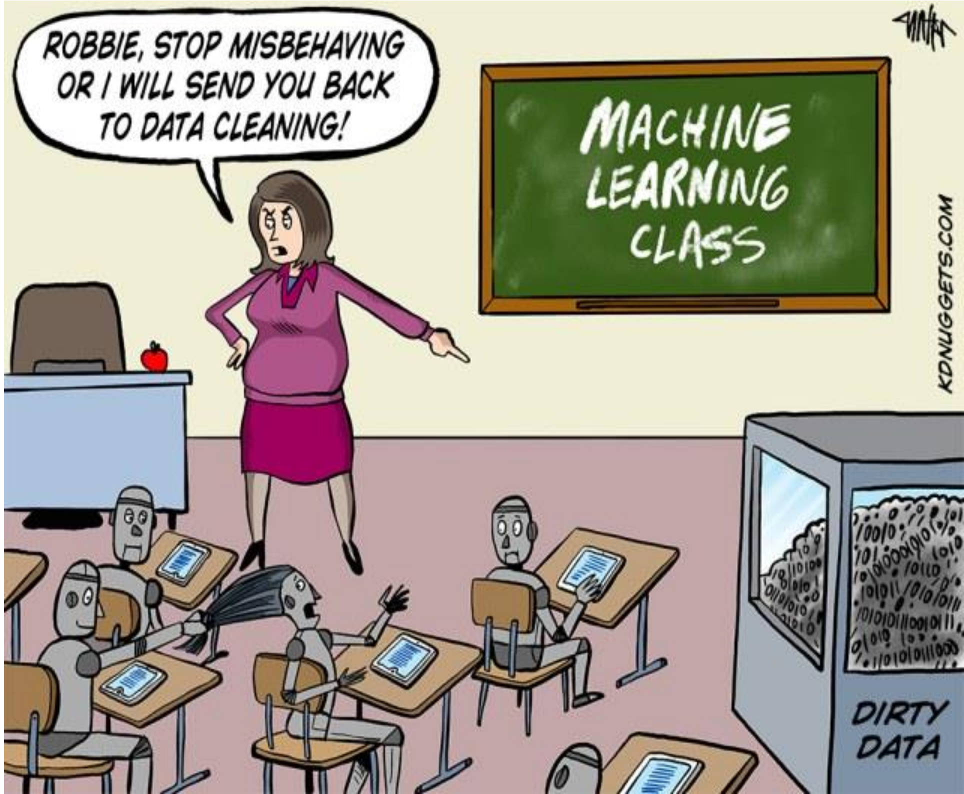
CONSORZIO RFX Ricerca Formazione Innovazione

Ciemat

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Machine Learning

- Machine Learning is the new frontier in Artificial Intelligence.
- It is likely that you interact with machine learning tools many times every day: suggestions by retailers, investment allocations, medical diagnosis etc.
- One of the most important challenges to Machine Learning and modern statistics in general is <u>learning in nonstationary conditions</u> (when the systems evolve).



Closed-World Learning

Traditional supervised Machine Learning is based on the *closed-world assumption:*

- The systems under study must be stationary. The i.i.d. assumption (data independent and identically distributed) means that the results are valid only if the pdf of the data are the same for the training set, the test set and the final application.
- All the classes in the test and final applications must have been seen in the training (with suitable number of examples).

  - **Excessive amounts of data for the training**
  - **Fast obsolescence**
  - **Lack of transferability**

Open-World Machine Learning

Adaptive learning: predictors are updated when appropriate to track the evolution of the phenomena to be predicted. Two main types of adaptation have been implemented for JET to reflect the different time scales involved during and between discharges.

a) Updates of the training sets (including de-learning) and decision functions between discharges

b) Trajectory learning during discharges.

Transfer learning: non supervised clustering to identify new classes (we also transferred one predictor from AUG to JET)

Non Stationary conditions

In Tokamaks there are two main historical effects which violate the i.i.d. assumption: a) Evolution of the experimental programme between discharges b) Memory effects during shots.

a) reasonable computational time available (discharge repetition time half an hour) but great variations

b) very limited amount of time available (on JET real time system cycle time is 2 ms)

Traditional predictors not considering non stationarity therefore are not very stable and start to behave sub optimally very rapidly

Possible solution: Adaptive Learning.

Outline

- Overview of Ensemble Classifiers
- Strategies of Adaptive Learning for predictions in non stationary conditions
- Results for the ILW
- Conclusions and future lines of investigation

Classification And Regression Trees: CART

- The methods implemented and refined to perform the studies described in the rest of the talk are based on the Classification And Regression Tree (CART) technology.

