EUROfusion

Predicting collapse: adaptive and transfer learning

Authors: A.Murari and J.Vega

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.

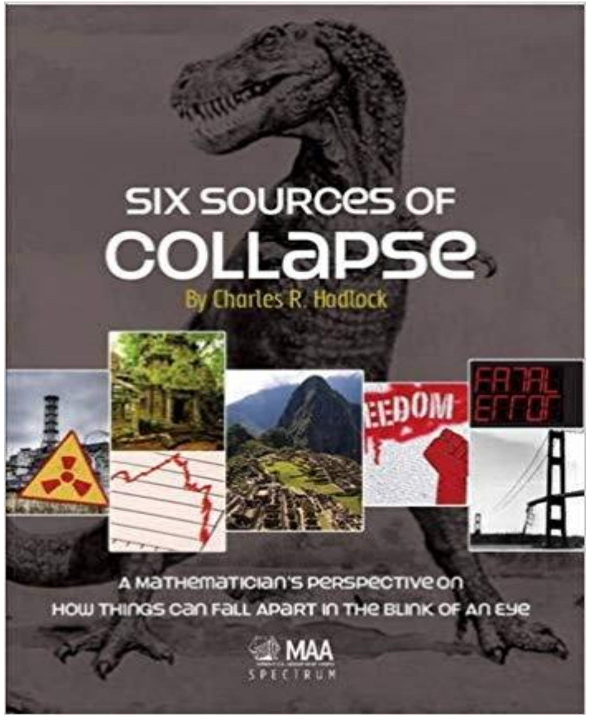
Investigation of Collapse

Many natural and man-made systems are stable for long times and look quite resilient but are nonetheless prone to catastrophic collapse.

Some of these collapses are quite straightforward to interpret and do not seem worthy of particular attention because, given the proper precautions, they are relatively easy to avoid. Others are very subtle and extremely difficult to predict. Earthquakes, and in general failures due to atmospheric phenomena, belong to the second category.

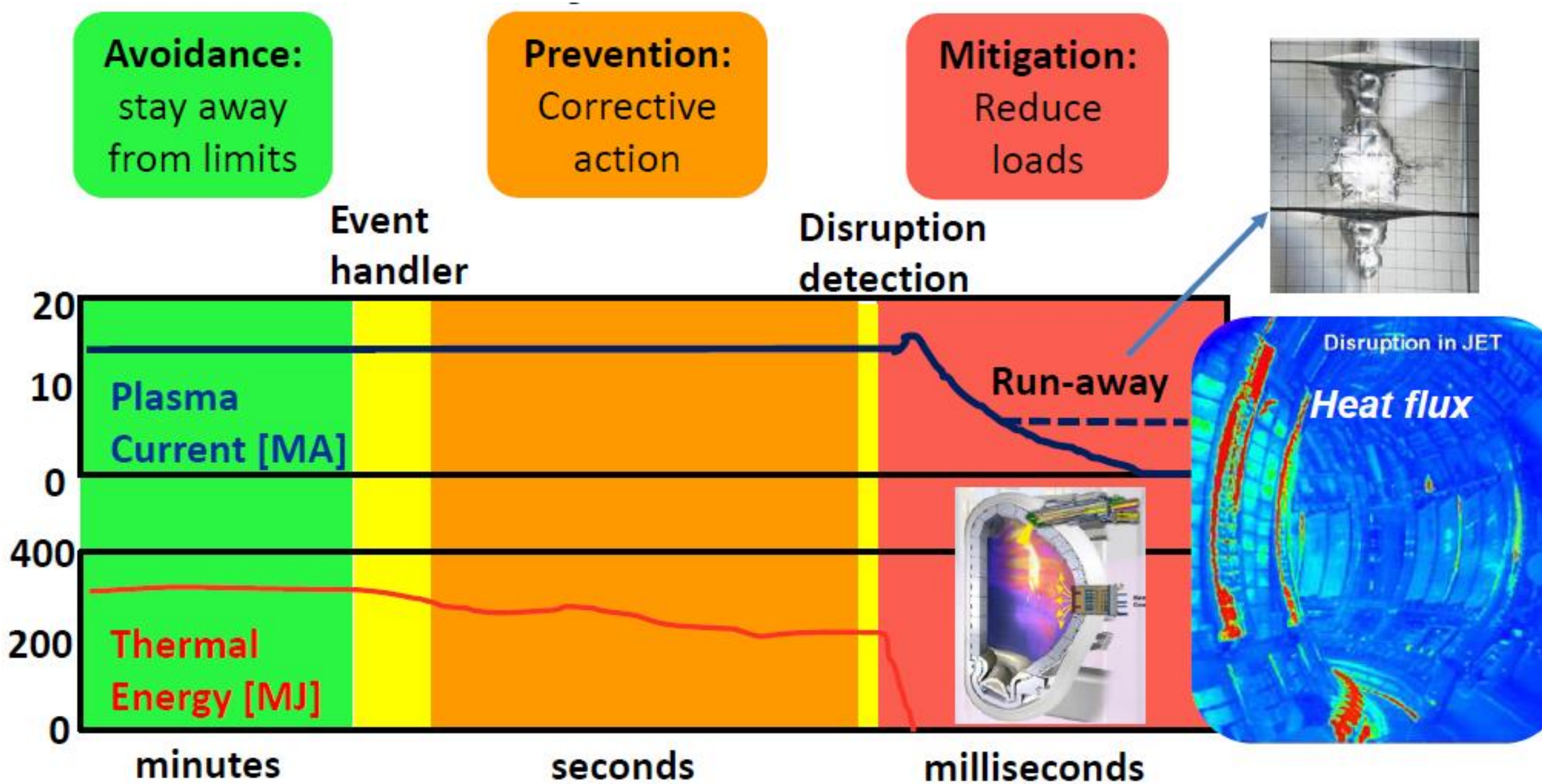
In the last years, various efforts have been devoted to developing mathematical tools more appropriate to investigating and predicting these catastrophic and rare phenomena.

Machine learning tools constitute an additional family in the arsenal of mathematical approaches which can be used to study catastrophic events



Main collapse in Tokamaks: disruptions

- Disruptions are a sudden loss of confinement and control, which lead to the extinction of the plasma current in a matter of ms.
- The damage to the devices can be severe and the problem scales badly with dimensions.



Learning in non stationary conditions

- One of the most important challenges to Machine Learning and modern statistics in general is learning in nonstationary conditions (when the systems evolve).
- The typical i.i.d. assumption at the basis of ML is therefore violated.
- 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.
- A typical violation are disruptions.

