EFDA–JET–PR(14)08

G. Farias, J. Vega, S. Dormido-Canto, A. Murari, R. Moreno, H. Vargas,
A. Valencia and JET EFDA contributors

Prediction of the Time to Disruption in JET with an ITER-Like Wall

Prediction of the Time to Disruption in JET with an ITER-Like Wall

G. Farias[1], J. Vega[2], S. Dormido-Canto[3], A. Murari[4], R. Moreno[2], H. Vargas[1], A. Valencia[1] and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]*Pontificia Universidad Católica de Valparaiso, Valparaiso, Chile*
[2]*Laboratorio Nacional de Fusión. CIEMAT, Madrid, Spain*
[3]*Dpto. Informática y Automática – UNED, Madrid, Spain*
[4]*Associazione EURATOM/ENEA per la Fusione, Consorzio RFX, 4-35127 Padova, Italy*
** See annex of F. Romanelli et al, "Overview of JET Results", (24th IAEA Fusion Energy Conference, San Diego, USA (2012)).*

Preprint of Paper to be submitted for publication in
Nuclear Fusion

ABSTRACT

Avoidance and mitigation of disruptions are crucial problems in ITER and are becoming increasingly relevant at JET with the installation of the new ITER-Like Wall (ILW). But it is important to emphasize that disruption prediction is a pre-requisite to put into operation any avoidance or mitigation methodology. The design of disruption predictors mainly takes into account the achievement of predictions with high success rate, low false alarm rate and enough anticipation time. An important issue in the estimation of the time to the disruption is to set bounds to the prediction error. An accurate estimation would allow the set-up of active avoidance/mitigation actions in direct dependence to the time to disruption from the prediction. This article describes a new generation of disruption predictors with very accurate determination of the time to disruption. The predictor has been tested with 1237 JET discharges during the ITER-like wall campaigns. The results show success rates of 100%, false alarm rates close to 0%, enough anticipation time (> 128 ms) and high accuracy in the time to disruption estimation (in the order of ms).

1. INTRODUCTION

In the context of guaranteeing safe operation of large Tokamaks, *avoidance* and *mitigation* (A/M) of disruptions have particularly high priority. The first one indicates the objective of ensuring disruption free operation. The second one is aimed at the alleviation of the disruption detrimental effects when the disruptive event is unavoidable.

Disruption prediction is a pre-requisite to start any A/M technique. Typically, disruption predictors are qualified by their success rate, false alarm rate and average warning time (or anticipation time). Ideal predictors are the ones that simultaneously meet the following conditions: success rates close to 100% (ITER requirement is >95%), false alarm rates close to 0% (ITER requirement is <5%) and enough average warning time (AWT) to activate A/M actions.

Figure 1 represents the sequence of events in typical real-time A/M procedures. A disruption predictor triggers an alarm when a disruptive behaviour is detected and this trigger can be used to activate any A/M mechanism. Of course, the A/M action will be effective after a certain delay (reaction time) with respect to the alarm. This reaction time, that includes the firing of technical systems, should be as short as possible but it is inevitable in the real world. After this time elapses, the plasma can start to experience the effects of the A/M techniques. Obviously, the reaction time has to be shorter than the warning time.

In general, different mitigation actions can be envisaged to reduce forces, to mitigate heat loads during the thermal quench and to avoid runaway electrons. Examples of mitigation methods can be the injection of a significant amount of gases through fast valves [1, 2], killer pellets [3, 4] or Electron Cyclotron Resonance Heating injection [5, 6]. Depending on the disruption type and the time available before the disruption, different strategies can be much more desirable than others.

A/M methods can have different characteristic times and, therefore, they cannot be applied if the reaction time is longer than the warning time. This is a crucial issue for the *active selections* of A/M

actions. The term *active selection* means to be able to choose, in real-time, among several potential A/M methods after the recognition of a forthcoming disruption. This decision can be optimised only on the basis of reliable warning times. However, so far, practical disruption predictors trigger an alarm after detecting a disruptive behaviour but they do not provide the time to disruption. In this respect, the off-line analysis of the predictions provides an average warning time with usually a large standard deviation. Fig. 2 shows results with the JET real-time APODIS predictor [7] corresponding to the first four ITER-like Wall (ILW) campaigns (years 2011-2013). On average, the resulting warning time is very good (423 ms) but there is a large standard deviation (827 ms); this of course makes very difficult the choice of which type of A/M strategy to implement once an alarm has been raised.

This article deals with the design and development of a new generation of disruption predictors, whose essential novelty is to provide not only the recognition of a disruptive behaviour but also the time to disruption. This Disruption Time Predictor (DTP) follows a multi-tier architecture similar to APODIS and has been tested with JET discharges corresponding to the first three ILW campaigns (years 2011-2012).

Section 2 presents a summary of data-driven models for disruption prediction, discusses the prediction confidence to accomplish successful mitigation actions and establishes the need of developing reliable time to disruption predictors (TTDPs). Section 3 explains the work done to date on TTDPs. Section 4 defines the architecture of a new TTDP (the so-called DTP). Section 5 is devoted to describing a particular implementation of the DTP for the three first ILW campaigns of JET and section 6 shows the results. To conclude, section 7 discusses about the DTP, its results and future work on this kind of predictors.

2. PREDICTION CONFIDENCE ON SUCCESSFUL MITIGATION ACTIONS WITH PRESENT PREDICTORS

Good theoretical models of plasma evolution would be ideal to guide operations. However, the existing models and simulation tools have performances far from those that are needed to guarantee that disruptions are avoided [8]. Some problems of these models are: incomplete description of the plasma, strong assumptions and unphysical boundary conditions. To overcome this lack of theoretical input, data-driven models have been investigated in the past and have provided much better practical results. Their objective is to achieve reliable predictions of forthcoming disruptions with enough warning time.

Disruption data-driven models are based on automatic classification techniques. In general, the plasma behaviour is represented at regular time intervals by a pair $(\mathbf{x}, y)$. The first element is a feature vector $\mathbf{x} \in \mathbb{R}^d$, where d is the feature space dimension. The vector components are measurements of plasma quantities at the time instants of the measurements. The second element is a label that represents the plasma behaviour: *disruptive* or *non-disruptive*. In this article, the disruptive and

non-disruptive behaviours are respectively represented by the labels *+1* and *–1*. In mathematical notation, $y \in \{+1, -1\}$.

Typically, disruption predictors are developed in a supervised way. A large dataset of feature vectors $\mathbf{x}_j \in^{o\ d} j = 1,..., N$ (the larger N the better) with well-known corresponding labels $y_j \in^{o\ d} j = 1,..., N$ are used as training set. The pairs $(\mathbf{x}_j, y_j)$ are chosen from a database of past discharges. The most important criterion to select the training feature vectors is to ensure that they cover the widest operational space of the tokamak. The training set allows dividing the feature space into two regions separated by a decision function. This decision function is determined through the so-called training process. The decision function just constitutes the disruption predictor (i.e. the disruption data-driven model). Given a new example to classify $\mathbf{x}_{new} \in^{o\ d}$, the label $y_{new} \in \{+1, -1\}$ is assigned depending on the region of the feature space (in relation to the decision function) to which $\mathbf{x}_{new}$ belongs (fig. 3).

Disruption predictors have to work in real-time for the whole duration of the discharge. Feature vectors are generated during the discharge on a periodic basis, $\mathbf{x}_{kTp}, k = 1,...,$F,where F is the number of feature vectors in each particular discharge and T_P is the temporal resolution of the predictor. Each feature vector is used as input to the predictor and the output is the predicted label $y_{kTp} \in \{+1, -1\}$at time kT_P. If no disruptive behaviour is detected, the output will be always *–1*. A forthcoming disruption is recognized when the output is *+1*. When this happens, it is usual to say that the predictor *triggers an alarm*.

Feature vectors can be generated from different sets of relevant signals. The raw signals are processed during a maximum time T_P to transform the raw data into features of distinctive nature, *i.e.* to generate a feature vector $\mathbf{x}_{kTp}$. In this article, the temporal interval T_P needed to process the raw signals and to form a feature vector will be called *signal processing interval* (SPI). Obviously, the data processing time within the SPI has to be shorter than T_P.

Several supervised machine learning methodologies have been reported during the years to develop disruption predictors. Some examples are neural networks [9, 10], Support Vector Machines (SVM) [11, 7], fuzzy logic [12], regression trees [13], discriminant analysis [14], self-organizing maps [15], manifold learning [16] and multiple threshold combinations [17].