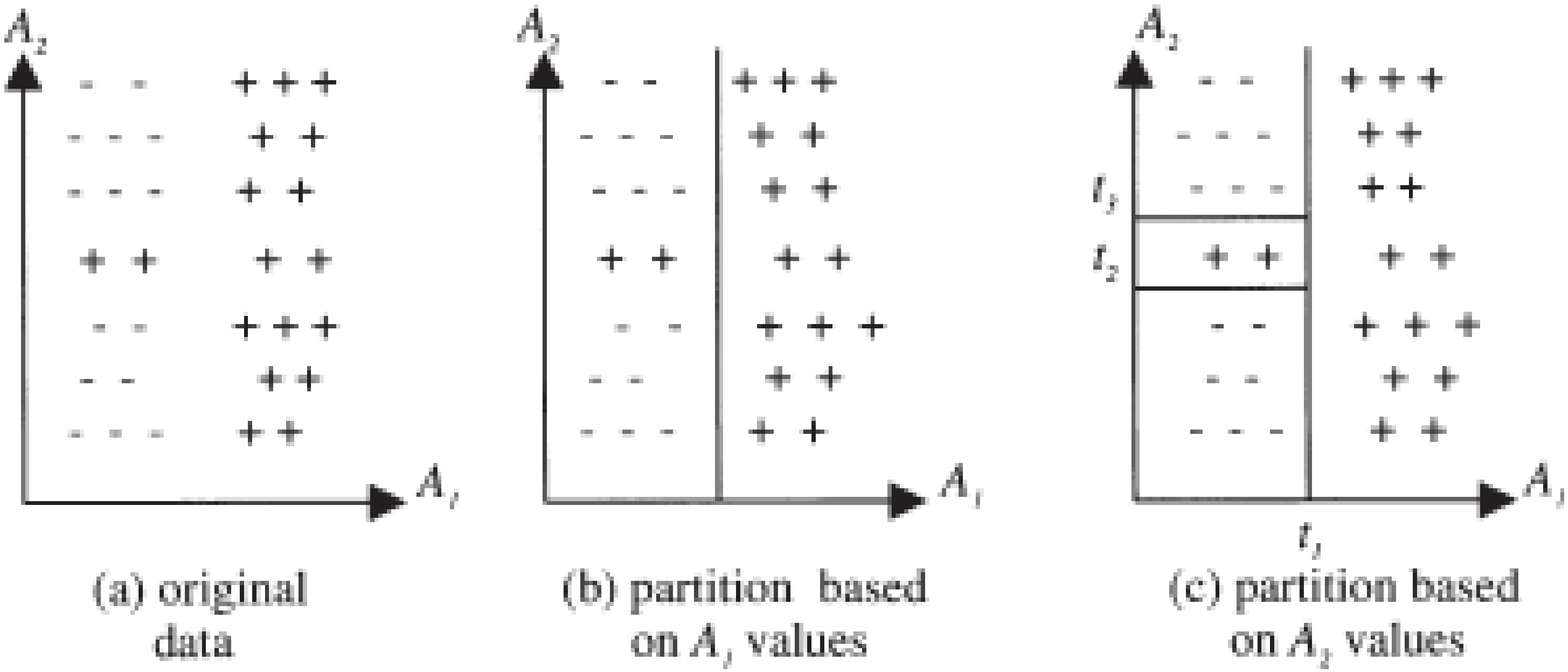
- CART is a supervised methodology. The models are derived directly from the available databases by recursively partitioning the data space and fitting a simple rule at each partition.

- The final partitioning, once properly optimised, consists typically of a series of “propositional logic” rules that can be represented graphically by a decision tree.

- Rule-based classifiers of the CART family are very powerful, easy to interpret and computationally efficient.

CART: recursive partitioning

During the training, the CART approach selects recursively the best variable to separate the examples of the various classes.



The final model can be represented either as a tree or a series of elementary rules of the type ***if……then……***

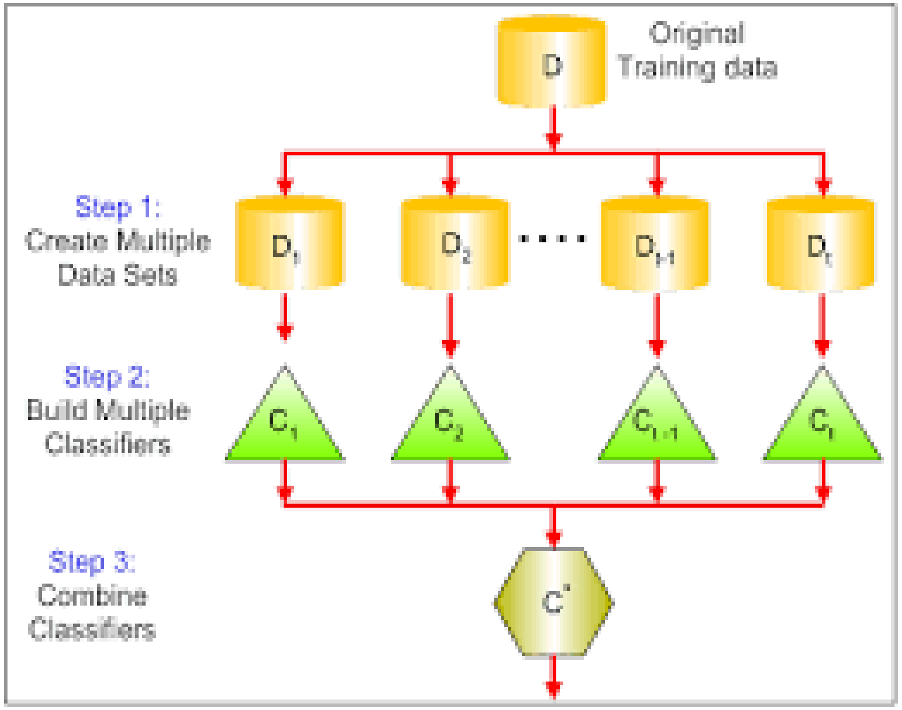
The results have the representational power of <u>propositional logic</u>

Weak Learning

CART trees are not very stable; small changes in the training set can result in major differences in the final trees and therefore in the final classification.