Closed-World Learning

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

- All the classes in the test and final applications must have been seen in the training (with suitable number of examples).
- 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.

Consequences: need for a lot of training data, inability to cope with the new situations, obsolescence etc. This is not the way humans learn.

Open-World Machine Learning

Motivations *for open-world learning:*

- Most systems are not stationary physical objects (*adaptive learning*).
- It would be advantageous to transfer knowledge from one problem to another (*transfer learning*).

- In Tokamaks there are two main historical effects which violate the stationarity assumption: a) Evolution of the experimental programme between discharges b) Memory effects during shots.
- Transfer Learning could be very important at the beginning of operation of new devices (different disr. types)

Decision Making under Uncertainty

"*Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns**—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.*" D.Rumsfeld

Missing: Unknowns Knowns or Prejudices

Outline

- Strategies of Adaptive and Transfer Learning for prediction in non stationary conditions
- Stacking of Classifiers
- Ensemble classifiers for Adaptive Learning
- Non supervised clustering for Transfer Learning
- Features, database and results for JET
- Conclusions and future lines of investigation

Adaptive 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.

The technology used to implement adaptive learning is ensemble of CART classifiers.

Transfer Learning

Transfer Learning consists of applying knowledge acquired to solve previous problems to new tasks.

In the present context, transfer learning is applied to the classification and identification of new disruption types.

Unsupervised clustering technologies are used to:

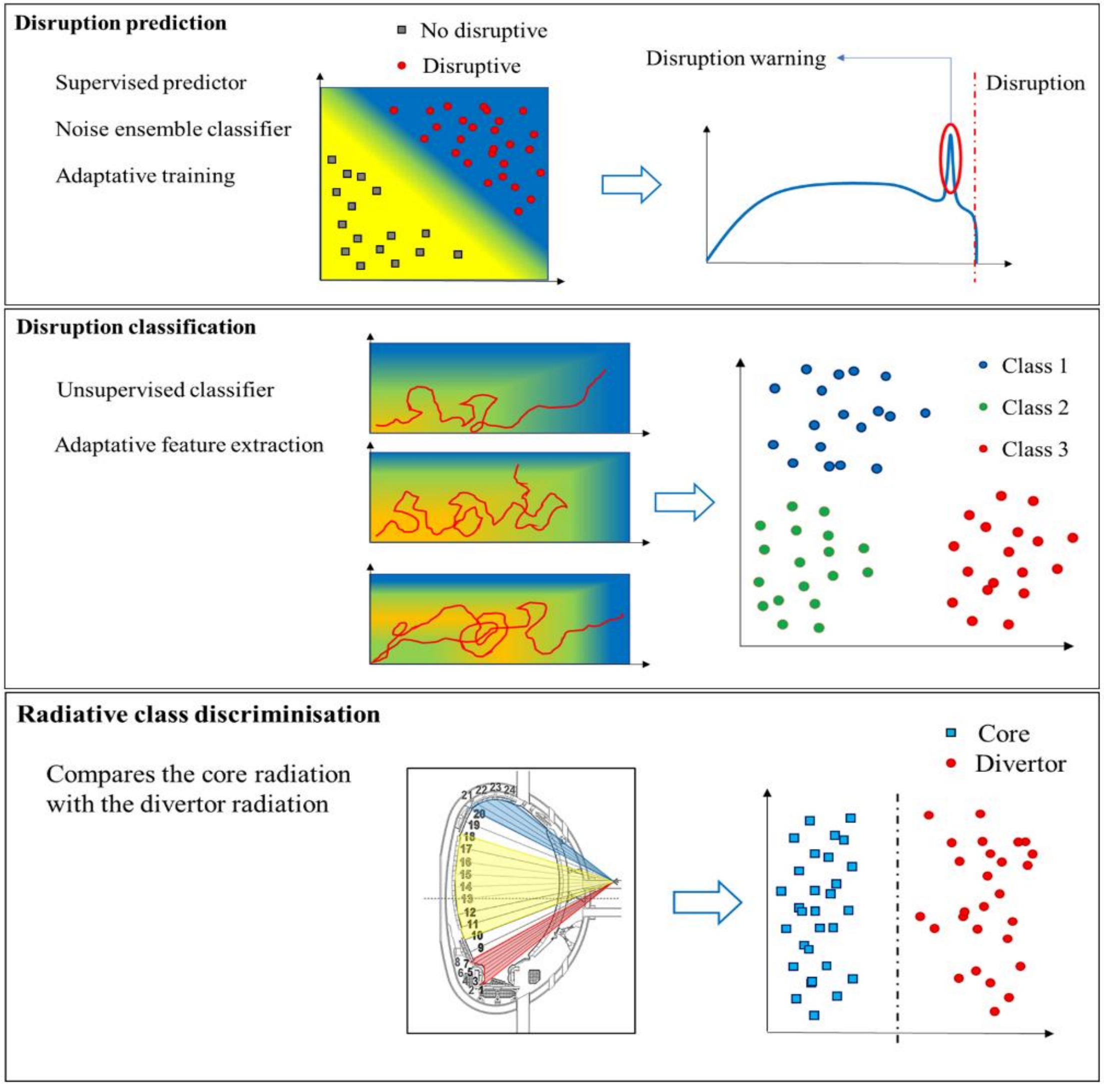
a) Attribute an incoming disruption, detected by the previous layer of the stack, to the appropriate class

b) Identify the need for a new class.

Outline

- Strategies of Adaptive and Transfer Learning for prediction in non stationary conditions
- Stacking of Classifiers
- Ensemble classifiers for Adaptive Learning
- Non supervised clustering for Transfer Learning
- Features, database and results for JET
- Conclusions and future lines of investigation

Stacking of Predictors and Classifiers



Three layers:

- Predict a disruption is about to occur
- Classify the disruption type
- If radiative disruption determine whether the problem is in the core or at the edge.

Stacking predictors or classifiers is an alternative approach to the usual philosophy of <u>general predictors</u>.