In general, the quality of a predictor is expressed in terms of the disruption success rate (DSR), *i.e.* a relation between true positive (TP) and false negative (FN) predictions, and the false alarm rate (FAR), i.e. a relationship between true negative (TN) and false positive (FP) predictions:

$$DSR = TP \Big/ (TP + FN) \tag{1}$$

$$FAR = FP \Big/ (FP + TN) \tag{2}$$

However, these rates are not sufficient by themselves to determine how good a real-time predictor is. Although a predictor has 100% success rate and 0% false alarm rate, it is useless if the warning

time is shorter than the reaction time to accomplish mitigation methods. Therefore, together with the previous rates, a predictor has to provide sufficient AWT. However, even if the AWT is larger than the minimum reaction time, it does not guarantee a confident implementation of A/W actions. Fig. 2 shows that on average, the alarms are triggered early enough. Nevertheless, the most important conclusion to extract from that figure is that 87.44% of the alarms take place 30ms in advance to the disruptions. Therefore, there is a confidence of 87.44% to apply successful A/M actions whose reaction time is less than 30ms. The confidence on having more time to implement successful mitigation actions with greater reaction time rapidly decreases. For example, A/M actions that would require a reaction time equal to the AWT, could only achieve a success confidence of 28.38%. In these conditions, it is therefore very difficult if not impossible to select the most appropriate A/M action once an alarm is triggered.

Therefore, to put into operation *active selection* of mitigation actions with a high level of confidence, a new generation of TTDPs are necessary. These disruption predictors have to be reliable enough to provide not only a high success rate and a low rate of false alarms but also *an accurate estimation of the time to disruption* when the alarm is triggered.

3. Discussion on previous time to disruption predictor proposals

This section describes three different works that deal with the problem of predicting the time to disruption in tokamaks.

The TTDP proposal from Pautasso et al [18] is based on a two-layer artificial neural network. The first layer contains 20 neurons and the network output (second layer) provides the time interval to the disruption. Therefore, the predictor is a temporal evolution signal whose value at each time instant is the time to disruption. The implementation of the artificial neural network in the ASDEX Upgrade (AUG) tokamak determined that the alarm threshold was 50 ms for 7.5 ms (3 consecutive feature vectors as the sampling period is 2.5 ms). In other words, if the predictor output was less or equal to 50 ms and this prediction was maintained for 7.5 ms, an alarm was triggered. The condition of maintaining the output below the alarm threshold for 7.5 ms allowed filtering false alarms. The value of 50 ms as the disruption alarm limit was chosen with the purpose of using an impurity pellet injector for disruption mitigation and to allow a margin to compensate for the error affecting the network prediction.

The feature vectors components ($\mathbf{x} \in^{o\ 13}$) were samples corresponding to temporal evolution signals and time derivatives. Feature vectors were formed on a periodic basis (2.5ms) and, therefore, the predictor output was obtained on this periodic basis. The off-line results reported in [18] with 65 disruptive discharges and 500 non-disruptive discharges showed that the success rate was 85% (55/65) and the false alarm rate was 1% (5/500). The predictor was tested on-line in open loop with 128 discharges (of which 28 disrupted). The success rate and the false alarm rate with disruptions of the same type that the ones used for training were 79% (22/28) and 7% (7/100) respectively.

A second work [19] related to a TTDP used a fuzzy framework only to achieve a suitable clustering

of the input space. The proposed predictor defined a complex structure of neural networks (NN). A first processing layer based on the radial basis function (RBF) NN scheme was basically used to decompose the original database into four parts (or subsets). The layer outputs were used to activate four multilayer perceptrons (MLP), trained exclusively on a subset of the original database. The output of the system was a single node linear layer that provided the estimated time to disruption (TTD) as well as an alarm when appropriated. As in [18], the predictor generated a temporal evolution signal (time period of 2.5 ms) and the alarm was triggered when the time to disruption was below 250ms.

The training database was made up of 62 disruptive discharges and the test set for off-line analysis had 46 disruptive shots. Also, an on-line test of the predictor was carried out in open loop, but the total number of discharges is not specified. The feature vector components ($\mathbf{x} \in {}^{o}{}^{24}$) were again samples of time series data and time derivatives. The probability of correctly activating an alarm taking into account the threshold of 250ms was 95% whereas the level of false alarms was 2% (it should be noted that the test database only contained disruptive shots and, therefore, a false alarm corresponded to an early detection of a disruption). In the predictor on-line version, a pre-alarm state was recognized when the TTD was in the range 50ms – 350ms. The alarm was triggered if the TTD was maintained in the range 250ms – 350ms for ten feature vectors. The use of these 10 additional feature vectors avoided false alarms.

A third work about predicting the time to disruption [20] was applied to the JET database and was also based on artificial neural networks. The best network configuration was composed of nine inputs, two hidden layers with six and five hidden neurons respectively, and one output. Only signals available in real-time were taken into account (this means that data relying on off-line equilibrium reconstruction or off-line processed data are not used). Twenty feature vectors ($\mathbf{x} \in {}^{o}{}^{9}$) per discharge with a sampling period of 20 ms were chosen. Eight of the components represented temporal evolution of physics quantities and the last one was a time derivative. The feature vectors with label *disruptive* corresponded to the features in the time interval $[t_D - 420ms, t_D - 40ms]$ of each disruptive discharge, where t_D was the disruption time. In [20] the disruption time was determined as the time when the current quench started. The feature vectors with label *non-disruptive* belonged to a randomly chosen time interval of 400ms from each non-disruptive discharge. The neural network output was a real number between 0 and 1 representing the risk of disruption. It was obtained every 20 ms (the time period of the input feature vectors). The network target for each disruptive discharge was a sigmoid in the time window of 400ms so as to represent a greater risk near the disruption.

The off-line analysis of the database included discharges with plasma current above 1.5MA, X-point configuration and flat-top plasma current. The training set consisted of 86 disruptive pulses and 400 non-disruptive ones, while the validation set consisted of 35 disruptive discharges and 246 non-disruptive ones. Finally, the test set had 62 and 132 disruptive and non-disruptive shots respectively. An alarm was triggered when the neural network output was above a certain threshold that was chosen minimizing a detection error function [20]. For disruptive pulses, the

neural network output was analyzed in the time window [t_D – 440ms, t_D – 100ms], where 100ms was a defined prediction interval before the disruption instant. The results with the test set showed no false alarms and a success rate of 83.9%. The remaining 16.1% were alarms that were raised late, *i.e.* in the interval [t_D – 100ms, t_D].

It is important to note that first the three above TTDPs were tested with a limited number of discharges and, secondly, not all types of disruptions were considered.

4. DTP architecture

As a first point, it should be emphasized that the TTDPs reviewed in section 3 were trained to produce temporal evolution signals of the time to disruption. When the TTD estimated is below a certain threshold, an assertion mechanism is used to filter false alarms and an alarm is raised. The DTP proposed in this article implements a different approach. Instead of generating a temporal evolution signal to predict the time to disruption and to trigger the alarm when special conditions are met, the objective is to predict a disruptive behavior and when the prediction is positive to simultaneously provide the time to disruption.

The DTP architecture is based on a two layer architecture like APODIS [7] (Fig.4) although many aspects are different and also some previous ideas described in [21] have been introduced. The first layer consists of a number of SVM classifiers (*SVM-Ci, i = 1, …, n*) trained to both identify disruptive behaviours and determine an expected time to disruption (ETTD).

The choice of SVM as base classifier is motivated by the fact that SVM is a very effective method for general purpose pattern recognition [22, 23]. In a few words, given a set of input vectors which belong to two different classes, SVM maps the inputs into a high dimensional feature space through some nonlinear mapping. An optimal separating hyper-plane is constructed in this feature space to minimize the risk of misclassification. The hyper-plane is determined by a subset of points of the two classes, called support vectors.

The difference among the n SVM classifiers of fig. 4 consists of their respective training processes, but this will be explained later. Due to the fact that the various classifiers can make different predictions about the plasma behaviour, it is necessary to combine all the predictions into a single one. This is the task of the second layer.

Figure 4 shows a feature vectorr $\mathbf{x}_t$ that is generated in the SPI corresponding to the interval $(t-T_p, t)$. This feature vector is used as input to all the first layer SVM classifiers. Each classifier, $SVM-C_i$, $i = 1,...,n$, generates as output a pair $(y_{t,i}, ETTD_i(t))$, where $y_{t,i} \in \{+1, -1\}$ is the label predicted for the feature vector $\mathbf{x}_t$ by the classifier i, and $ETTD_i(t)$ is its corresponding ETTD. It should be noted that $ETTD_i(t)$ only makes sense if $y_{t,i} = \{+1\}$, *i.e.* only when the classifier recognizes a disruptive behaviour. The combination of all pairs $(y_{t,i}, ETTD_i(t)), i = 1,...n,$ into a single one is accomplished in the second layer through a *decision function* $f_d(t) \equiv f_d((y_{t,i}, ETTD_i(t)),..., ((y_{t,n}, ETTD_n(t)) = (y_t, ETTD(t))$. Therefore, the *decision function* determines whether or not the combination of the first layer classifiers recognizes a disruptive behaviour at time t. The output of the *decision function*

is also a pair (y_t, *ETTD* (t)) that provides the label of the decision $y_t \in \{+1,-1\}$ together with the respective ETTD (if the prediction is '*disruptive*').