- A 'weak' learner (either classifier or predictor) is just a machine learning tool, which produces a model that performs relatively poorly but is computationally not too demanding.
- The relatively limited computational resources required allow training various versions of such weak learners which can then be pooled together to create a "strong" **ensemble classifier**.

The trick is to increase diversity by training with slightly different sets.



The basic classifiers used as weak learners are CART trees.

Ensemble Classifiers: diversity

Three ensemble classifiers have been implemented: Bagging, Random Forests and Noise-based ensembles.

Bagging

- Generation of many random sub-samples of the original dataset with replacement.

Random Forests

- Sample the original dataset at random with replacement to create a subset of the data (as a bag).
- At each node also select at random a subset of predictor variables from all the predictor variables

Noise based Ensembles

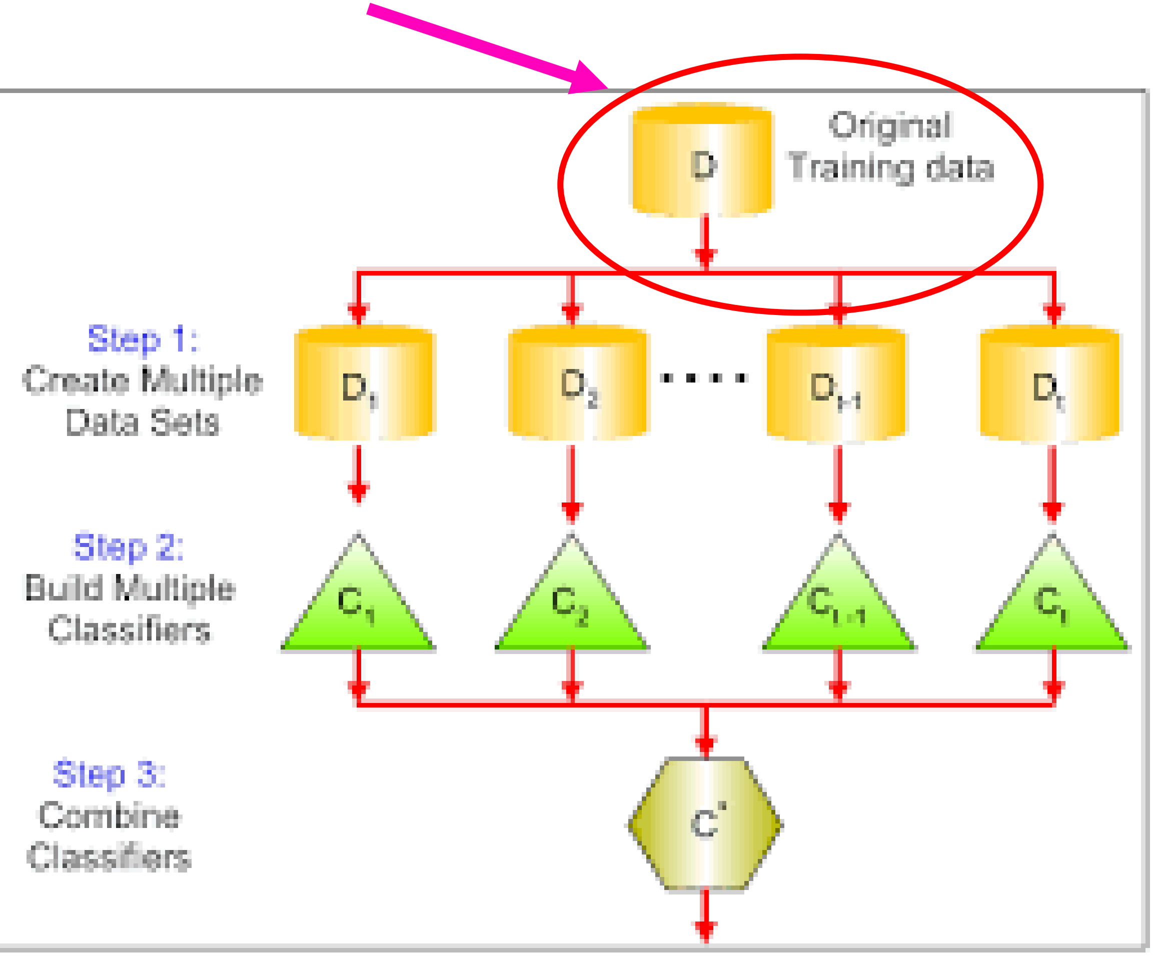
- The idea consists again of collecting ensembles but not with subsets of the original data; on the contrary the various training sets are obtained by the original one summing random noise to each entry.