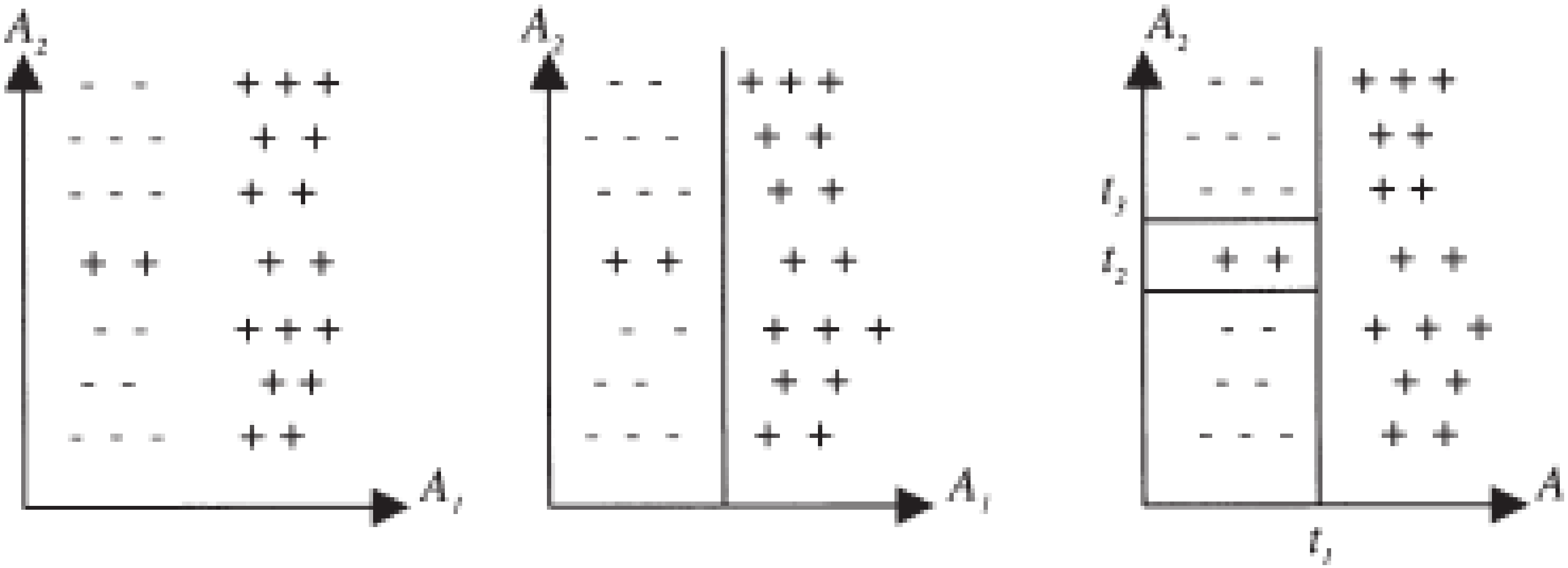
Stacked predictors remain <u>universal</u> in the sense that they are applied to entire campaigns without any a priori selection of disr.

Outline

- Strategies of Adaptive and Transfer Learning for prediction in non stationary conditions
- Stacking of Classifiers
- Ensemble classifiers for Adaptive Learning
- Non supervised clustering for Transfer Learning
- Features, database and results for JET
- Conclusions and future lines of investigation

CART: recursive partitioning

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



(a) original data

(b) partition based on A_1 values

(c) partition based on A_2 values

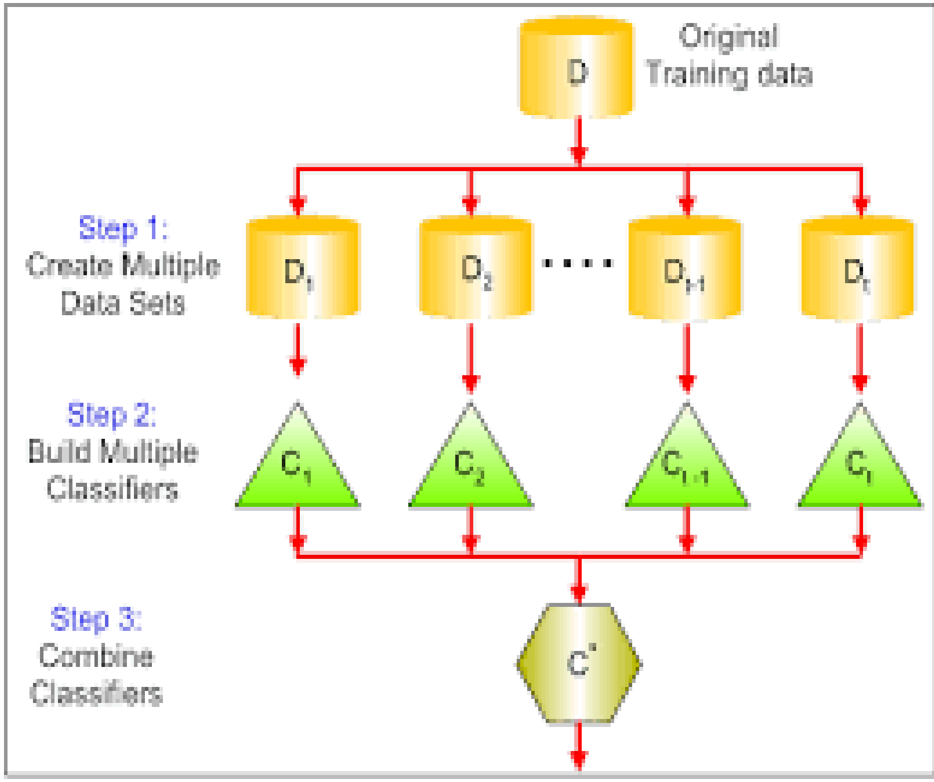
The final model can be represented either as a tree or a series of elementary rules of the type ***if……then……***

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

Ensembles of CART Classifiers

Individual CART classifiers are often 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 .