It is important to emphasise that different *decision functions* can be defined to combine the classifier outputs into a single one. Therefore, each particular DTP can implement its own decision function. The specific implementations have to deal with two aspects: the combination of labels into a single one and the combination of the respective ETTDs to provide a resultant one.

An issue of any disruption predictor can be to maintain a low rate of false alarms. This is very important because a predictor with high sensibility can result incompatible with standard operations of the tokamaks to achieve high performance plasmas. Therefore, in order to filter spurious alarms with the DTP of figure 4, a *filtering criterion* has to be introduced. From a conceptual point of view, the *filtering criterion* establishes to trigger an alarm only when N_C consecutive disruptive behaviours are detected (figure 5), *i.e.* when N_C consecutive labels verify $y_{t-(N_C-1)\cdot T_P} = ... = y_{t-T_P} = y_t = \{+1\}$. From a mathematical point of view, the *filtering criterion* about to trigger or not an alarm at time t can be determined as the pair $(y_t^{(N_C)}, ETTD^{(N_C)}(t))$ where

$$y_t^{(N_C)} = \begin{cases} \{+1\} \text{ iff } y_{t-(N_C-1)\cdot T_P} = ... = y_{t-T_P} = y_t = \{+1\} \\ \{-1\} \text{ otherwise} \end{cases}$$

and

$$ETTD^{(N_C)}(t) = ETTD(t)$$

It should be noted that fig. 4 corresponds to a DTP with N_C=1.

So far, the DTP architecture has been described at a high level but more specific details are required. In particular, it is necessary to explain both the rationale of choosing an architecture like the one shown in fig. 4 and the individual training procedure of the SVM classifiers. The former is accomplished in section 4.1 and the latter is carried out in section 4.2.

4.1. PHYSICS OF DISRUPTIONS AND ITS INFLUENCE ON THE DTP DESIGN

This section gives answers to two important questions in relation to fig. 4:

Q1: Why are needed a set of n classifiers that make predictions in parallel with the same feature vector $\boldsymbol{x}_t$?

Q2: How many classifiers have to be considered?

The response to Q1 is the central idea for the estimation of the time to disruption. It was mentioned in section 2 that the feature vectors are generated during a SPI whose time length is T_P. The objective of the SVM classifier C_1 is to predict a disruptive behaviour just for the next SPI. In other words, the expected time to disruption with C_1 for a feature vector generated at time t would be $t < ETTD_1(t) \leq t + T_P$. With regard to the classifier C_2, the aim is to predict a disruptive behaviour for 2 SPIs later at the latest, *i.e.* if the feature vector is generated at time t, the expected time to disruption

with C_2 would be $t < ETTD_2(t) \leq t + 2T_P$. In general, given a feature vector $\mathbf{x}_t$, the expected time to disruption with the classifier C_n would take place within the time interval of n SPIs, which means $t < ETTD_n(t) \leq t + nT_P$.

With this structure and assuming a number of 5 classifiers, table 1 reproduces the ideal outputs obtained with the different classifiers from a time t with a temporal resolution T_P in the case that a disruption will happen within the time interval $(t < 4T_P, t + 5T_P)$.

According to the central idea described above, a classifier can predict with an anticipation of several SPIs (for instance, the *SVM-*C_n classifier ideally predicts with an anticipation of n SPIs). However an individual SVM classifier is not enough to be used as a general predictor. Different types of disruptions are characterized by different particular precursors that show specific characteristic times. For example, the fastest disruptive events are expected to be predicted only by the classifier *SVM-*C_1. Other types of disruptions can be predicted well in advance by *SVM-*C_8, *SVM-*C_9 or so. Therefore, due to the existence of different characteristic times for the development of the disruptive event, several SVM classifiers have to be simultaneously considered to ensure an early prediction (the earlier the better) of any type of disruption. The existence of simultaneous SVM classifiers compels to use a *decision function* to give a single prediction at each time instant.

The answer to Q2 is not easy because there is not a threshold time from which disruption precursors appear. So, the use of a high n can be misleading because actually disruptive precursors can be absent n SPIs in advance and, therefore, the training of these classifiers can be ambiguous and non-reliable. Each particular application of the DTP will have to determine a number of classifiers.

4.2. SELECTION OF THE TRAINING DATASETS

This section describes the way in which the several classifiers can be trained to predict a disruptive behaviour 1, 2, …, n SPIs in advance. The selection of the training sets is very innovative and this selection process has been responsible of the success of the present TTDP.

Before describing the selection process of the training sets for the different SVM classifiers, it is necessary to describe the way in which the discharge signals are pre-processed. To this end, it should be emphasized that the main requirement is to have available a database of disruptive discharges with a high reliable estimation of the disruption time.

The pre-processing of discharges requires as a first step the normalization of the temporal evolution signals that are considered for the classifiers. The normalization process takes into account the whole database of past discharges and it is carried out to avoid that quantities that can differ by several orders of magnitude can have higher weights in the classification system. Typical normalization techniques are the ones that limit the signal amplitude to the range [0, 1] or the ones that transform the signals into time series data with mean 0 and variance 1.

The second step is the resampling of all needed signals to a common sampling period. Usually, this is accomplished by interpolation methods.

The third step is related to the pre-processing of disruptive discharges. The disruption time of

each discharge is the reference time to form SPIs backwards whose temporal length is T_P. Figure 6 shows the sampling times of the signals (T_S) after the resampling process. It is important to note that the disruption time remains outside of the first SPI (I_1). The several SPIs are numbered according to its location relative to the disruption time: I_1, I_2, I_3 and so on. As it was mentioned in section 2, each SPI is used to generate a feature vector $\mathbf{x} \in {}^{o}{}^{d}$ that condenses the plasma behaviour into *d* components. Each component is obtained after some kind of signal processing with selected temporal evolution signals.

To conclude the discharge pre-processing, the non-disruptive discharges are also split into SPIs T_P long, from the plasma breakdown to the plasma end. Again, a feature vector $\mathbf{x} \in {}^{o}{}^{d}$ per SPI is generated.

Once the feature vectors of the training discharges have been determined, it is necessary to describe how to select disruptive and non-disruptive feature vectors to train each individual SVM classifier. The selection process is the key aspect of this article. First of all, it is important to note that disruptive and non-disruptive discharges are not treated the same.

Top of Fig.7 shows a non-disruptive discharge with its resulting SPIs. As stated previously, a different training feature vector is generated in each SPI. However, they have been represented in the same way, $\mathbf{x}_{ND}$ where ND makes reference to its non-disruptive character.

At the bottom of Fig.7, the breakdown of a disruptive discharge into its SPIs is presented. As in the non-disruptive case, each SPI is represented by a training feature vector $\mathbf{x}_{-m}$, where the subindex is related to its corresponding SPI number ($\mathbf{x}_{-m}$ is the feature vector corresponding to SPI I_m). Therefore, the subindex is linked to the SPI location in relation to the disruption. The use of the negative sign is necessary to avoid confusion with the notation used in the article about feature vectors at time *t* ($\mathbf{x}_t$).

It is important to note that three types of SPIs have been considered for the disruptive discharges. The first type of SPIs (type I) consists of the closest ones to the disruption. As it is explained later, the corresponding feature vectors will be used sometimes with label '*disruptive*' and sometimes with label '*non-disruptive*'. The second type of SPIs (type II) defines feature vectors that can be always classified as '*non-disruptive*'. It is generally accepted that *far* from the disruption, the feature vectors do not show a disruptive behaviour. This is justified by the fact that disruption precursors have a limited anticipation time. This limit is blurred and, therefore, a gap can be present between the SPIs of type I and type II. The third type of SPIs (type III) is really far away from the disruption and its corresponding feature vectors are never considered for training purposes.

The training process of the SVM classifiers (*SVM-Ci, i = 1, …, n*) of a DTP (Fig.4) has been defined to use the same number of feature vectors per classifier. However, although this number is constant, the difference between classifiers appears in the number of disruptive and non-disruptive feature vectors.

