IOP Publishing and International Atomic Energy Agency Nuclear Fusion
Nucl. Fusion **47** (2007) 1559–1569 doi:10.1088/0029-5515/47/11/018

A prediction tool for real-time application in the disruption protection system at JET

B. Cannas[1], A. Fanni[1], P. Sonato[2], M.K. Zedda[1] and JET-EFDA contributors[a]

[1] Electrical and Electronic Engineering Department, University of Cagliari, Piazza D'Armi, 09123, Cagliari, Italy
[2] Consorzio RFX, Associazione Euratom-ENEA sulla Fusione, Corso Stati Uniti, 4, I-35127 Padova, Italy

Received 2 March 2007, accepted for publication 22 August 2007
Published 19 October 2007
Online at stacks.iop.org/NF/47/1559

Abstract
A disruption prediction system, based on neural networks, is presented in this paper. The system is ideally suitable for on-line application in the disruption avoidance and/or mitigation scheme at the JET tokamak.

A multi-layer perceptron (MLP) predictor module has been trained on nine plasma diagnostic signals extracted from 86 disruptive pulses, selected from four years of JET experiments in the pulse range 47830–57346 (from 1999 to 2002).

The disruption class of the disruptive pulses is available. In particular, the selected pulses belong to four classes (density limit/high radiated power, internal transport barrier, mode lock and h-mode/l-mode).

A self-organizing map has been used to select the samples of the pulses to train the MLP predictor module and to determine its target, increasing the prediction capability of the system.

The prediction performance has been tested over 86 disruptive and 102 non-disruptive pulses. The test has been performed presenting to the network all the samples of each pulse sampled every 20 ms. The missed alarm rate and the false alarm rate of the predictor, up to 100 ms prior to the disruption time, are 23% and 1%, respectively.

Recent plasma configurations might present features different from those observed in the experiments used in the training set. This 'novelty' can lead to incorrect behaviour of the predictor. To improve the robustness and reliability of the system, a novelty detection module has been integrated in the prediction system, increasing the system performance and resulting in a missed alarm rate reduced to 7% and a false alarm rate reduced to 0%.

PACS numbers: 52.35.Py, 52.55.Fa

1. Introduction

Disruptions are critical events in which the plasma energy is lost within a time span of few milliseconds. They present serious problems for the operation of a tokamak. Firstly, they limit the accessible range of operational parameters and performance. Secondly, major disruptions, taking place at high plasma current, expose the tokamak first wall components and vacuum vessel to severe thermo-mechanical and electro-mechanical stresses. This problem is even more serious for the new generation of tokamaks such as ITER where the design margins should be reasonably minimized to reduce the costs and the dimensions, whereas, in contrast, the availability and the reliability of the device must be maximized to demonstrate the consistency of the fusion nuclear device as a power generating plant. Therefore, avoiding plasma disruptions or predicting a forthcoming disruption in order to mitigate its effects is a major issue for the operation of the next tokamak devices.

Disruptions have different physical causes most of which have been identified [1–7]. Nevertheless, the understanding of the underlying mechanisms is not exhaustive for the implementation of a deterministic disruption prediction system. Thus, a black box approach that predicts disruption directly from a set of plasma parameters measured during a discharge seems to be suitable because it does not require one to explicitly define the problem.

The following considerations have to be taken into account when designing a disruption prediction system:

- the prediction success rate should be greater than those of the existing alarm systems already available on tokamak machines, generally based on thresholds applied on single signals;
- the prediction time should sufficiently anticipate the starting of the disruption, in order to allow the mitigation and shut-down systems to safely intervene;

[a] See annex of Pamela J. *et al* 2004 Overview of JET results *Proc. 20th IAEA Fusion Energy Conf. (Vilamoura, Portugal, 2004).*

- the false alarm rate should be limited;
- at the same time, an as low as possible missed alarm rate should be obtained;
- the prediction system should be able to forecast different types of disruptions, characterized by different operational scenarios and dynamics;
- the prediction system should be able to operate in real-time.