Outline

- Overview of Ensemble Classifiers
- Strategies of Adaptive Learning for predictions in non stationary conditions
- Results for the ILW
- Conclusions and future lines of investigation

Updating the Training Set

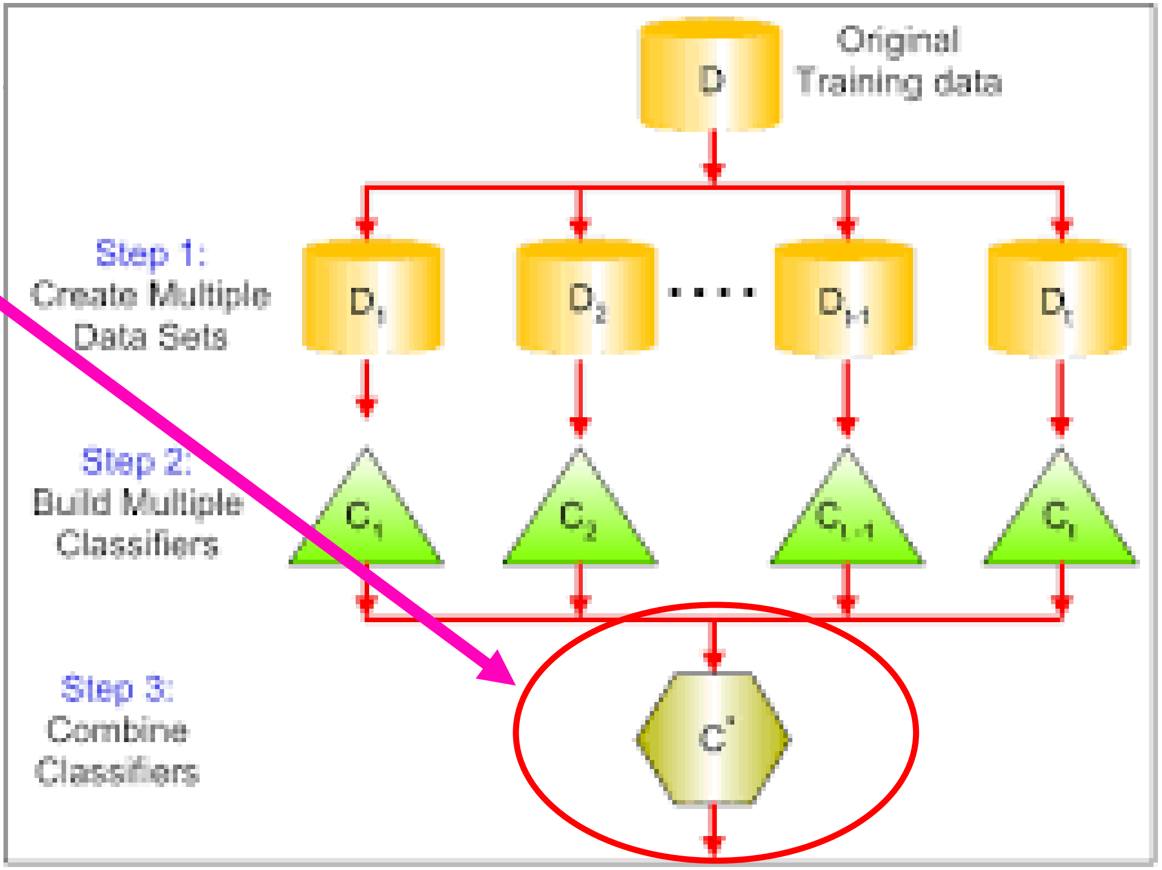
The *training set* is updated according to two different criteria.



- When there is a error in the prediction (for example a missed or a tardy alarm).
- To implement de-learning: old examples are discarded when they become obsolete and therefore misleading.

Updating the decision function

The ensembles are pooled and the final output is obtained with a *decision function*