Given N_d disruptive discharges and N_s safe (or non-disruptive) discharges for training purposes, the training set of each SVM classifier includes N_s features vectors from non-disruptive discharges

and N_{FV} + 1 feature vectors per disruptive discharge. On the one hand, the N_s feature vectors of class '*non-disruptive*' correspond to 1 feature vector randomly chosen from the SPIs of each safe discharge (top of Fig.7). In general, the training set of the several SVM classifiers will have different feature vectors from the safe discharges. On the other hand, there are N_{FV} type I (bottom of Fig.7) feature vectors per disruptive discharge and they are the same for the n classifiers ($x_{-m}, m = 1,2..., N_{FV}$) but have different labels as described below. Other feature vector is selected at random from the type II SPIs of a disruptive discharge (bottom of fig. 7) and its label is '*non-disruptive*'. The same vector is just included in the training set of all the classifiers

According to the description so far, N_{FV} + 1 feature vectors per disruptive discharge are exactly the same in the training set of all classifiers. The difference in the datasets resides in the labels (disruptive/non-disruptive) that are assigned to the feature vectors. Table 2 shows how the type I feature vector labels are defined per disruptive discharge for the several classifiers. The reason of this assignment is based on the objective that the classifier C_k has to predict a disruptive behaviour within a time interval of k SPIs from the prediction time. Therefore, for the classifier C_1, only the feature vector corresponding to the SPI I_1 is labelled as '*disruptive*' and all other feature vectors are labelled as '*non-disruptive*'. In the case of C_2, the feature vectors from SPIs I_1 and I_2 are considered '*disruptive*', whereas the remaining ones up to N_{FV} are marked as '*non-disruptive*'. The generalization of this reasoning is straightforward.

It is important to emphasize the innovative idea of using as either 'disruptive' or 'non-disruptive' the same feature vectors depending on the classifier. The objective is to look for specialised classifiers that are able to recognise disruptive features and to identify the remaining time to the disruption.

5. DEVELOPMENT OF A DTP FOR THE JET ILW CAMPAIGNS

A DTP for 1237 JET ILW discharges (201 disruptive and 1036 non-disruptive) between years 2011 and 2013 has been developed. Different training/test datasets can be generated with a discharge proportion 60%/40% respectively.

Table 3 shows the twelve plasma quantities that have been taken into consideration for the DTP. The nine first signals are JET real-time signals that are read from the JET database and have been resampled at 1 ksample/s. The remaining ones are computed from the previous quantities. The signal amplitudes have been normalized to the range [0, 1].

SPIs 32ms long are used to take into account that the minimum time in JET to perform mitigation actions is 30 ms. Also, in a similar way to APODIS, two different features per signal are obtained in each SPI. The first one is the signal mean value in the corresponding SPI and the second one is the standard deviation of the power spectrum (after removing the DC component) during the SPI.

In this way, the total number of features that are available for the DTP is 24.

In this article, a maximum number of ten SVM classifiers have been used for the DTP first layer: *SVM-Ci, i = 1, …, 10*. This value, according to section 4.1, would allow the prediction of disruptive behaviors 10 SPIs in advance, *i.e.* 320ms in advance.

Concerning the training of the classifiers, N_{FV} = 20 SPIs of type I samples per disruptive discharge have been considered, which means to cover a time interval of 640 ms before the disruptions. Type II SPIs are chosen to deal with a period of time between 1 s and 1.6 s before the disruptions. All of the SVM classifiers are trained with the same radial basis function (RBF) kernel [23]

$$H(\mathbf{x},\mathbf{x}') = \exp\left\{-\frac{\left|\mathbf{x}-\mathbf{x}'\right|^2}{\sigma^2}\right\}.$$

Several values of the kernel parameter have been tested $\sigma = \{10^{-5},\ 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and the best results are obtained with $\sigma = 10^{-1}$. Moreover, all the SVM classifiers are trained with the same regularization constant [23] in their respective optimization processes. Due to the fact that all the software has been written in Matlab (http://www.mathworks.com), the functions related to SVM that are included in the Matlab Statistical Toolbox have been used. With regard to the regularization constant, the default value provided by Matlab has been used.

As N_{FV} is 20, 21 feature vectors per disruptive discharge are considered. Taking into account that with 1237 discharges (which include 201 disruptive and 1036 safe) a training set uses 60% of the available ones (121 disruptive shots and 622 safe), it implies that the number of feature vectors to train each classifier is $121*(N_{FV}+1)+ N_s = 3163$.

The DTP has been tested with 40% (80 disruptive and 414 non-disruptive) of the 1237 available discharges by simulating the real-time data acquisition and signal processing. The DTP decision function that has been implemented establishes:

$$f_d(t) \equiv f_d\left(\left(y_{t,1}, ETTD_1(t)\right),...,\left(y_{t,n}, ETTD_{10}(t)\right)\right) = \left(y_t, ETTD(t)\right)$$

where

$$y_t = \begin{cases} \{+1\} \text{ iff } \exists i, i = 1,...10, / y_{t,i} = \{+1\} \\ \{-1\} \text{ otherwise} \end{cases}$$

If $y_t = \{+1\}$ then

$$ETTD(t) \in \left(t + (k-1)\cdot T_P,\ t + k\cdot T_P\right] \tag{3}$$

where

$$k = \min(j), \mathrm{j} = 1,...10 / \mathrm{y}_{t,j} = \{+1\}$$

Therefore, this decision function is aimed at recognizing a disruptive behaviour at time t (fig. 4) when anyone of the 10 first layer classifiers identifies a disruptive event. But it is important to mention that more than one classifier can make a disruptive prediction. If this happens, the present

decision function gives an ETTD that is related to the classifier with smallest index (the one closer to the disruption time), which predicts the disruptive behaviour.

At this point, a couple of questions arise: why the smallest identifier? What does eq. mean?

Concerning the first question, it is important to take into account that the smaller identifier the lesser number of SPIs is predicted for the disruption. Therefore, this is the most restrictive criterion to put into operation A/M actions and it provides a very conservative option for these purposes. With regard to the meaning of eq. , it can be explained with an example. Let's assume that only three of the classifiers recognize a forthcoming disruption, for instance $y_{t,9} = y_{t,7} = y_{t,4} = \{+1\}$. The smallest index corresponds to the classifier *SVM-*C_4. Figure 8 shows that *SVM-*C_4 provides the following ETTD: $t < ETTD_4(t) \leq t = 4 \cdot T_P$. However, *SVM-*$C_1$, *SVM-*$C_2$ and *SVM-*C_3 do not recognize a disruptive behaviour for the next 3 SPIs. A simple combination of this information means to choose as ETTD the interval $t + 3 \cdot T_P$. Eq. just generalises this reasoning.

Finally, it is important to mention that the TTD assigned to a DTP is defined as the remaining time to the upper bound of the ETTD interval, *i.e.* if the TTD: $ETTD\,(t) \in (t + (k-1) \cdot T_P, t + k \cdot T_P)$, the corresponding $TTD(t) = k \cdot T_P$.

6. JET DTP RESULTS

As mentioned previously, the discharges of the training set are chosen in a random way from the 1237 JET ILW shots between years 2011 and 2013. After selecting a training set (622 non-disruptive and 121 disruptive discharges), the remaining ones (414 non-disruptive and 80 disruptive) are used as test set to obtain the results. To take into account the potential influence of the random selection of discharges, 100 different datasets of trainining/test sets have been generated: DS(1), DS(2), … DS(100). In other words, each dataset of training/test sets DS(m), m = 1, …, 100 obtains a disruption success rate (DSR(m)), a false alarm rate (FAR(m)) and time to disruption prediction (TTD(m)). The final results of a DTP are computed according to the following expressions:

$$DSR = \frac{\sum_{m=1}^{100} DSR(m)}{100}$$
$$FAR = \frac{\sum_{m=1}^{100} FAR(m)}{100} \qquad (4)$$
$$TTD = \frac{\sum_{m=1}^{100} TTD(m)}{100}$$

The error in the estimation of the TTD is the standard deviation of the elements TTD(m), m = 1, …, 100.

The DTP assessments have taken into account feature vectors of dimension 14 ($\mathbf{x}_t^{\,o\,14}$) and 24 ($\mathbf{x}_t^{\,o\,14}$). The first case corresponds to use the signals 1-7 of table 3 and two features per signal in each

SPI 32 ms long: the mean value and the standard deviation of the power spectrum after removing the DC component. These 14 features are the ones used by APODIS in the JET real-time network and, therefore, APODIS results can be compared with the results of the DTP. The case with 24 features uses all the signals of table 3 and two features per signal as in the preceding case. Different number of classifiers in the first layer has been evaluated.

It should be reminded that each test set contains 414 non-disruptive and 80 disruptive discharges. In all cases, the test discharges are analysed every 32 ms, from plasma breakdown to end. A discharge is recognised as non-disruptive when all feature vectors of the discharge (taken at regular times of 32 ms) are identified as non-disruptive. A discharge is labelled as disruptive when one alarm is triggered, *i.e.* when the DTP predicts N_C consecutive disruptive behaviours. A discharge is considered as missed alarm with it is disruptive but no alarm is triggered. Finally, a discharge is classified as false alarm when it is non-disruptive but the DTP predictor has triggered an alarm.

Due to the fact that the test sets consist of 414 non-disruptive discharges, each 1% of false alarm rate approximately means 4 false alarms. On the other hand, owing to the use of test sets with 80 disruptive discharges, table 4 shows, on average, the number of missed alarms as a function of the DSR.