As in [8] and [9], in this paper the performance of the prediction system is evaluated in terms of the percentage of false alarms (*PFA)* and the percentage of missed alarms (*PMA)*, where *PFA* is defined as the ratio between the number of non-disruptive pulses predicted by the system as disruptive pulses and the total number of non-disruptive pulses, in per cent, while *PMA* is defined as the ratio between the number of disruptive pulses predicted as non-disruptive pulses and the number of disruptive pulses, in per cent. Furthermore, the prediction success rate is defined in the present paper as the success rate of the predictor in correctly predicting both disruptive and non-disruptive pulses. Note that a disruption prediction is considered successful if the system is able to correctly predict the disruption up to $t_{\rm pred}$ prior to the disruption time, where $t_{\rm pred}$ depends on the tokamak machine considered. At JET, the prediction time $t_{\rm pred}$ has been conventionally set equal to 100 ms before $t_{\rm D}$, where $t_{\rm D}$ is the time of disruption, which is the time instant corresponding to a plasma current time derivative larger than 6.6 MA s^{-1}. That time of disruption is stored in the JET disruption database. According to JET disruption experts, the choice of $t_{\rm pred}$ should allow to undertake, in advance, an adequate mitigation strategy.

In this paper, a multi-layer perceptron (MLP) has been trained to forecast an impending disruption from a set of nine plasma parameters measured during disruptive and non-disruptive discharges. In particular, 172 disruptive pulses and 102 non-disruptive pulses in the pulse range 47830–57346 were available from four years of JET experiments, from 1999 to 2002.

The disruption class is available for 86 disruptive pulses. In particular, these pulses belong to four classes (density limit/high radiated power, internal transport barrier, mode lock, and h-mode/l-mode) [10, 11].

In a plasma pulse different disruption phases can be observed: pre-precursor phase, precursor phase, fast phase and quench phase [6]. At JET, the fast phase and the quench phase typically last less then 40 ms [12]. Since during these phases the diagnostic signals could not be completely reliable, due to the presence of high induced currents and magnetic field variations, the time window, from $t_{\rm D}$ −40 ms to $t_{\rm D}$, has not been monitored in the present work.

Defining $t_{\rm prec}$ as the time instant that discriminates between stable and unstable states of the plasma, some disruption precursors are expected to appear in the time window from $t_{\rm prec}$ to $t_{\rm D}$ −40 ms. Thus, two consecutive phases can be identified before and after $t_{\rm prec}$: the non-disruptive phase and the disruptive phase. Unfortunately, the identification of $t_{\rm prec}$ is often a very difficult task. Different physic criteria have been investigated, based on the analysis of the H–L transition, the loop voltage signal, the relative τ signal and the locked mode signal. Therefore, none of these criteria enabled us to unambiguously identify the two phases. Moreover, $t_{\rm prec}$ does not have a unique value for the different pulses, and presently, indices of the transition from a phase to the other are not available.

Nevertheless, due to the fact that MLPs are supervised neural networks, for each disruptive pulse we have to distinguish between samples belonging to the non-disruptive phase and samples belonging to the disruptive phase, in order to associate them with a different output of the neural network. In particular, the samples of a disruptive pulse, which belong to the non-disruptive phase, will be associated with a null value of the network output, as well as all the samples of the non-disruptive pulses, whereas the samples of a disruptive pulse, which belong to the disruptive phase, will be associated with a network output equal to one. This is one of the main issues in the design of a disruption prediction system and, in particular, in the training set generation (when the predictor is an MLP).

In this paper, we have given up trying to identify the time instant $t_{\rm prec}$, and an uncertainty band (or *transition region*), which discriminates between the non-disruptive and disruptive phases, is introduced by clustering the samples of each pulse by means of a self-organizing map (SOM). The clustering procedure identifies three sets: the first set contains samples that are supposed to belong to the non-disruptive phase, the second set contains samples that are supposed to belong to the disruptive phase, while the third set contains the samples belonging to the *transition region*. These last samples are excluded from the training phase, in order to avoid giving ambiguous information to the MLP. Details on the use of the SOM to identify the uncertainty band are reported in section 5.

A second issue to be performed is the selection of a limited number of samples sufficient to train the MLP predictor. In a non-disruptive pulse, the number of samples is indeed too large to be used in the training set. Moreover, the number of samples in the non-disruptive phase of a disruptive pulse is much larger with respect to the number of samples available in the disruptive phase. A balance is needed between the number of samples selected to describe the disruptive phase and those selected to describe the non-disruptive phase. In fact, in neural learning, if data are unbalanced among different classes, the features representing the classes that have a small number of samples in the training set can be ignored by the system, because of the large number of learning samples in the remaining classes presented to the neural network during the learning.