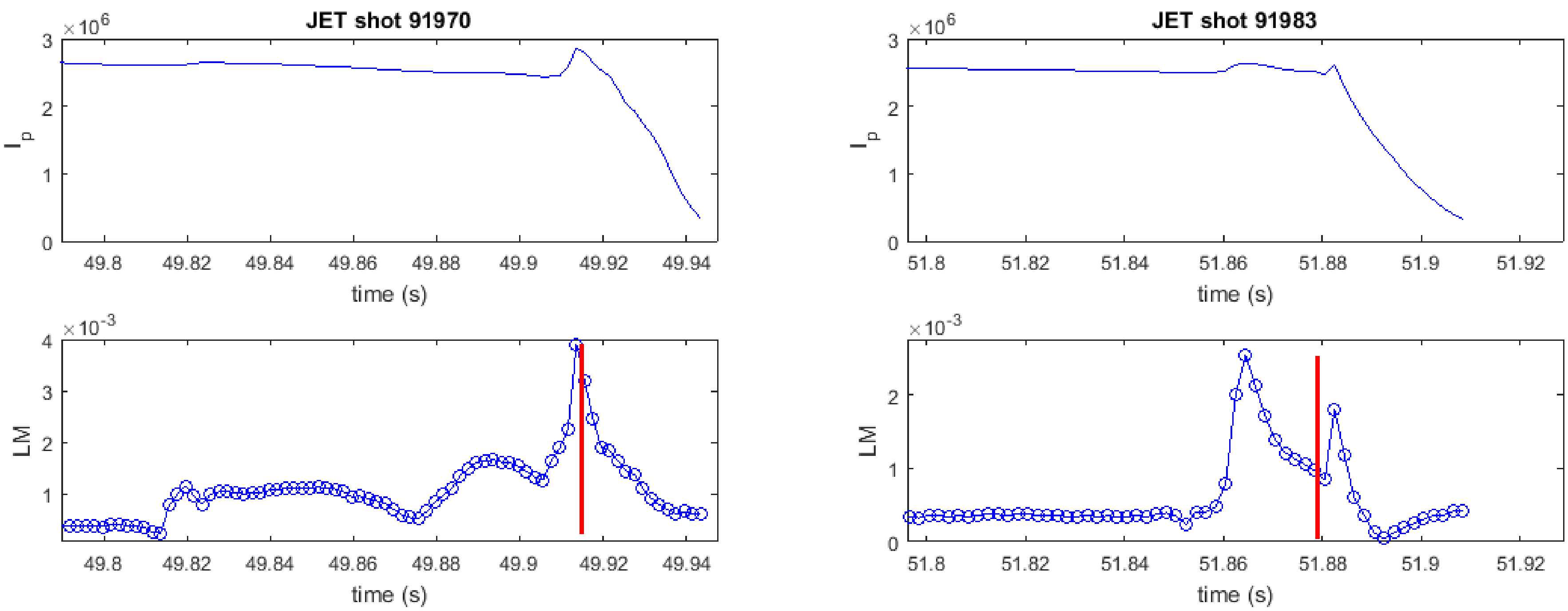


Various decision functions are run in parallel and the one with the best results so far is used to generate the alarm.

At this stage one can optimise *de-learning*, the rejection of old and therefore misleading examples.

Trajectory learning

In ***trajectory learning***, the training set contains the history of the data (sequence of samples) so that the predictors can learn the system trajectory in the feature space.



Statistically, the trajectory of the LM amplitude can be different depending on the shot.

Outline

- Overview of Ensemble Classifiers
- Strategies of Adaptive Learning for predictions in non stationary conditions
- Results for the ILW
- Conclusions and future lines of investigation

Database with the ILW wall and settings

The DB analyzed covers campaigns C28-C32 (430 disruptions and 1998 safe shots) with 1 ms time resolution and all time slices with Ip > 750 kA.

Tardy alarms: if the alarm is triggered less than 5 ms from the beginning of the current quench.

Machine learning tools: single CART, Ensembles of CARTS using the locked mode and internal inductance as inputs.

The first model is obtained after the first disruption (from scratch).

Trajectory learning: three points in the last 50 ms for both LM and li are sufficient

De-learning: only the last 10 shots in the training set

Results with CART and Ensembles

Method	Success Rate	Missed	Early	Tardy	False	Mean [ms]	Std [ms]
Single CART	93.24 (400/429)	0.47 (2/429)	3.50 (15/429)	2.80 (12/429)	14.39 (289/2008)	385	397
BAG	89.04 (382/429)	0.70 (3/429)	6.53 (28/429)	3.73 (16/429)	11.13 (225/2021)	369	407
RF	95.10 (408/429)	0.93 (4/429)	0.93 (4/429)	3.26 (14/429)	8.37 (167/1996)	371	401
Noise+RF	94.87 (407/429)	1.17 (5/429)	0.47 (2/429)	0.47 (2/429)	7.17 (143/1995)	364	397
Noise+BAG	95.10 (408/429)	0.70 (3/429)	0.70 (3/429)	3.73 (16/429)	7.97 (159/1996)	374	400

Decision function: majority voting.
Results competitive but weak on the front of the false alarms.