Table 5 summarizes the outcomes with $N_C = 1$, *i.e.* the first disruptive behaviour is enough to trigger an alarm. Up to 10 sequential classifiers in the first layer have been tested. The DSR is above 97.9% in all cases, which means that practically all disruptive discharges are successfully predicted. With regard to the false alarms, the more classifiers in the first layer the higher FAR. A DTP with 24 features reduces this impact but some mechanism to filter false alarms is necessary.

Table 5 shows that up to three classifiers can be combined for successful predictions with FAR below 5%. But it should be noted that the ETTD with three classifiers (taking into account that processing windows of 32 ms are being used) vary between (64, 96] and (0, 32] ms. Therefore, to improve not only the ETTD but also to diminish the FAR, the filtering criterion with $N_C > 1$ (fig. 5) has been applied to the same 100 datasets of training/test sets: DS(m), m = 1, …, 100.

Table 6 shows the results with different number of classifiers in the first layer, several cases of consecutive alarms (N_C parameter) and both 14 and 24 features. Each line of the table shows a different DTP that differs from the others in the first layer classifiers and/or the N_C parameter. All possible combinations of first layer classifiers (between 10 and 5) and N_C have been tested. However, only the cases in which the FAR is below 5% (either with 14 or 24 features) are shown. The values DSR, FAR and TTD are defined in eq. . Finally, min(TTD) represents the smallest element TTD(m), m = 1, …, 100 of a DTP.

As it is shown in table 6, the DSR is quite close to 100% in all cases. In fact, the minimum DSR with 14 features is 98.2% (this means, according to table 4, to miss 2 disruptions) and 97.1% in the case of 24 features (3 missed alarms). On average, there is about 3% (12 discharges) of false alarms in the DTPs with 14 features and about 2% (8 discharges in the case of 24 features.

The columns called TTD represent the average time to disruption obtained with the predictions.

Elaborating on the false alarms, it is obvious that the rates decrease as N_C increases. This means that N_C, as expected, can be used to filter false alarms. Demanding the prediction of several consecutive disruptive behaviours before triggering an alarm makes a DTP both robust (the DSR is maintained) and insensitive to noise (false alarms are filtered).

With regard to the TTD, it should be noted that all the predictions are compatible with the ETTD. Moreover, it is important to emphasize that the minimum TTD in all cases is always greater than or equal to the time length of one SPI. This is essential to ensure that all predictions have at least 32ms to carry out A/M actions. It is important to mention that the minimum time in JET to carry out mitigation actions is 30 ms. Therefore, the DTP results do not show tardy detections.

As mentioned, each row of table 6 shows the average results over 100 datasets of training/test sets. The results show high DSR, low FAR and accurate predictions of the time to disruptions. The outcomes, which are the average on a high number of training/test sets, prove the robustness of the DTP and confirm that there is no dependence on a specific selection of discharges. This means, that the training/test datasets can be chosen at random because all of them provide similar results.

To put into operation a real-time DTP, instead of computing results with 100 different predictors, it is necessary to select a particular training set and to assess the corresponding DTP with a test set. Table 7 shows the results with the best predictors of each row in table 6. But it is important to clarify the term 'the best predictors of each row'. To obtain the best of the 100 predictors in each row, three filters have been used. The first one is to select the predictor whose test set provides the smallest standard deviation in the computation of the TTD. In case of similar values, the second filter establishes to choose the predictor that provides the smallest FAR. If a third filter is necessary, the predictor with the greatest DSR is selected.

Table 7 practically shows 100% of success rate in all the cases (14 and 24 features). The false alarm rate is almost 0% (on average, less than 1 discharge). The TTDs are compatibles with the ETTDs and it is important to note the high accuracy of the predictions (in some cases, the error estimation is below 1 ms). Moreover, in all the cases, the TTD definition (according to the largest possible time within the ETTD) is satisfactory. For practical purposes, DTPs with 14 features are preferred in comparison to the ones with 24 features. The former essentially provide the same results and are less sensible to signal failures (only 7 signals are required).

According to table 7, a DTP with seven classifiers ($C_1, \ldots, C_7$), $N_C = 3$ and 14 features provides the maximum TTD with the minimum FAR and 100% of DSR. Figure 9 shows the accumulated fraction of detected disruptions with this DTP. It is important to note that 1, 75 and 4 predictions show TTDs of 192ms, 160ms and 128 ms respectively. All anticipation times are very concentrated around a single value, which makes the DTP very accurate. This behaviour is the standard one for all the DPTs of table 7. Also, it is important to emphasise that all the TTDs are far enough from the time threshold of 30ms to accomplish mitigation actions in JET.

In the same figure, the real warning times at the time instants when the DTP triggers the alarm have been plotted. The AWT is 161ms and the estimation error is ±4 ms. It is clear from the plot

that both the TTD and the real warning times coincide in 95% (76/80) of the cases. But it should be noted that the discrepancies predict a shorter time to disruption than the real one. A shorter TTD is not an issue because if A/M actions are carried out by considering a shorter time, the disruptions will not be missed.

In general, in all the DTPs of table 7, the TTD and the DTP warning times match in 95% of the cases, and the disagreements show shorter TTDs.

Fig. 10 compares the real results of APODIS obtained with the same discharges of Fig.9. The success rate without considering tardy detections is 72.5% (58/80) and the false alarm rate is 0.5% (2/414). The AWT is 289ms and its range of variation is from 30 ms to 1.835 s (which implies a standard deviation of 332ms). These results are quite different from the ones obtained with the DTP. On the one hand, 100% of the disruptions are predicted with enough anticipation time. On the other hand, there is a large accuracy in the TTD predictions that according to fig. 9 means a large accuracy in the warning times.

7. DISCUSSION AND FUTURE WORK ON DTPS

The DTP described in this article is a kind of time to disruption predictor that successfully copes with three essential aspects of disruption prediction: high success rate and low false alarm rate together with an early and very accurate estimation of the time to disruption (table 7 and Fig.9). The tests carried out in JET with 1237 discharges show much better results of the DTP than the ones obtained with the JET APODIS predictor (Fig.10). These results include 100% of DSR, FAR near to 0%, enough anticipation time (> 128ms) and high accuracy in the TTD estimation (in the order of ms). As part of future work, the DTP has to be tested with the large database of JET corresponding to the previous Carbon Fiber Composite (CFC) wall. Also, an important aspect to investigate is the possibility of increasing the TTD estimation but maintaining the accuracy. To this end, the analysis of new decision functions, different interval lengths (T_P) of the SPIs, effect of increasing the number of 1st layer classifiers and implications of varying the number of type I and type II SPIs will have to be carried out.

The so accurate determination of the TTD will allow including the time as a new variable to search precursors of disruptive behaviours. From tables 6 and 7, it is clear that even the fastest disruptions in JET (for instance, vertical stabilization) could be predicted with at least 32ms. This means that the footprint of the disruptive event would be present and, therefore, a data mining analysis of the discharges can be crucial to establish physics models about disruptions. But it is important to note that during the ILW campaigns, the plasma scenarios that have been investigated did not produce a large number of fast disruptive events. So, the analysis of CFC wall discharges and the ones that will be produced in the next ILW experimental campaigns will be of primary importance in this respect.

Also, it is important to note the robustness of the predictor even with a reduced number of signals. Here, the term robustness means that there is no dependence on specific discharges (table 6) to train the DTP whenever a large number of discharges are available for training purposes. However,

recent analyses have raised the need of developing reliable disruption predictors from scratch [24, 25, 26] for next fusion devices such as ITER or DEMO. Therefore, an interesting research topic will be the development of a DTP from scratch.

Other possibilities that can be explored in the future are the use of new plasma quantities and the introduction of probabilistic classifiers to create the DTP.

To conclude, it is important to emphasize that the results of these article give the unprecedented opportunity of designing active A/M actions to be included in the development of plasma scenarios.

ACKNOWLEDGMENTS

This work was partially funded by the Chilean Ministry of Education under the Project FONDECYT 11121590. This work was partially funded by the Spanish Ministry of Economy and Competitiveness under the Projects No ENE2012-38970-C04-01 and ENE2012-38970-C04-03. This work, supported by the European Communities under the contract of Association between EURATOM/CIEMAT, was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES

[1]. M. Lehnen, A. Alonso, G. Arnoux, N. Baumgarten, S.A. Bozhenkov, S. Brezinsek, M. Brix, T. Eich, S.N. Gerasimov, A. Huber, S. Jachmich, U. Kruezi, P.D. Morgan, V.V. Plyusnin, C. Reux, V. Riccardo, G. Sergienko, M.F. Stamp and JET EFDA contributors. "Disruption mitigation by massive gas injection in JET". Nuclear Fusion **51** (2011) 123010 (12pp).

[2]. M. Bakhtiari, G. Olynyk, R. Granetz, D.G. Whyte, M.L. Reinke, K. Zhurovich and V. Izzo. "Using mixed gases for massive gas injection disruption mitigation on Alcator C-Mod". Nuclear Fusion **51** (2011) 063007 (9pp).