In the current approach, only disruptive pulses have been used to train the MLP predictor, while both disruptive and non-disruptive pulses have been used to test the predictor performance. This approach belongs to the disruption proximity prediction methods, which has been investigated in the literature for several tokamaks [12–14].

As previously mentioned, in this paper a clustering procedure is used, based on SOMs [15, 16], which allows us to automatically separate the samples belonging to the disruptive phase from those belonging to the non-disruptive phase. Moreover, the clustering procedure allows us to automatically select a limited number of significant samples from the whole length of the pulse. The predicting performance of the MLP predictor is quite good, with a very limited number of false alarms.

The drawback of the approaches based on neural techniques is that the trained network could produce a not

reliable output when the input comes from an entirely new distribution. This could be the case in JET, and it has also been experienced in other tokomaks, where new plasma configurations or improved operational boundaries can lead to not recognized discharges.

An improvement can be made using novelty detection techniques [17]. Several novelty detection methods have been proposed in the literature to determine the degree of novelty of a given input based both on statistical and neural network approaches [18–20].

In this paper, a neural network approach based on SOMs is used to determine the 'novelty' of an input of the MLP predictor module. In the on-line application, the novelty detection should be used to assess the reliability of the network output, i.e. samples having a low confidence have to be discarded and used off-line to update the disruption predictor.

The paper is organized as follows. In section 2 a survey on disruption prediction systems presented in the literature is reported, while section 3 is dedicated to the database selection and to the description of the diagnostic signals used as input of the prediction system. Section 4 briefly introduces MLP neural networks and SOMs, used in this paper. In section 5 the proposed disruption predictor is presented and the results are discussed. In section 6 novelty detection methods are described and an implementation through SOMs is reported. Finally, section 7 summarizes the conclusions.

2. Related studies

In the literature several papers have described the operational limits of a tokamak and the theoretical stability limits of the plasma: the Greenwald plasma density limit [1], the high-β limit [6], the l_i–q_ψ diagram [3] and the ratio of the radiated power to the input power [4]. Nevertheless none of them have led to the development of a reliable predictive model of disruptions. For this reason in the last 15 years, there have been several studies for the prediction of disruptions using neural networks. In particular some papers are focused on predicting the proximity of the plasma to disruption by using an artificial output.

The high β disruption boundary was modelled in [8] using 33 input magnetic measurements for DIII-D. A three-layer MLP has been trained on 56 disruptive discharges in order to predict the value of normalized β_N (i.e. the value of β referred to the toroidal field). The ratio between the network output and the actual value of β_N is an index of the stability of plasma confinement; therefore, it can be used as a β limit detection parameter. The probability of correct disruption prediction and the probability of a false alarm can be, respectively, maximized and minimized by choosing a proper alarm threshold for this ratio. At least 90% of disruptions in the test set, composed of 28 disruptive shots, were successfully predicted many tens of milliseconds before the major disruption, but false alarms were generated in about 20% of the cases. In this study, 94 inputs, consisting of 47 diagnostic signals and their normalized time derivatives, were initially adopted. By employing the sensitivity analysis, the number of inputs was reduced to 33. All time derivatives were excluded due to their low contribution to the prediction.

In [21] two buffered MLPs have been trained to forecast Mirnov measurements that identify $m = 2$ MHD instabilities in the TEXT tokamak. The first buffered MLP is trained to produce the Mirnov coil measurement at the time t using the past measurements at time steps $t - \tau$, $t - 2\tau$, ..., $t - N\tau$, where τ is the sampling time of the diagnostic system. Subsequently, the neural network works in the autonomous mode (i.e. the output at time t is fed back for the calculation of the output at time $t + \tau$) to obtain the prediction at time $t + K\tau$. The second buffered MLP is trained to produce the Mirnov coil measurement directly at time $t + K\tau$ using the past measurements $t - \tau, t - 2\tau, \ldots, t - N\tau$.