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

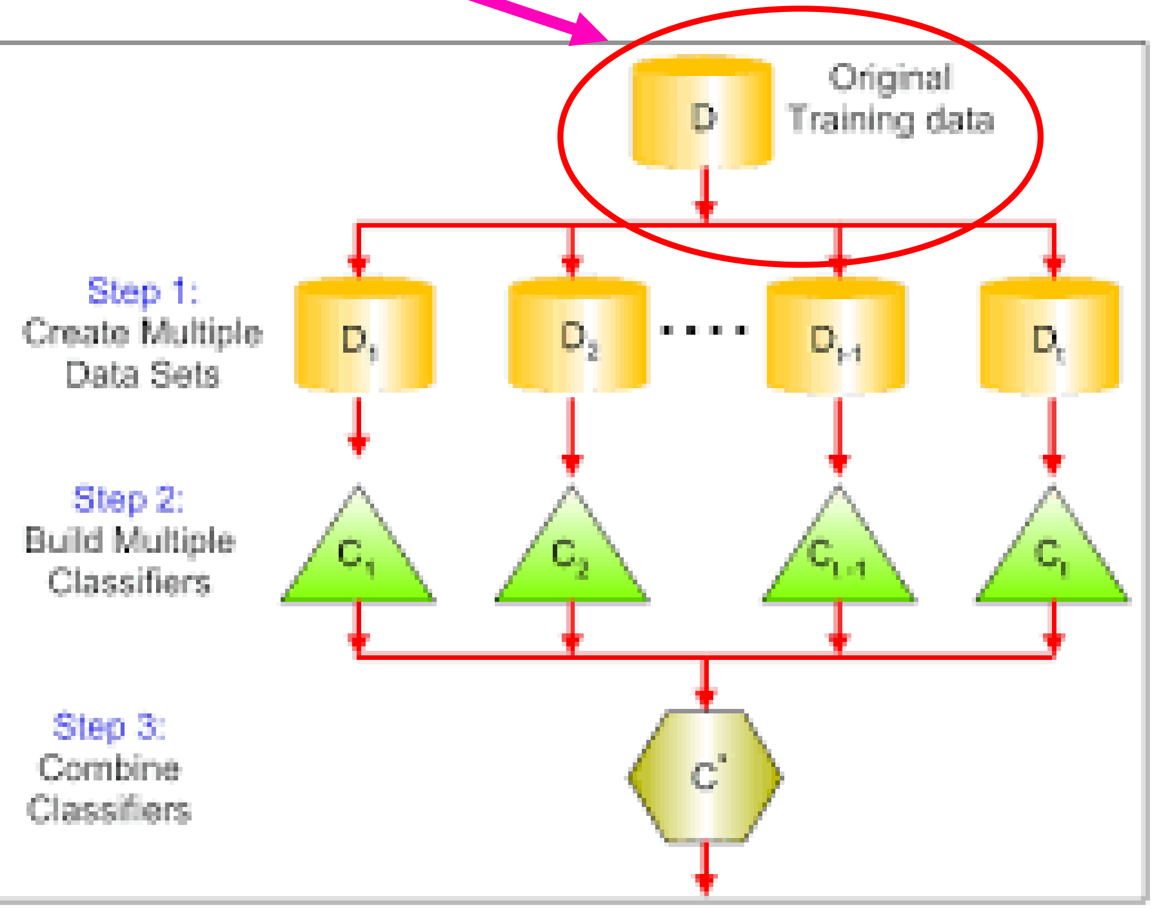
- Sample the original dataset at random with replacement to create a subset of the data (as a bag).
- At each node also select at random a subset of predictor variables from all the predictor variables

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.

Updating the Training Set between discharges

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



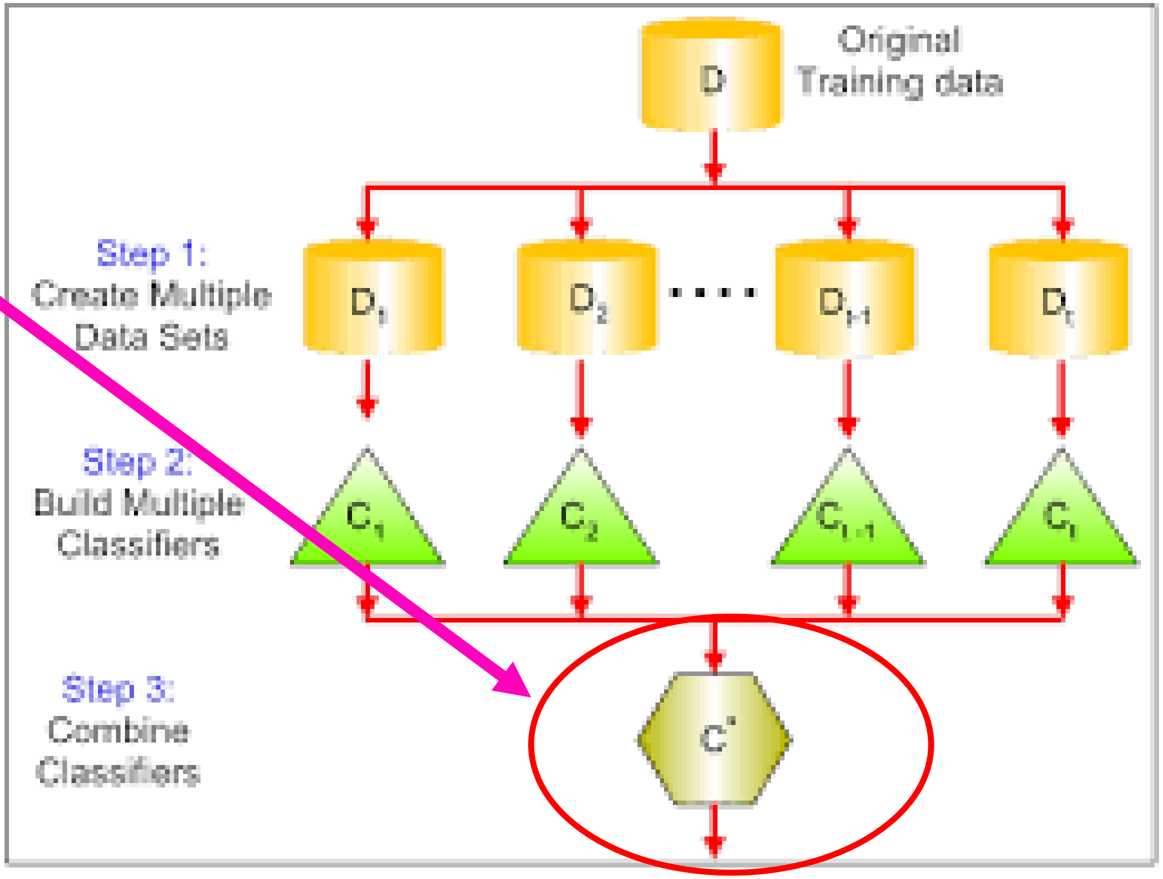
- When there is a error in the prediction (for example a missed or a tardy alarm).
- To implement de-learning: old examples are discarded when they become obsolete and therefore misleading.

Violation of Occam's razor (more complex models less generalising)

Updating the decision function between discharges

The ensembles are pooled and their 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.