[3]. G. Pautasso, K. Buchl, J.C. Fuchs, O. Gruber, A. Herrmann, K. Lackner, P.T. Lang, K.F. Mast, M. Ulrich, H. Zohm and ASDEX Upgrade Team. "Use of impurity pellets to control energy dissipation during disruption". Nuclear Fusion, **36**, 10 (1996) 1291-1297.

[4]. N. Commaux, L.R. Baylor, S.K. Combs, N.W. Eidietis, T.E. Evans, C.R. Foust, E.M. Hollmann, D.A. Humphreys, V.A. Izzo, A. N. James, T.C. Jernigan, S. J. Meitner, P. B. Parks, J.C.Wesley and J. H. Yu. "Novel rapid shutdown strategies for runaway electron suppression in DIII-D". Nuclear Fusion **51** (2011) 103001 (9pp).

[5]. B. Esposito, G. Granucci, P. Smeulders, S. Nowak, J. R. Martín-Solís, L. Gabellieri, FTU and ECRH teams. "Disruption Avoidance in the Frascati Tokamak Upgrade by Means of Magnetohydrodynamic Mode Stabilization Using Electron-Cyclotron-Resonance Heating". Phys. Rev. Lett. 100, 045006 (2008).

[6]. B. Esposito, G. Granucci, M. Maraschek, S. Nowak, A. Gude, V. Igochine, E. Lazzaro, R. McDermott, E. Poli, J. Stober, W. Suttrop, W. Treutterer, H. Zohm, D. Brunetti and ASDEX Upgrade Team. "Avoidance of disruptions at high β_N in ASDEX Upgrade with off-axis ECRH". Nuclear Fusion **51** (2011) 083051 (9pp).

[7]. J. Vega, S. Dormido-Canto, J.M. López, A. Murari, J. M. Ramírez, R. Moreno, M. Ruiz, D. Alves, R. Felton and JET-EFDA Contributors. "Results of the JET real-time disruption predictor in the ITER-like wall campaigns", Fusion Engineering and Design **88** (2013) 1228-1231.

[8]. A.H. Boozer. Physics of Plasmas. **19**, 058101 (2012).

[9]. C. G.Windsor, G. Pautasso, C. Tichmann, R.J. Buttery, T.C. Hender, JET-EFDA Contributors and the ASDEX Upgrade Team. "A cross-tokamak neural network disruption predictor for the JET and ASDEX Upgrade tokamaks". Nuclear Fusion **45** (2005) 337–350.

[10]. B. Cannas, A. Fanni, G. Pautasso, G. Sias and P. Sonato. "An adaptive real-time disruption predictor for ASDEX Upgrade". Nuclear Fusion **50** (2010) 075004 (12pp).

[11]. G.A. Rattá, J. Vega, A. Murari, G. Vagliasindi, M. F. Johnson, P. C. de Vries and JET-EFDA Contributors. "An Advanced Disruption Predictor for JET tested in a simulated Real Time Environment" Nuclear Fusion. **50** (2010) 025005 (10pp).

[12]. G. Vagliasindi, A. Murari, P. Arena, L. Fortuna, M. Johnson, D. Howell and JET-EFDA Contributors. "A Disruption Predictor Based on Fuzzy Logic Applied to JET Database". IEEE Transactions on Plasma Science. **36** (1) 2008 (253-262).

[13]. A. Murari, G. Vagliasindi, P. Arena, L. Fortuna, O. Barana, M. Johnson and JET-EFDA Contributors. "Prototype of an adaptive disruption predictor for JET based on fuzzy logic and regression trees". Nuclear Fusion **48** (2008) 035010 (10pp).

[14]. Y. Zhang, G. Pautasso, O. Kardaun, G. Tardini, X.D. Zhang and the ASDEX Upgrade Team. "Prediction of disruptions on ASDEX Upgrade using discriminant analysis". Nuclear Fusion **51** (2011) 063039 (12pp).

[15]. B. Cannas, A. Fanni, G. Pautasso, G. Sias, the ASDEX Upgrade Team. "Disruption prediction with adaptive neural networks for ASDEX Upgrade". Fusion Engineering and Design **86** (2011) 1039–1044.

[16]. B. Cannas, A. Fanni, A. Murari, A. Pau, G. Sias and JET-EFDA Contributors. "Manifold learning to interpret JET high-dimensional operational space". Plasma Phys. Control. Fusion **55** (2013) 045006 (11pp).

[17]. S.P. Gerhardt, D.S. Darrow, R.E. Bell, B.P. LeBlanc, J.E. Menard, D. Mueller, A. L. Roquemore, S.A. Sabbagh and H. Yuh. Nuclear Fusion **53** (2013) 063021 (19pp).

[18]. G. Pautasso, C. Tichmann, S. Egorov, T. Zehetbauer, O. Gruber, M. Maraschek, K.-F. Mast, V. Mertens, I. Perchermeier, G. Raupp,W. Treutterer, C.G.Windsor, ASDEX Upgrade Team. "On-line prediction and mitigation of disruptions in ASDEX Upgrade". Nuclear Fusion **42** (2002) 100-108.

[19]. F.C. Morabito, M. Versaci, G. Pautasso, C. Tichmann, ASDEX Upgrade Team. "Fuzzy-neural approaches to the prediction of disruptions in ASDEX Upgrade". Nuclear Fusion **41**, 11 (2001) 1715-1723.

[20]. B. Cannas, A. Fanni, E. Marongiu, P. Sonato. "Disruption forecasting at JET using neural networks". Nuclear Fusion **44** (2004) 68-76.

[21]. G. Farias, J. Vega, S. González, A. Pereira, X. Lee, D. Schissel, P. Gohil. "Automatic determination of L/H transition times in DIII-D through a collaborative distributed environment", Fusion Engineering and Design, Volume **87**(2012) 2081-2083.

[22]. V. Vapnik. "The Nature of Statistical Learning Theory". Second edition. Springer. (1999).

[23]. V. Cherkassky, F. Mulier. "Learning from data". John Wiley & Sons, Inc. (1998).

[24]. S. Dormido-Canto, J. Vega, J.M. Ramírez, A. Murari, R. Moreno, J.M. López, A. Pereira and JET-EFDA Contributors. "Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER". Nuclear Fusion.**53** (2013) 113001 (8pp).

[25]. R. Aledda, B. Cannas, A. Fanni, G. Sias, G. Pautasso and ASDEX Upgrade Team. "Multivariate statistical models for disruption prediction at ASDEX Upgrade". Fusion Engineering and Design **88** (2013) 1297-1301.

[26]. J. Vega, A. Murari, S. Dormido-Canto, R. Moreno, A. Pereira, A. Acero. "Adaptive high learning rate probabilistic disruption predictors from scratch for the next generation of tokamaks". Submitted to Nuclear Fusion. (2014).

Classifier	Classifier outputs at the prediction times				
	t	$t + T_P$	$t + 2T_P$	$t + 3T_P$	$t + 4T_P$
SVM-C5	({+1}, ≤5 SPIs)	({+1}, ≤5 SPIs)	({+1}, ≤5 SPIs)	({+1}, ≤5 SPIs)	({+1}, ≤5 SPIs)
SVM-C4	({-1}, -)	({+1}, ≤4 SPIs)	({+1}, ≤4 SPIs)	({+1}, ≤4 SPIs)	({+1}, ≤4 SPIs)
SVM-C3	({-1}, -)	({-1}, -)	({+1}, ≤3 SPIs)	({+1}, ≤3 SPIs)	({+1}, ≤3 SPIs)
SVM-C2	({-1}, -)	({-1}, -)	({-1}, -)	({+1}, ≤2 SPIs)	({+1}, ≤2 SPIs)
SVM-C1	({-1}, -)	({-1}, -)	({-1}, -)	({-1}, -)	({+1}, ≤1 SPI)

Table 1: The table shows the pairs (y_t, $ETTD_t(t)$) provided by the classifiers at the several prediction times. The first element in the pair is the predicted label and the second one is expressed as the expected number of SPIs from the prediction time. When the predicted label is {–1}, no ETTD is output.

Classifier	Labels					
	$\mathbf{x}_{-1}$	$\mathbf{x}_{-2}$	$\mathbf{x}_{-3}$	$\mathbf{x}_{-4}$	$\mathbf{x}_{-5}$	…
SVM-C1	+1	-1	-1	-1	-1	…
SVM-C2	+1	+1	-1	-1	-1	…
SVM-C3	+1	+1	+1	-1	-1	…
SVM-C4	+1	+1	+1	+1	-1	…
SVM-C5	+1	+1	+1	+1	+1	…
…	…	…	…	…	…	…

Table 2: Labels of the feature vectors from type I SPIs of a disruptive discharge.