A similar approach has been developed by using soft x-ray measurements to predict the disruption three time steps earlier than the best network fed with Mirnov coil measurements alone [22].

In [13] an on-line predictor of the time to disruption installed on the Asdex Upgrade tokamak is presented. The prediction system uses a neural network trained on eight plasma parameters and some of their time derivative extracted from 99 disruptive discharges. The non-disruptive phase was defined as the L-mode phase following the H-mode phase before the disruption, or the phase starting just before a MARFE and ending with a disruption, for the plasma which has been in the L-mode for more than 0.8 s. The system was implemented and tested for real-time mitigation, showing satisfactory prediction capability. However the authors highlight the deterioration of the network performance on on-line tests, due to the slight difference between the real-time signal and the stored ones. Moreover, it has been shown that new experiments, which belong to operational spaces different from those used for training, are not well predicted in the on-line implementation, thus presenting the so-called 'ageing' of the neural network.

The work presented in [12] has been performed on flat-top JET scenarios characterized by a single null plasma. The authors trained a MLP to forecast disruptive events at JET, up to 100 ms in advance. Pulse samples have been selected in a temporal window of 400 ms. For disruptive pulses, the window is constituted of the last 400 ms of the discharge and the artificial output is a sigmoid representing the risk of disruption. In the paper a saliency analysis is also presented to validate the suitability of the selected input signals.

The major disruptions caused by the density limit, the plasma current ramp-down with high internal inductance, the low density locked mode and the β-limit in JT-60U have been investigated in [14]. The 'stability level', proposed in the paper is calculated from nine plasma parameters by an MLP, and the occurrence of a major disruption is predicted when the stability level decreases to a certain level, named the 'alarm level'. In particular, the onset of the major disruption is determined as the start of the positive current spike followed by the plasma current quench. The β-limit disruption prediction performance has been improved in [9] with a cascade of specialized MLPs.

3. Diagnostic signals and database selection

The diagnostic signals for training and testing the neural predictor were selected in the pulse interval 47830–57346, produced at JET between March 1999 and October 2002. The

Table 1. Diagnostic signals.

Signal name	Units
1. Plasma current	[A]
2. Locked mode	[T]
3. Radiated power	[W]
4. Plasma density	[1/m^3]
5. Input power	[W]
6. Internal inductance	
7. Safety factor	
8. Poloidal beta	
9. Plasma centroid vertical position	[m]

different disruption classes in the JET machine have been manually classified by a team of experts trying to identify the following ones: mode lock (ML), density limit (DL), high radiated power (RP), H-mode/L-mode transition (HL), internal transport barrier (IT) and vertical displacement (VD). On the basis of a deeper analysis, some classes have been discarded or merged [10, 11].

For example, it is worth noting that all but one VD disruptive pulses stored in the available JET database are provoked by the control system. For this reason, predicting VD disruptions is not an interesting task at JET. Moreover, the ideal beta limit and the low q limit disruptions are not considered since they are very rare at JET. More precisely, only one beta limit disruption and no low q limit disruptions are present in the large database considered in this paper.

As described in a previous paper [12], the signals used as input for the predictor are extracted from the available set of real-time signals in the JET JPF database. The sampling rate of the signals differs from diagnostic to diagnostic and therefore, for the use in the predictor, the input data set must be resampled at the lowest sampling rate which for the considered set of signals is 20 ms. This resampling is sufficient, as shown in the results, considering that the target of the predictor is to generate the alarm 100 ms before the disruption occurrence.

The discharges included in the network database satisfy the following requirements:

- plasma current $I_{\rm pla} > 1.5$ MA;
- X-point configuration;
- flat-top plasma current profile.

Discharges with $I_{\rm pla}$ below 1.5 MA were discarded as they generally have little impact on subsequent conditioning and operation of the device.

Nine diagnostic signals have been selected to describe the plasma regime during the discharge flat-top. These signals represent the input of the neural network prediction system. The choice of the signals takes into account physical considerations and the availability of real-time data. Moreover, previous experiences on disruption prediction confirm the appropriateness of the chosen input variables [10, 11, 23].

Table 1 shows the selected diagnostic signals.

The whole database consists of 172 disruptive pulses and 102 non-disruptive pulses. The information on the disruption class is available for 86 pulses.