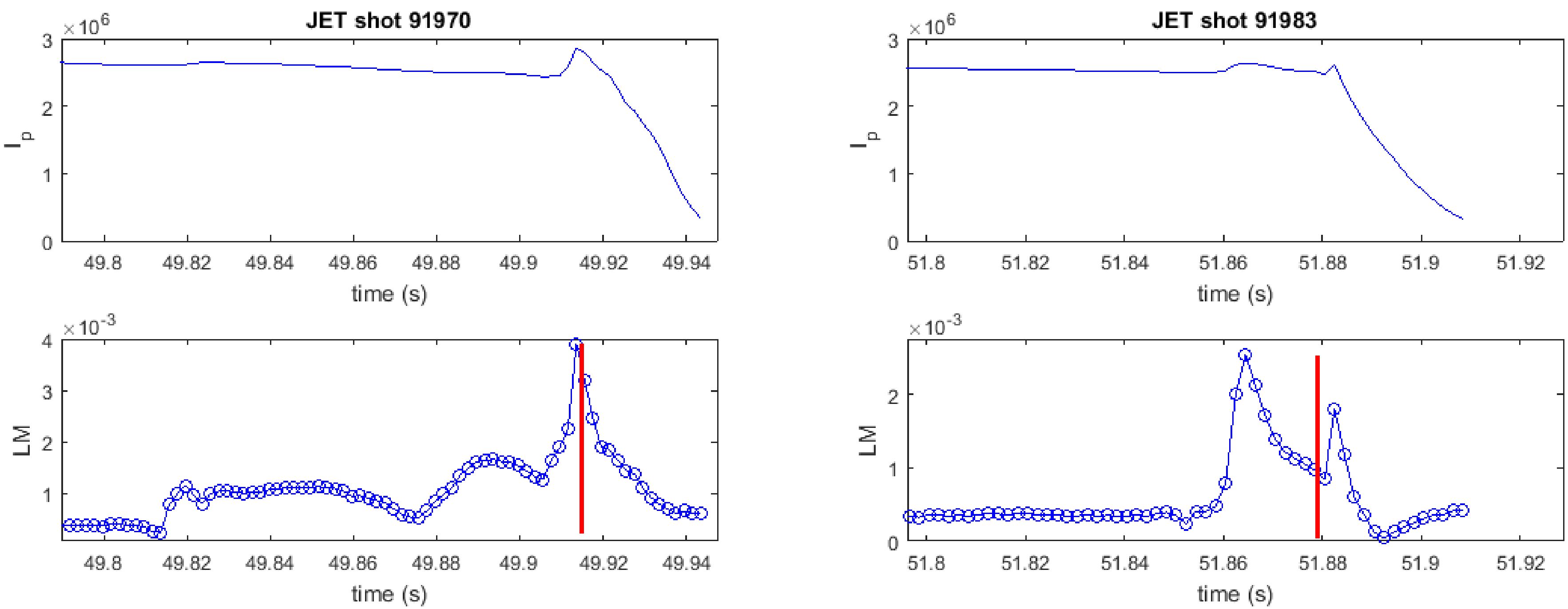


At this stage one can optimise *de-learning*, the rejection of old and therefore misleading examples.

Other prejudice: the more you know the better.

Trajectory learning during discharges

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



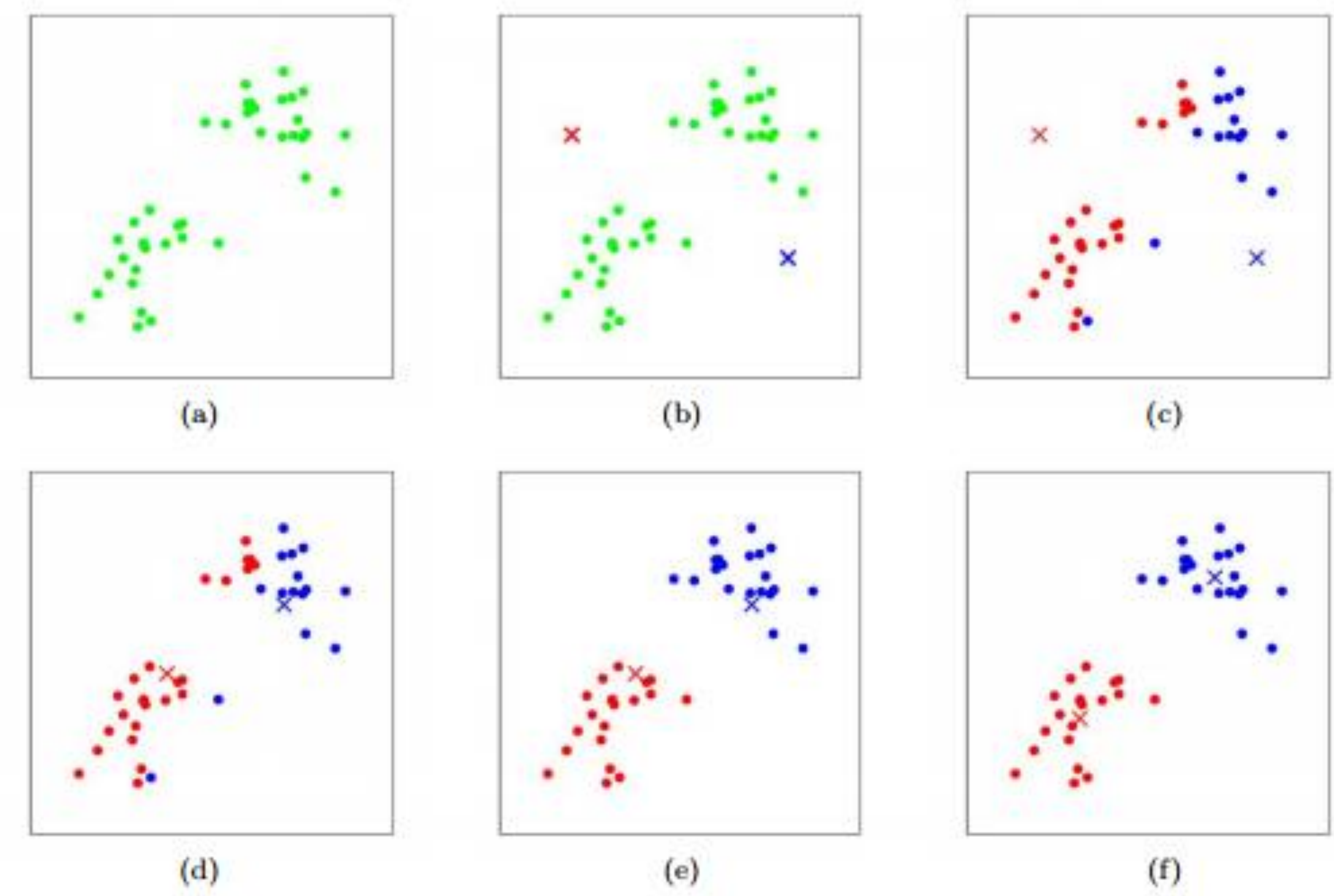
Statistically, the trajectory of the ML amplitude can be different depending on the shot.