Signal id.	Signal name	Units
1	Plasma current	A
2	Mode locked amplitude	T
3	Plasma internal inductance	
4	Plasma density	m^{-3}
5	Stored diamagnetic energy time derivative	W
6	Radiated power	W
7	Total input power	W
8	Poloidal beta	
9	Plasma vertical centroid position	m
10	Plasma inductance time derivative	s^{-1}
11	Poloidal beta time derivative	s^{-1}
12	Plasma vertical centroid position time derivative	m/s

Table 3: List of signals to characterize the disruptive/non-disruptive status of JET plasmas with a DTP.

DSR(%)	Missed alarms
100	0
$98.75 \leq DSR < 100$	1
$97.50 \leq DSR < 98.75$	2
$96.25 \leq DSR < 97.50$	3

Table 4: Average number of missed alarms in the test sets deduced from the DSR.

	14 features		24 features	
1st layer classifiers	DSR(%)	FAR(%)	DSR(%)	FAR(%)
C1	100	0.00	97.93	0.00
C2, C1	100	0.06	98.40	0.06
C3, C2, C1	100	1.38	98.51	0.84
C4, …, C1	100	8.49	99.28	6.36
C5, …, C1	100	11.68	99.29	8.60
C6, …, C1	100	13.22	99.40	10.24
C7, …, C1	100	16.99	99.49	10.93
C8, …, C1	100	21.05	99.68	12.16
C9, …, C1	100	29.79	99.78	16.21
C10, …, C1	100	41.86	99.83	19.74

Table 5: The DTP triggers an alarm when any first layer classifier predicts a disruptive behaviour. DSR and FAR are defined in eq. and respectively. Results are shown with 14 and 24 features and correspond to the average from 100 datasets: DS1, DS2, ..., DS100. For simplicity, the names of the classifiers in the first column do not show "SVM-".

			14 features				24 features			
1st layer classifiers	N_C	ETTD (ms)	DSR(%)	FAR(%)	TTD (ms)	min(TTD) (ms)	DSR(%)	FAR(%)	TTD (ms)	Min(TTD) (ms)
C10, …, C1	8	(64, 96]	99.4	6.3	97±16	32	98.4	4.3	97±15	32
C10, …, C1	9	(32, 64]	99.4	4.7	65±11	32	98.1	2.9	65±12	32
C10, …, C1	10	(0, 32]	98.2	3.3	33±10	32	97.1	1.7	33±14	32
C9, …, C1	7	(64, 96]	99.9	5.5	97±14	32	98.5	4.0	97±14	32
C9, …, C1	8	(32, 64]	99.3	4.2	66±14	32	98.1	3.1	66±17	64
C9, …, C1	9	(0, 32]	99.2	2.8	33±8	32	97.6	1.9	33±12	32
C8, …, C1	5	(96, 128]	100	5.2	127±10	32	99.1	4.1	127±7	64
C8, …, C1	6	(64, 96]	99.9	3.3	97±11	32	98.5	2.5	96±10	32
C8, …, C1	7	(32, 64]	99.9	2.3	65±9	32	98.4	1.5	65±11	32
C8, …, C1	8	(0, 32]	98.6	1.4	34±8	32	97.5	0.9	33±12	32
C7, …, C1	3	(128, 160]	100	7.2	158±10	96	99.1	4.7	159±7	96
C7, …, C1	4	(96, 128]	100	5.2	127±10	32	99.1	2.9	127±6	64
C7, …, C1	5	(64, 96]	100	3.1	96±8	32	99.1	1.5	96±8	64
C7, …, C1	6	(32, 64]	99.9	1.5	64±6	32	98.4	0.7	64±8	32
C7, …, C1	7	(0, 32]	99.8	0.8	33±6	32	97.7	0.2	33±8	32
C6, …, C1	2	(128, 160]	100	6.5	158±10	32	99.1	4.8	158±7	96
C6, …, C1	3	(96, 128]	100	2.5	127±8	32	99.1	1.8	127±6	64
C6, …, C1	4	(64, 96]	100	0.8	95±5	32	98.8	0.7	96±6	32
C6, …, C1	5	(32, 64]	100	0.2	64±4	32	98.2	0.1	64±8	32
C6, …, C1	6	(0, 32]	99.9	0.1	32±5	32	97.7	0.0	32±8	32
C5, …, C1	2	(96, 128]	100	4.8	127±8	32	99.1	2.5	127±5	64
C5, …, C1	3	(64, 96]	100	1.7	96±7	64	98.3	0.6	96±6	64
C5, …, C1	4	(32, 64]	100	0.5	65±8	32	98.2	0.1	65±8	32
C5, …, C1	5	(0, 32]	99.5	0.1	33±8	32	97.6	0.0	33±11	32

Table 6: at least 2 consecutive disruptive behaviours are necessary to trigger an alarm. Again, DSR and FAR are defined in eq. and respectively. ETTD is the expected time to disruption interval.

			14 features			24 features		
1st layer classifiers	N_C	ETTD (ms)	DSR(%)	FAR(%)	TTD (ms)	DSR(%)	FAR(%)	TTD (ms)
C10, …, C1	8	(64, 96]	100	0.7	95±7	98.7	0.2	95±3
C10, …, C1	9	(32, 64]	100	0.4	64±0	98.7	0.2	64±0
C10, …, C1	10	(0, 32]	100	0.4	32±0	98.7	0.0	32±0
C9, …, C1	7	(64, 96]	100	0.9	97±6	100	0.0	96±3
C9, …, C1	8	(32, 64]	100	0.0	64±3	98.7	0.0	64±0
C9, …, C1	9	(0, 32]	100	0.0	32±3	98.7	0.0	32±0
C8, …, C1	5	(96, 128]	100	0.4	128±5	98.7	0.4	128±0
C8, …, C1	6	(64, 96]	100	0.0	96±6	100	0.0	96±5
C8, …, C1	7	(32, 64]	100	0.0	64±5	100	0.0	64±3
C8, …, C1	8	(0, 32]	100	0.2	32±5	100	0.9	32±0
C7, …, C1	3	(128, 160]	100	0.2	160±3	100	0.0	160±0
C7, …, C1	4	(96, 128]	100	0.2	128±3	100	0.0	128±0
C7, …, C1	5	(64, 96]	100	0.0	96±3	100	0.0	95±3
C7, …, C1	6	(32, 64]	100	0.0	64±3	100	0.2	64±3
C7, …, C1	7	(0, 32]	100	0.0	32±3	100	0.7	32±0
C6, …, C1	2	(128, 160]	100	0.4	159±6	100	0.0	159±3
C6, …, C1	3	(96, 128]	100	0.0	128±0	100	0.2	128±0
C6, …, C1	4	(64, 96]	100	0.0	96±0	100	0.0	96±3
C6, …, C1	5	(32, 64]	100	0.0	64±0	100	0.0	64±0
C6, …, C1	6	(0, 32]	100	0.0	32±0	100	0.0	32±0
C5, …, C1	2	(96, 128]	100	0.2	128±0	100	0.0	128±0
C5, …, C1	3	(64, 96]	100	0.0	96±3	100	0.0	96±3
C5, …, C1	4	(32, 64]	100	0.0	64±0	100	0.2	64±0
C5, …, C1	5	(0, 32]	100	0.0	32±3	100	0.0	32±0

Table 7: each row gives the best predictor in each row of table 6.

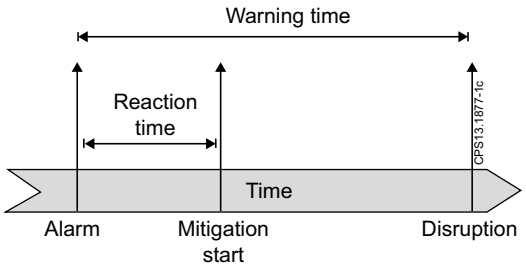

Figure 1: The reaction time can include, firstly, any computation time to decide between implementing avoidance or mitigation actions (and the selection of a specific methodology), secondly, the necessary time to fire technical systems and, thirdly, the plasma response time to the A/M actions.

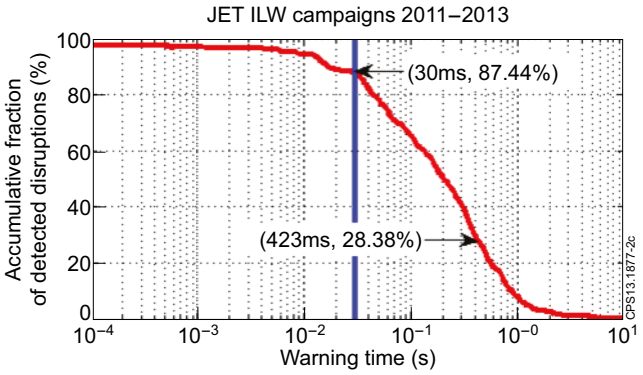

Figure 2: APODIS results during the 2011-2013 JET ILW campaigns (1078 non-disruptive discharges and 564 unintentional disruptions): success rate 97.34% (549/564)) and false alarm rate 1.64% (12/1078). The AWT is 423ms but there is a large uncertainty (827ms) in its determination. The solid vertical line at 30ms represents the minimum time in JET to carry out A/M actions.