4. Neural network algorithms

An artificial neural network (ANN) is a system composed of simple processing elements operating in parallel. The processing ability of the network is stored in the inter-unit connection strengths (weights), obtained by a process of adaptation to a set of training patterns (learning).

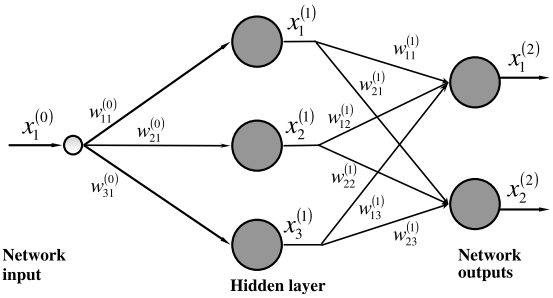

Figure 1. MLP with one input, one hidden layer, three hidden neurons and two outputs.

Today, ANNs are applied to an increasing number of real world problems of considerable complexity. They offer ideal solutions to a variety of classification problems such as pattern recognition, speech, character and signal recognition, as well as functional prediction and system modelling. ANNs may also be applied to control problems, where the input variables are measurements used to drive an output actuator, and the network learns the control function.

There are many different types of ANNs. In this paper, a traditional MLP has been trained to estimate the disruption risk from diagnostic data, whereas a SOM has been used for both clustering and novelty detection tasks.

4.1. The multi-layer perceptron

The MLP is the most widely used type of neural network. An MLP network consists of n_I inputs, one or more hidden layers of neurons and one output layer with n_O neurons. The neurons in each layer are connected with all the neurons of the previous layer.

The output of the ith neuron in the lth layer $x_i^{(l)}$ is a non-linear function of the weighted sum of the previous layer outputs:

$$x_i^{(l)} = f\left(\sum_{j=1}^{n^{(l-1)}} w_{ij}^{(l-1)} x_j^{(l)}\right),$$

where $i = 1, \ldots, n^l, l = 1, \ldots, L$; n^l is the number of neurons in the lth layer, L is the number of layers, $w_{ij}^{(l)}$ is the connection weight between neuron j in the $(l-1)$th layer and neuron i in the lth layer and f is a non-linear, usually sigmoidal, function [24].

Figure 1 shows an MLP with one input, one hidden layer, three hidden neurons and two output neurons.

Thus, an MLP with one hidden layer basically performs a linear combination of sigmoidal functions of the inputs. A linear combination of sigmoids is useful because of the following.

- It can approximate any continuous function of one or more variables. This is useful for obtaining a continuous function fitting a finite set of points when no underlying model is available [25].
- If the network is trained with a binary target, its outputs can be interpreted as posterior probabilities. This is very useful for classification tasks, as it gives a certainty measure on the classification performance [24].

The connection weights are determined, during the so-called *learning phase*, by applying a set of actual input–output values (the training set) to the network and comparing, through the error function, the network output with the desired output.

To ensure good out-of-sample generalization performance, a cross-validation technique can be used during the training phase, based on monitoring the error on an independent set, called the validation set [24].

In this paper, the available data have been divided into three subsets.

- The training set: it is used to update the network weights and biases.
- The validation set: the error on the validation set is monitored during the learning process. The validation error will normally decrease during the initial phase of learning, as the training set error does. However, when the network begins to over-fit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the learning is stopped, and the weights and biases at the minimum of the validation error are returned.
- The test set: as the training set and the validation set play a key role in selecting the model, their reliability as an independent reference to evaluate the performance of the model is therefore compromised. To correctly estimate the performance of the model, a third set (the test set) is used (at least when the available data allow it). The generalization capability of the neural model is tested with the test set, to ensure that the results on the training and validation sets are reliable, and not artefacts of the training process.

4.2. The self-organizing map

The SOM, developed by Kohonen [15], is a type of ANN, especially suitable for data survey because it has prominent visualization properties. It projects high dimensional input space on a low-dimensional (usually two-dimensional) regular grid that can be effectively utilized to visualize and explore properties of the data, for example, the cluster structure.

Let us consider an I-dimensional input space X. The SOM defines a mapping from X onto a regular array of neurons, preserving the topological properties of the input.