Outline

- Strategies of Adaptive and Transfer Learning for prediction in non stationary conditions
- Stacking of Classifiers
- Ensemble classifiers for Adaptive Learning
- Non supervised clustering for Transfer Learning
- Features, database and results for JET
- Conclusions and future lines of investigation

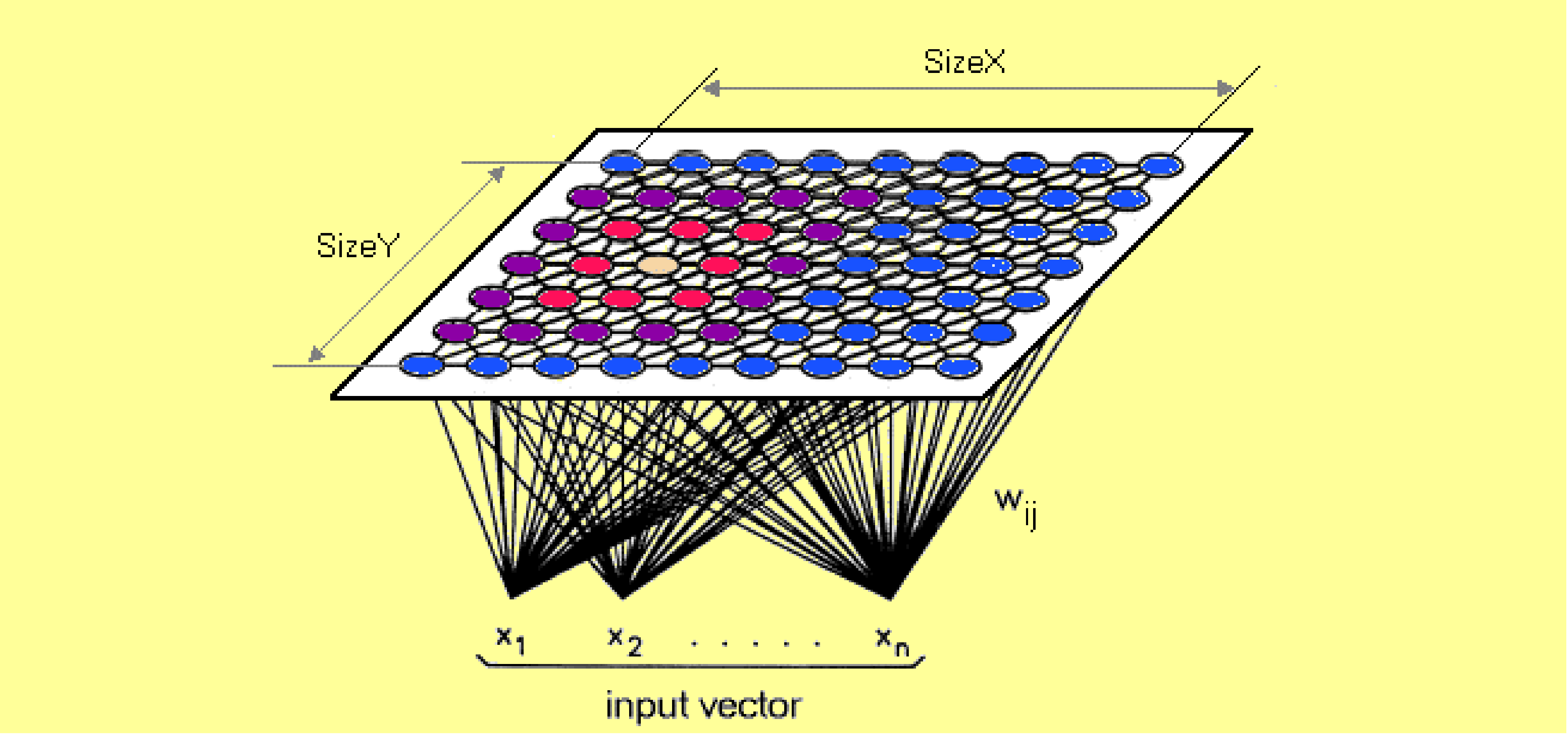
K-Means

K-means is an unsupervised clustering method implemented with the expectation maximization algorithm.



SOM: self organizing maps

. A **self-organizing map** (**SOM**) is a type of ANN that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a **map.**



Outline

- Strategies of Adaptive and Transfer Learning for prediction in non stationary conditions
- Stacking of Classifiers
- Ensemble classifiers for Adaptive Learning
- Non supervised clustering for Transfer Learning
- Features, database and results for JET
- Conclusions and future lines of investigation

Database JET with the ILW wall and settings

The DB analyzed covers campaigns C28-C32 and C36-C36b (overall 588 disruptions and 3019 safe shots) with 1 ms time resolution and all time slices with Ip > 750 kA.

Tardy alarms: if the alarm is triggered less than 10 ms from the beginning of the current quench.

Early alarms: triggered more than 3 s from the beginning of the current quench.

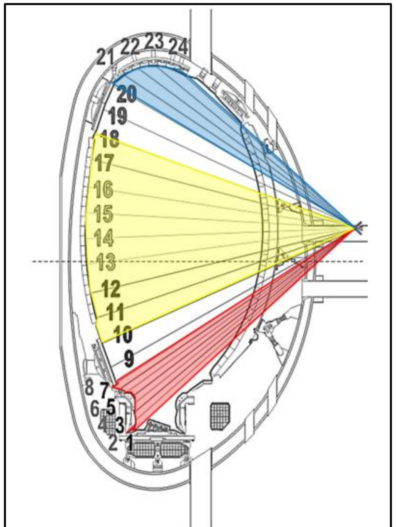
The first model is obtained after the first disruption and 5 safe discharges (<u>from scratch</u>). <u>No selection</u> of the discharges (universal), except for the intentional disruptions. Procedure <u>fully automated</u>: no human intervention.

Features and Classes

The features used as inputs to the predictors have to satisfy various conditions: 1) being reliable 2) being available in real time 3) being sufficiently selective.

For mitigation the inputs used are: normalized locked mode amplitude (and its std deviation), internal inductance, line integrated density and the percentage of energy radiated Λ .