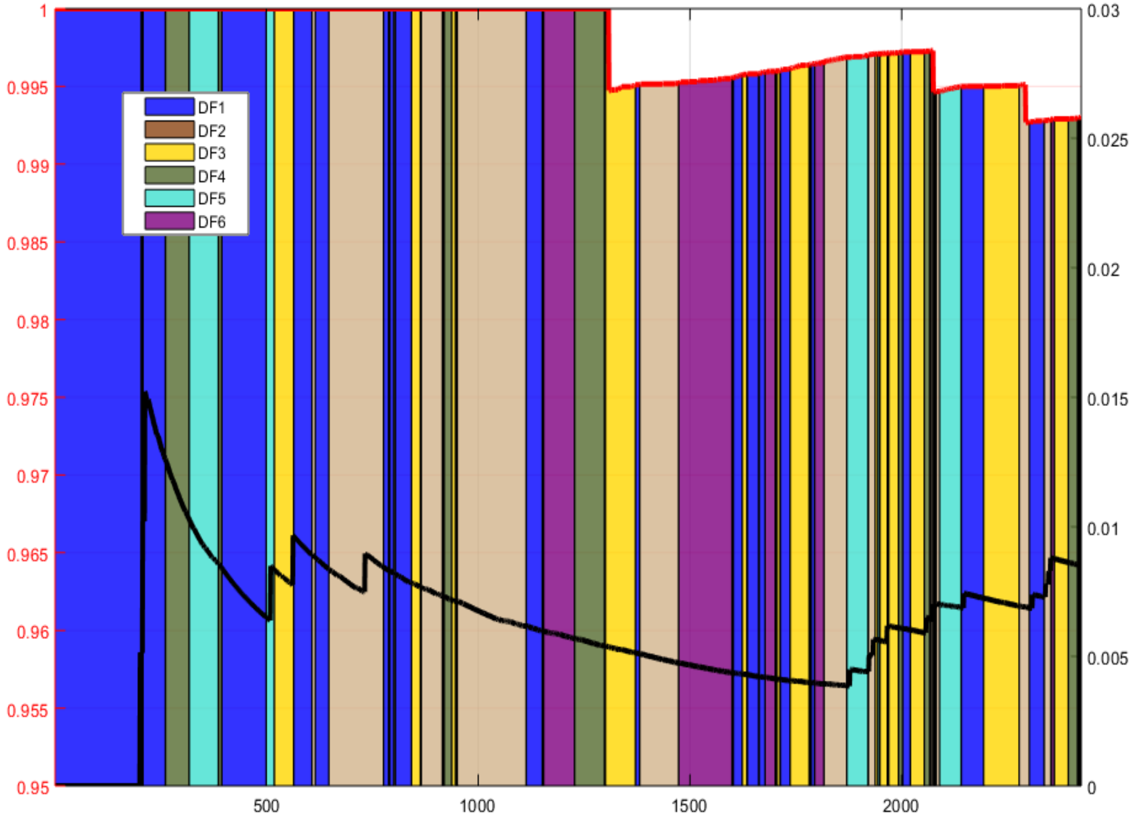
Results with Trajectory Learning

Decision Function	Success Rate	Missed	Early	Tardy	False	Mean [ms]	Std [ms]
4 RF	96.97 (416/429)	1.17 (5/429)	0.23 (1/429)	1.63 (7/429)	6.37 (127/1994)	349	358
4 BAG	97.90 (420/429)	0.93 (4/429)	0.23 (1/429)	0.93 (4/429)	6.32 (126/1994)	342	360
5 RF	97.67 (419/429)	0.70 (3/429)	0.47 (2/429)	1.17 (5/429)	7.07 (141/1995)	377	402
5 BAG	97.67 (419/429)	1.63 (7/429)	0.23 (1/429)	0.47 (2/429)	5.52 (110/1994)	348	353
6 RF	96.97 (416/429)	1.40 (6/429)	0.23 (1/429)	1.40 (6/429)	4.76 (95/1994)	335	351
6 BAG	96.27 (413/429)	1.86 (8/429)	0.23 (1/429)	1.63 (7/429)	4.61 (92/1994)	336	352

Results improve significantly both in terms of success rate and false alarms.

Results with Switching Decision Function and Delearning

Success Rate	99.3 %
False alarms	0.85 %



- Different colours indicate different decision functions
- Success rate always above 99% (overall 3 errors)
- False alarms always below 1.5 % (overall 15 errors)
- Things better than they look

Conclusions

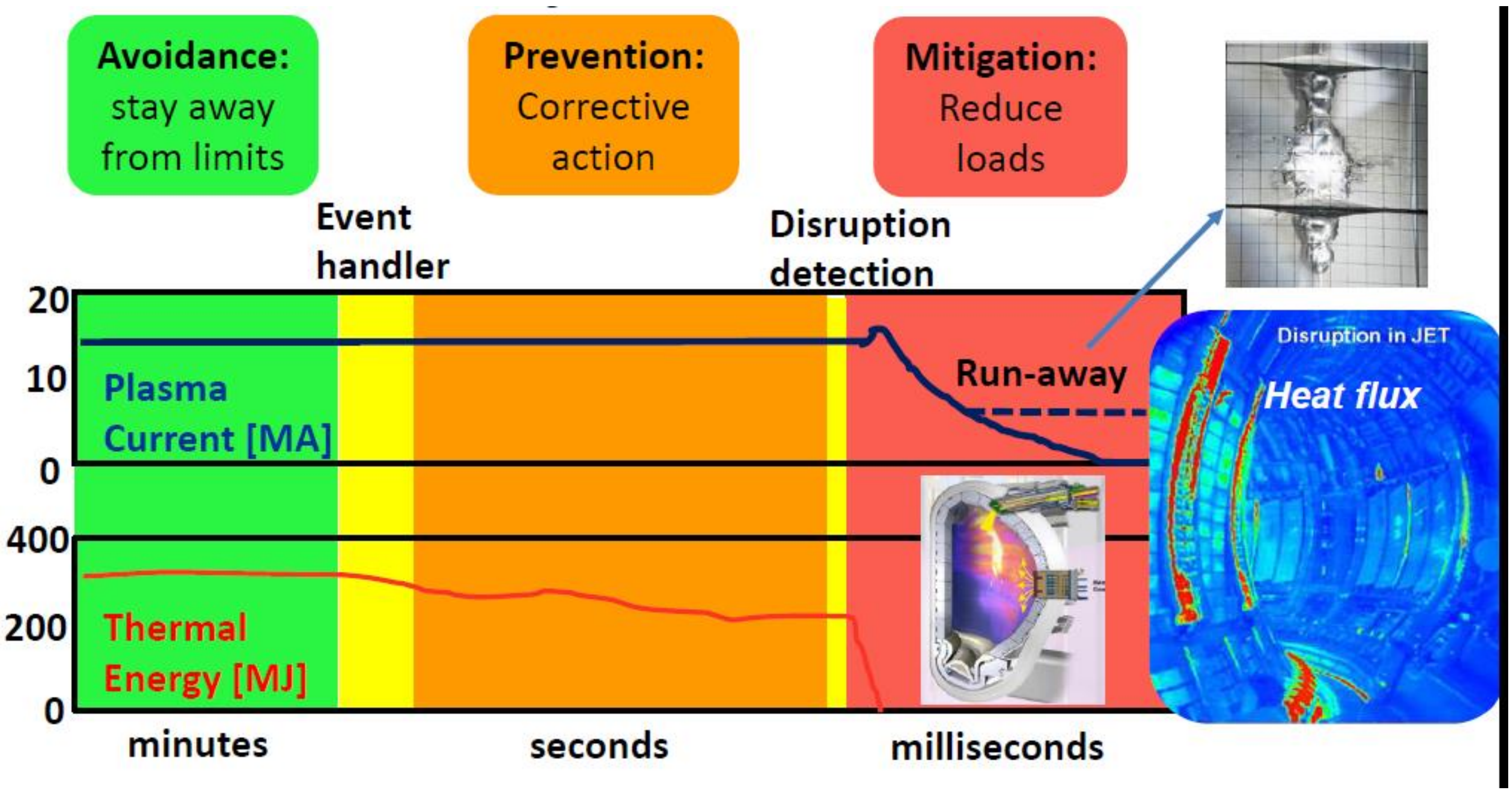
- Learning in non stationary conditions is becoming an important issue in Tokamak physics, particularly for disruption prediction.
- The innovative approach of Ensembles has proved to be sufficiently flexible to implement complex strategies of adaptive learning.
- The developed techniques of adaptive learning are being applied very successfully to JET data at the beginning of operation with the new ITER Like Wall.
- Since up to now translating predictors from one device to another has proved to be a very challenging task, the proposed approach of adaptive learning is very interesting for future devices particularly at the beginning of their operation (when no much data is available for training).
- The adaptive learning approach is a good fall-back solution for ITER given the great variations in its operational scenarios (range of currents, isotopic composition etc.)