Figure 2 shows an example of mapping of a three-dimensional input set onto a two-dimensional SOM.

A key feature of SOMs is that the neurons in the map are arranged such that neighbour neurons represent similar patterns, and neurons that are well separated represent different patterns. This means that patterns close to each other in the input space are mapped on the same or neighbour neurons in the output space, defining a clustering of the I-dimensional data.

The number of map neurons, which typically varies from a few dozen up to several thousand, determines the granularity of the map and affects the accuracy and generalization capability of the SOM: increasing the number of neurons enhances the accuracy but deteriorates the generalization capability, which are contradictory goals.

The SOM is iteratively trained using competitive learning. During the SOM learning, a weight vector with the same dimensionality as the input space is associated with each

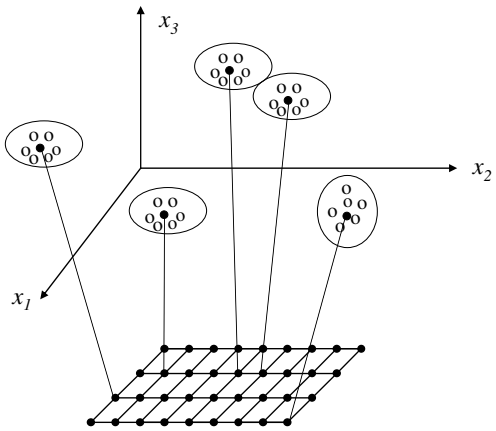

Figure 2. Projection of a three-dimensional input set into a two-dimensional SOM with 36 neurons.

neuron. The weights of the neurons are first initialized to small random values. Then, at each training step, a pattern is randomly chosen from the input data set and its Euclidean distance to all the weight vectors is computed. The neuron with the weight vector closest to the selected input pattern is called the best matching unit (BMU). The weights of the BMU and of the neurons close to it in the SOM are adjusted towards the chosen input. The magnitude of the change, defined by a neighbourhood function, decreases with time and it is smaller for neurons physically far away from the BMU. The update formula for a neuron with weight vector $\boldsymbol{w}(t)$ is

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \vartheta(\boldsymbol{w}, t)\alpha(t)\left[\boldsymbol{x}(t) - \boldsymbol{w}(t)\right],$$

where $\alpha(t)$ is a monotonically decreasing learning coefficient. The neighbourhood function $\vartheta(\boldsymbol{w}, t)$ depends on the distance between BMU and weight $\boldsymbol{w}$. In the simplest form, it is set equal to 1 for all neurons close enough to BMU and zero for the others (but a Gaussian function is a common choice too).

Learning generally proceeds in two broad stages: a shorter initial training phase in which the map reflects the coarser and more general patterns in the data, followed by a much longer fine tuning stage in which the local details of the organization are refined.

This process is repeated for each input pattern for a usually large number of cycles. When training is completed, each neuron of the map corresponds to a cluster of patterns in the input data set.

The map provides a convenient visualization surface showing the cluster structure of the data. However, the visualization can only be used to obtain qualitative information. To produce quantitative description of data properties other methods for giving good candidates for map unit clusters are required [16].

5. Disruption predictor

In the literature, the disruption prediction problem has been investigated for different tokamak machines [8, 12–14, 21–23], and several methods have been proposed to solve it, most of which are based on neural networks. Even if the results obtained are encouraging for the off-line analysis, further

investigations are necessary to carry out an effective real-time implementation. For example, a straightforward extension of the off-line predictor proposed in [12] to the real-time implementation, where the complete sequence of the samples of a pulse has to be presented to the predictor, considerably deteriorates the prediction performances.

In this paper, a disruption prediction system is proposed, developed with the goal of optimizing the prediction performance, which has been quantified in terms of maximizing the prediction success rate and minimizing the false alarm rate, when the entire pulses, rather than samples selected from limited time windows [12], are presented to the predictor.

Here, the disruption predictor is a traditional MLP. It takes as input the nine values of the diagnostic signals at the generic time instant t. Hence, the MLP has nine input nodes, while it has a single output node, with a logistic activation function.

Training and validation sets are composed of 69 and 17 disruptive pulses, respectively. For these pulses the corresponding disruption class is available.