For prevention a profile factor of bolometry has been added (ratio of lower versus middle lines of sight).



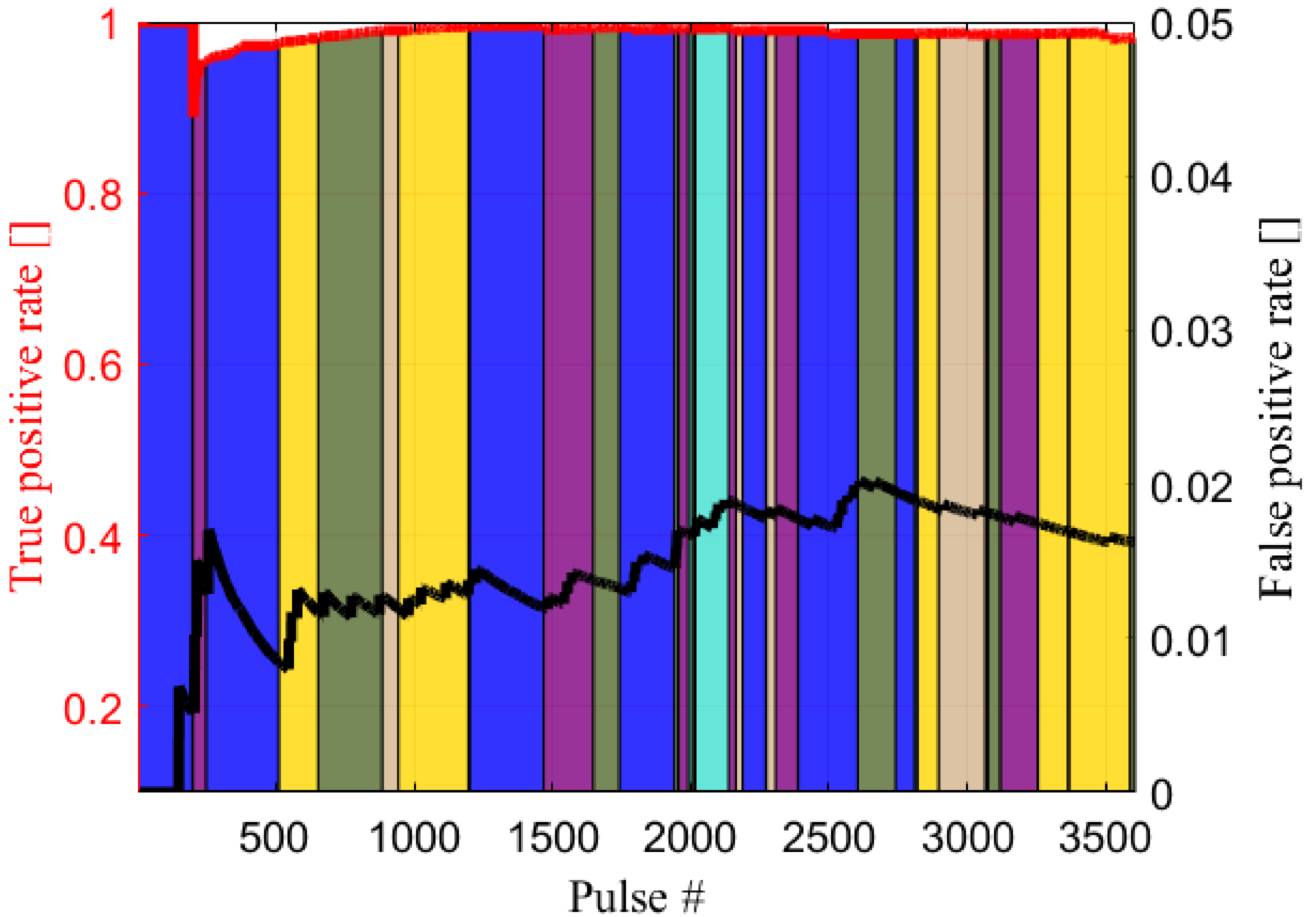
Four operational classes of disruptions:

Locked mode Density limit

Cooling edge Core radiation

Results for prediction

	Good	Missed	Early	Tardy	All D	False ND	False Alarms	All ND
Counts	576	10	1	0	587	47	48	3014
Percentage	98.13%	1.70%	0.17%	0.00%		1.56%	1.59%	

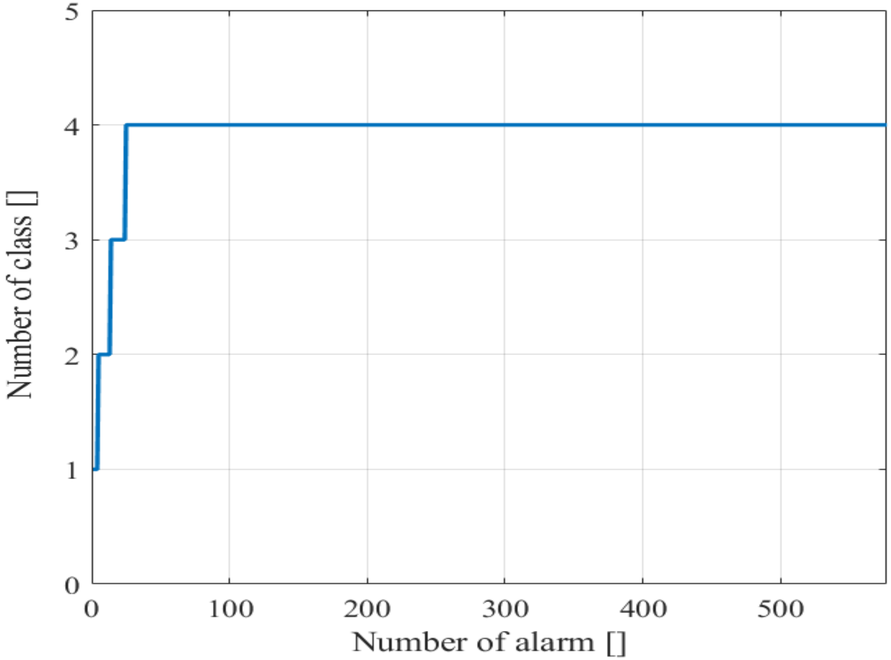


Success rate always above 90% and false alarms never much above 2%.

Statistics conservative.