Thanks for Your Attention!

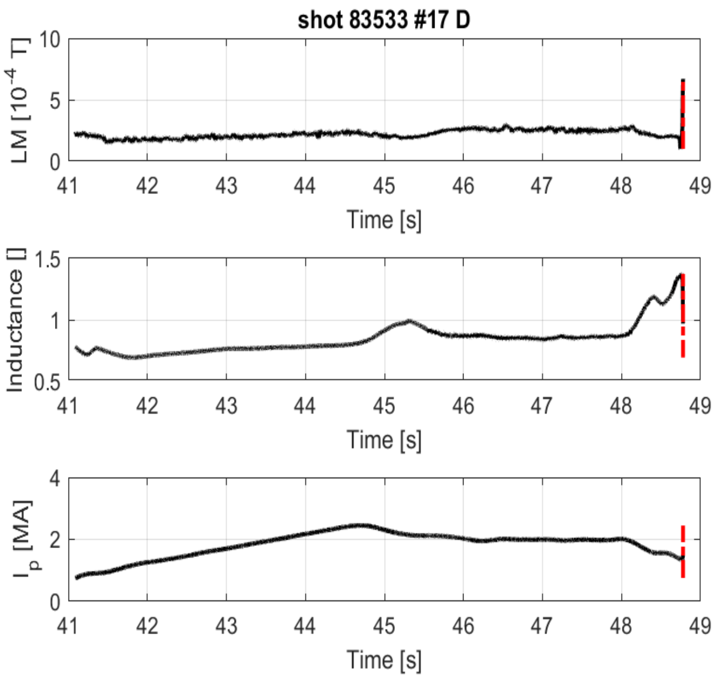
QUESTIONS?

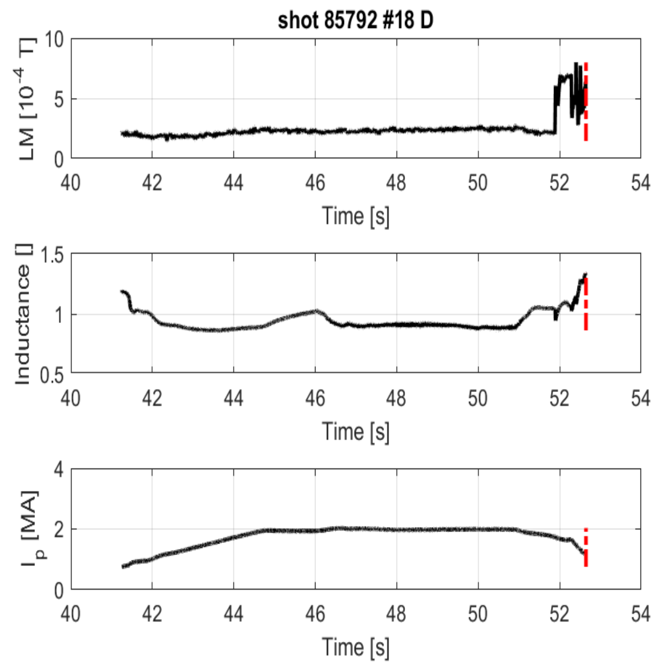
Future work

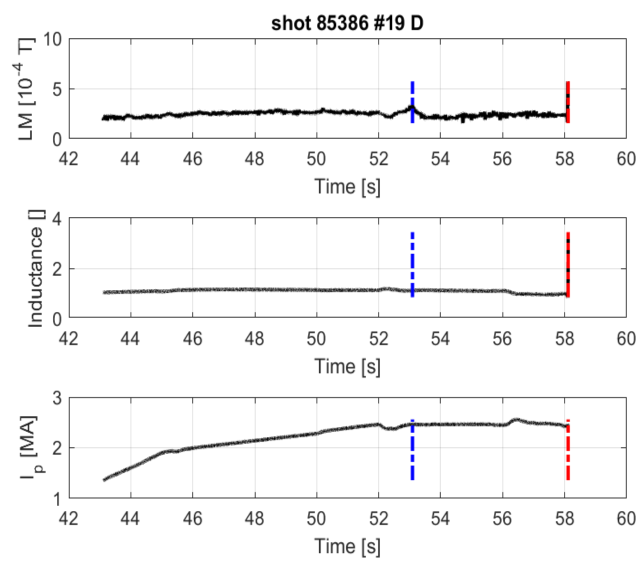
- Test the potential of the adaptive learning approach for prevention