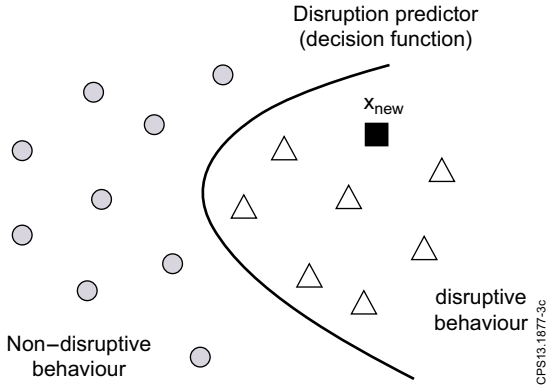

Figure 3: Figure caption.

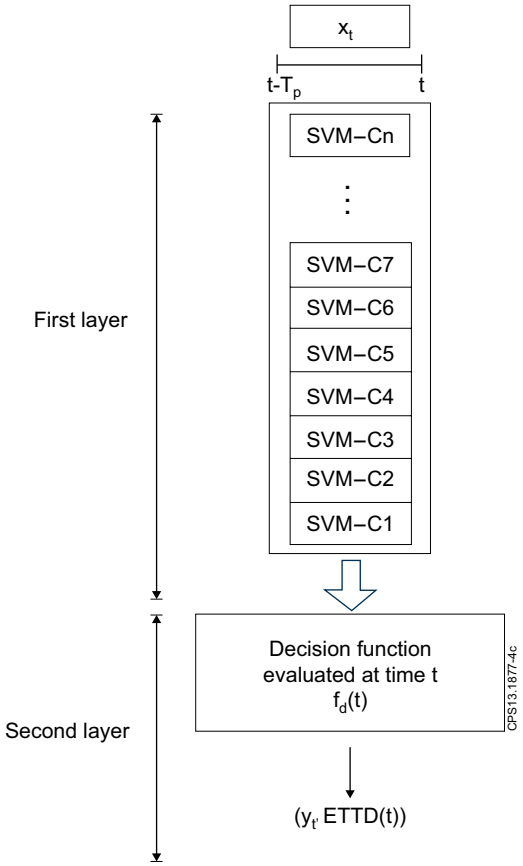

Figure 4: DTP architecture. The prediction takes place at time t from a feature vector $\mathbf{x}_t$ generated in the SPI

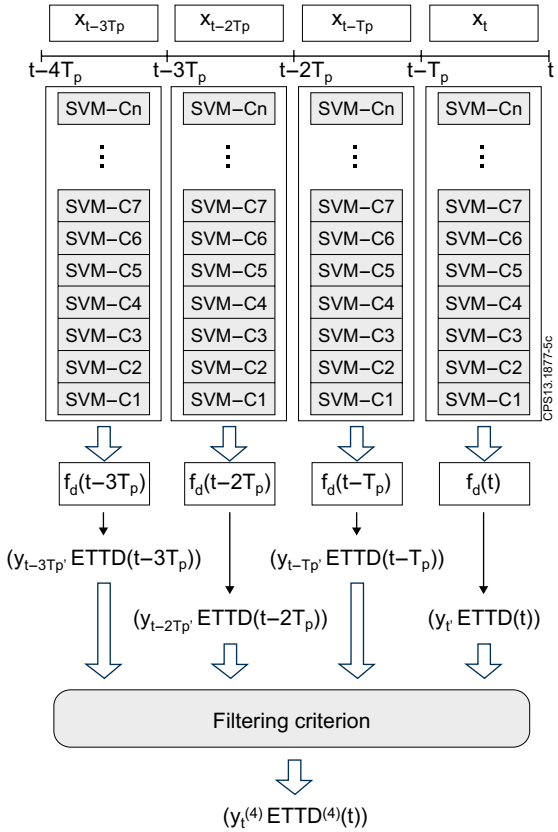

Fig. 5: Filtering criterion. The predictor only triggers an alarm after the recognition of N_C consecutive disruptive behaviours in the N_C consecutive DTPs. In this figure $N_C = 4$ and $y_t^{(4)} = \{+1\}$ if $y_{t-3TP} = y_{t-2TP} = y_{t-TP} = y_t\{+1\}$. Moreover, $ETTD^{(4)}(t) = ETTD(t)$.

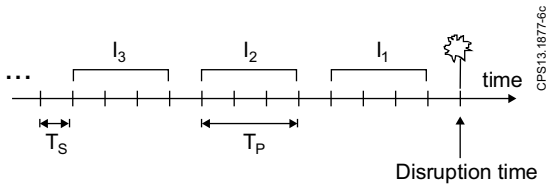

Figure 6: Sampling times after resampling in a disruptive discharge. The SPIs ($I_1, I_2, I_3, \ldots$) are numbered depending of its position in relation to the disruption time.

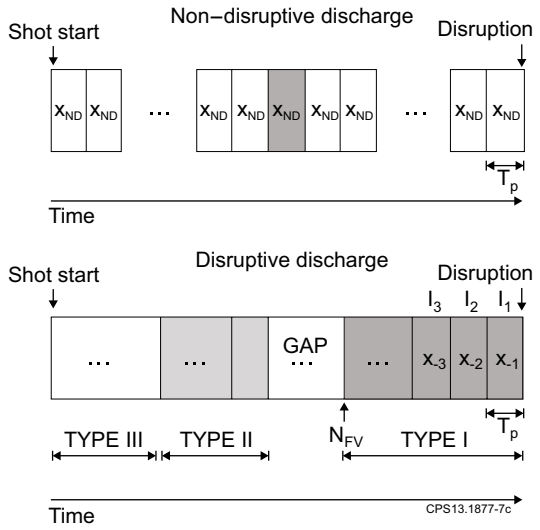

Figure 7: Description of SPIs for non-disruptive and disruptive training discharges.

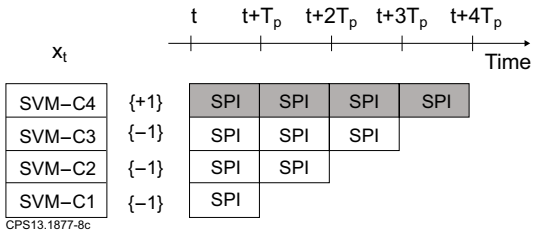

Figure 8: The only classifier that predicts a disruptive behaviour is SVM-C4. The combination of intervals determines that ETTD(t)∈(t+3·T_P,t+4·T_P).

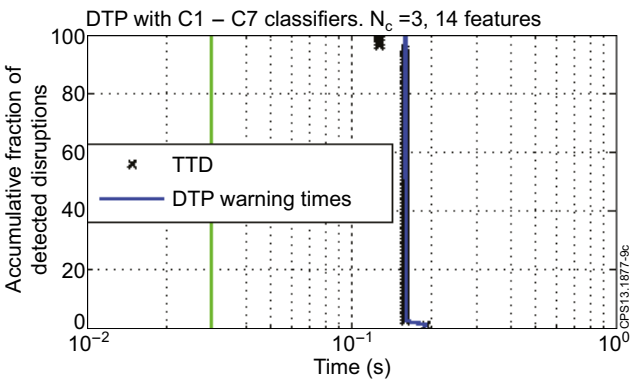

Figure 9: DTP results during the 2011-2013 JET ILW campaigns. The training was carried out with 622 and 121 non-disruptive and disruptive discharges respectively. The test set consists of 414 non-disruptive and 80 disruptive discharges: success rate 100% (80/80) and false alarm rate 0.2% (1/414). The solid vertical line at 30ms represents the minimum time in JET to carry out mitigation actions.

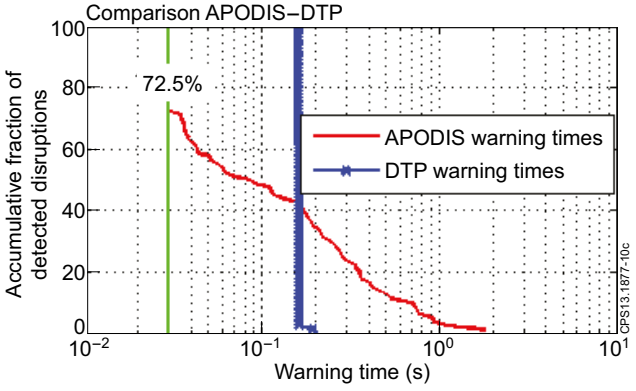

Figure 10: APODIS real results with the discharges of Fig.9: success rate after removing tardy detections 72.5% and false alarm rate 0.5%. The AWT is 289ms and its standard deviation is 332ms. The solid vertical line at 30 ms again represents the minimum time in JET to carry out mitigation actions.