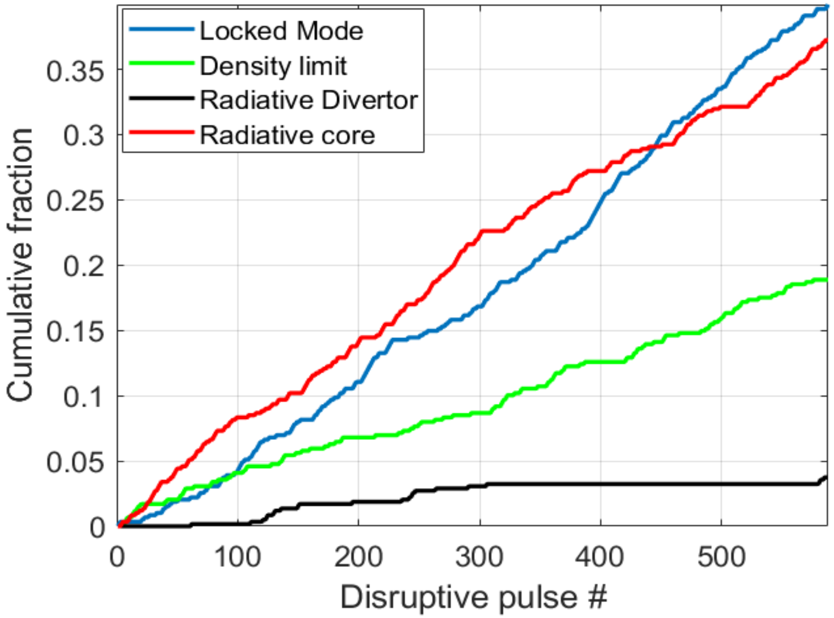
Results for Classification (K-means)

A disruption is attributed to the class of the first Instability Factor crossing the stability threshold.



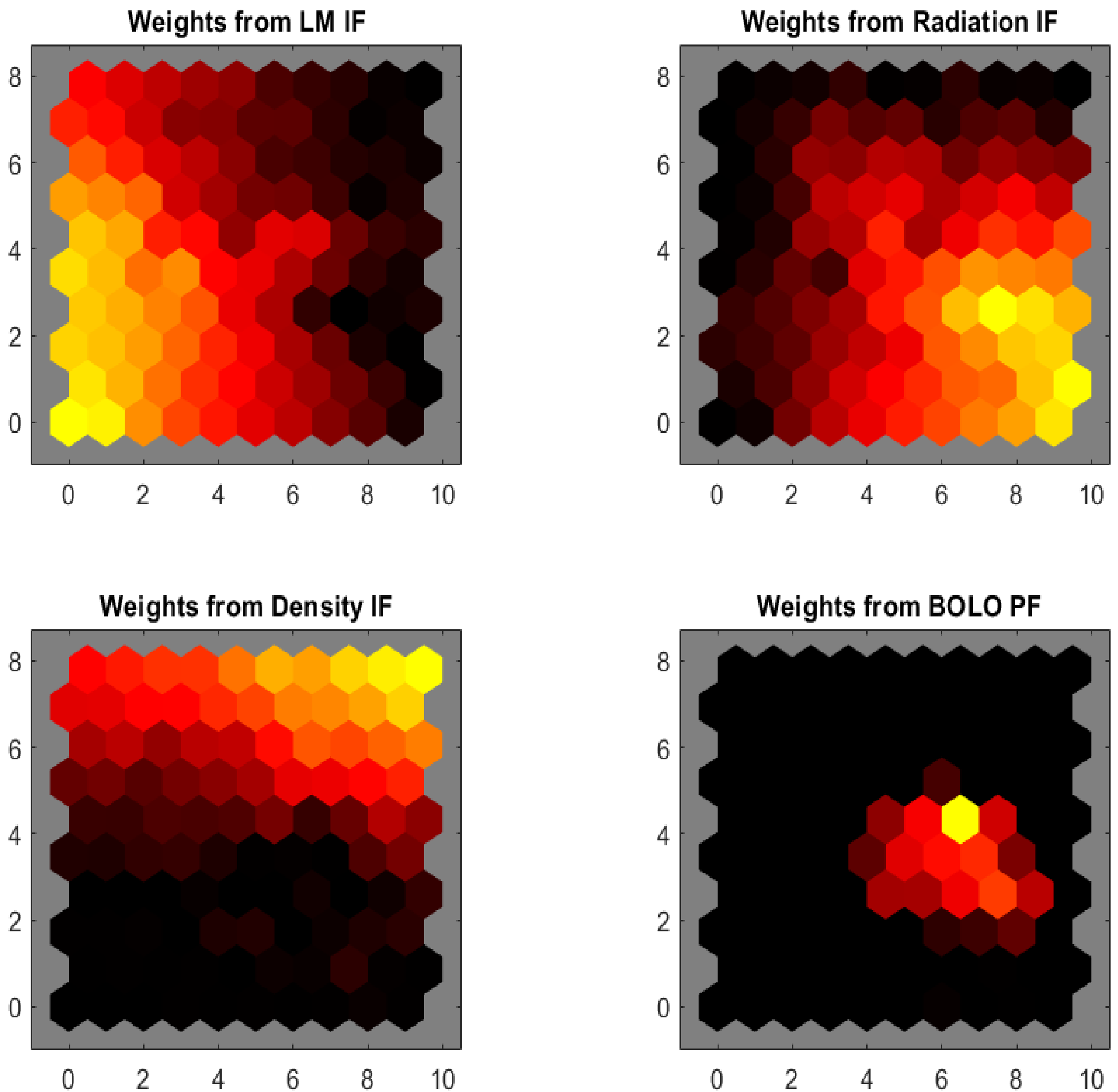
The unsupervised classifier converges rapidly to the four classes expected (in about 40 discharges).

The cumulative plot of the types of disruption



Good agreement with the expert classification

Results for classification (SOM)



- Even the weights of the SOM map identify clearly the four classes of disruption.
- This is a mathematically completely independent method.
- Operationally these disruption types are clearly distinguishable

Conclusions

- Adaptive and Transfer Learning are becoming important in many fields including Tokamak physics, particularly for disruption prediction.
- In addition to helping in solving many problems, Open-World learning also is forcing the community to revisit some ideas and concepts too acritically accepted.
- The technologies of Stacks of predictors, Ensembles of classifiers, SOMs and K-means have proved to be sufficiently flexible to implement complex strategies of adaptive learning and type classification.
- The developed techniques of adaptive/transfer learning have been quite successful in predicting and classifying disruptions on JET at the beginning of operation with the new ITER Like Wall. They have also maintained their performance for more recent campaigns.

Thanks for Your Attention!

QUESTIONS?