Misclassification: Disruptions

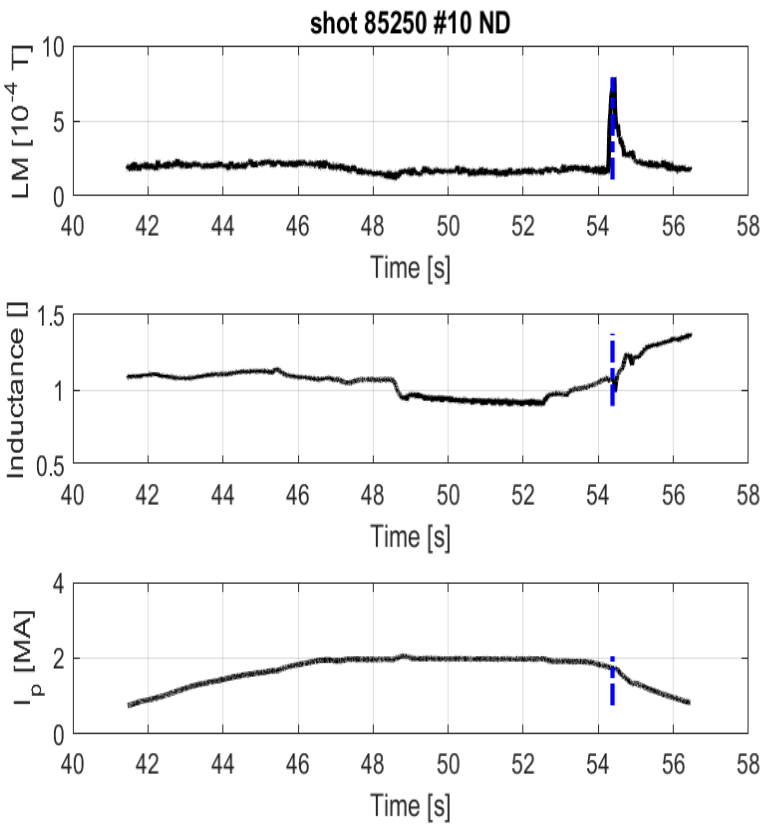


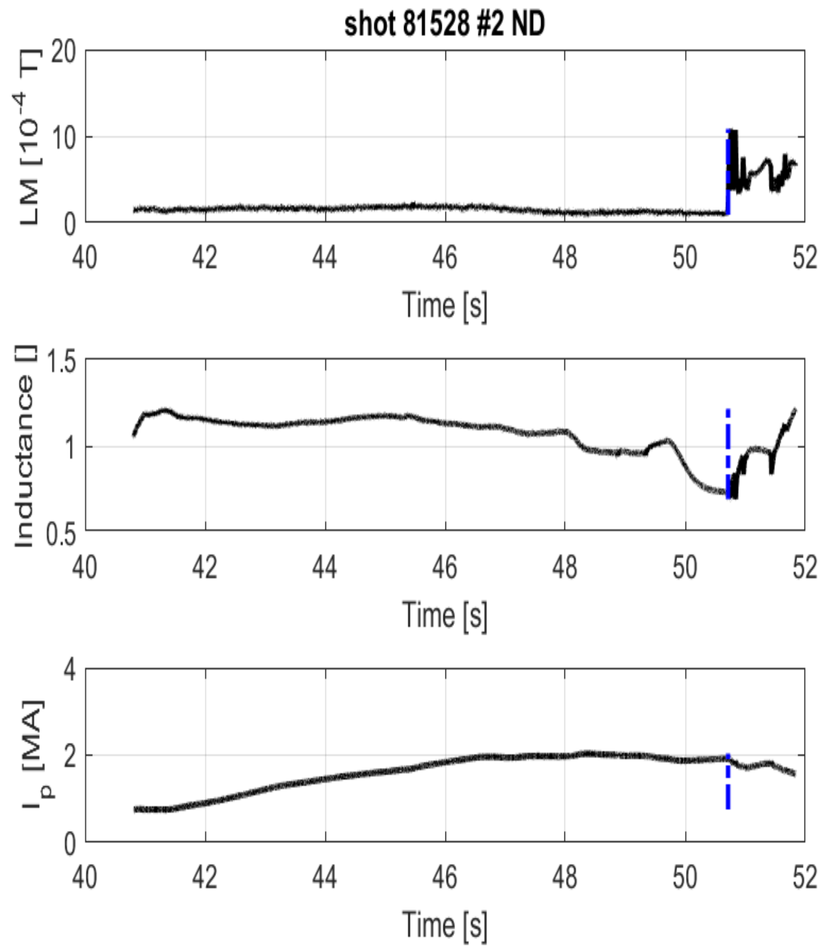




- Two late detections (top)
- One early (right)

Misclassification: false alarms





10 out of 15 false alarms are in the current decay phase of the discharge and are triggered by minor disruptions