During the training phase a null value is associated with the output node for all the samples belonging to the non-disruptive phase of a disruptive pulse. In contrast, a value equal to one is associated with the network output for all the samples of the disruptive phase of a disruptive pulse. In the following, details on the MLP architecture and on the training procedures used to obtain the prediction system are reported.

As previously mentioned, the most critical phase in developing a supervised neural network predictor, such as the MLP predictor considered here, is the learning phase. During this phase a suitable set of input–output couples $(\boldsymbol{x}^i, \boldsymbol{y}^i)$ (the training set) has to be selected in order to set the connection weights of the neural network. For each disruptive pulse, selected from the database for the training set, the ith input sample $(\boldsymbol{x}^i)$ is characterized by the nine diagnostic signals (see table 1). Input data have been normalized in the [0; 1] interval using the following expression:

$$x_{j_{\mathrm{norm}}} = \frac{x_j - m_j}{M_j - m_j}, \qquad j = 1, \ldots, 9,$$

where M_j and m_j are the maximum and the minimum absolute values of the jth training signal, respectively.

The corresponding output training pattern y^i will assume a value equal to 0 or 1 depending on whether the sample belongs to the non-disruptive phase or to the disruptive phase, respectively.

As previously cited, one of the main issues in the training set generation is the separation of the samples belonging to the disruptive phase from those belonging to the previous non-disruptive phase.

A second issue to be considered in the training set generation is the selection, for each pulse, of a limited number of samples sufficient to unambiguously describe the operational domain of the experiments.

In this paper, these two issues have been approached using the information derived from a clustering procedure.

The architecture of the predictor during the training phase consists of a cascade of a SOM and an MLP, as shown in figure 3.

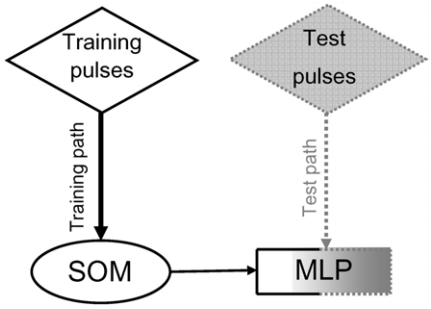

Figure 3. Architecture of the prediction system. Continuous lines: training path; dashed line: test path.

In order to build the training set for the MLP predictor, 86 two-dimensional SOMs have been constructed, one for each pulse in the training set and in the validation set. The nine-dimensional input patterns are mapped into two dimensions, allowing the visualization of the spatial distribution of the data.

It is worth noting that non-disruptive pulses have not been used during the training phase, assuming that stable states could be extrapolated from the non-disruptive phase of the disruptive pulses. Non-disruptive pulses will be used only during the test phase.

The SOM networks take as input all the samples of the disruptive pulses and give as output a two-dimensional map of clusters. As a SOM preserves the input topology, the map will be arranged such that samples close in the nine-dimensional input space are mapped in the same or in adjacent clusters in the two-dimensional map. Note that clusters that are well separated are supposed to contain samples belonging to different states of the plasma.

The identification of samples belonging to the disruptive phase and those belonging to the previous non-disruptive phase (first issue) and the selection of a limited number of samples belonging to the non-disruptive phase (second issue) have been performed using the following procedure.

All the samples belonging to the cluster that contains the sample at the time instant (t_{D} −40 ms), which is surely a disrupted sample, are considered *disrupted* samples. All the disrupted samples are included in the training set and the corresponding network target is set equal to 1.

For each pulse, an ambiguous region of the map, corresponding to the transition between the non-disruptive phase and the disruptive phase, is identified. The samples belonging to this region are considered *transition* samples and they are excluded from the training phase since they cannot be classified with sufficient confidence as either disrupted or non-disrupted samples. These samples belong to the clusters surrounding the cluster of disruptive samples.

The third set is the set of the *non-disruptive* samples. These samples belong to the remaining cluster units in the SOM and the corresponding network outputs have been set equal to 0.

In order to reduce the data dimensionality, only one sample for each of these clusters is considered for the training phase.

In figure 4, an example is reported for pulse no 53041, composed of 270 samples. The cluster that contains the 270th sample, corresponding to the time instant t_{D} −40 ms, is easily identified. This cluster is chequered in figure 4. This cluster

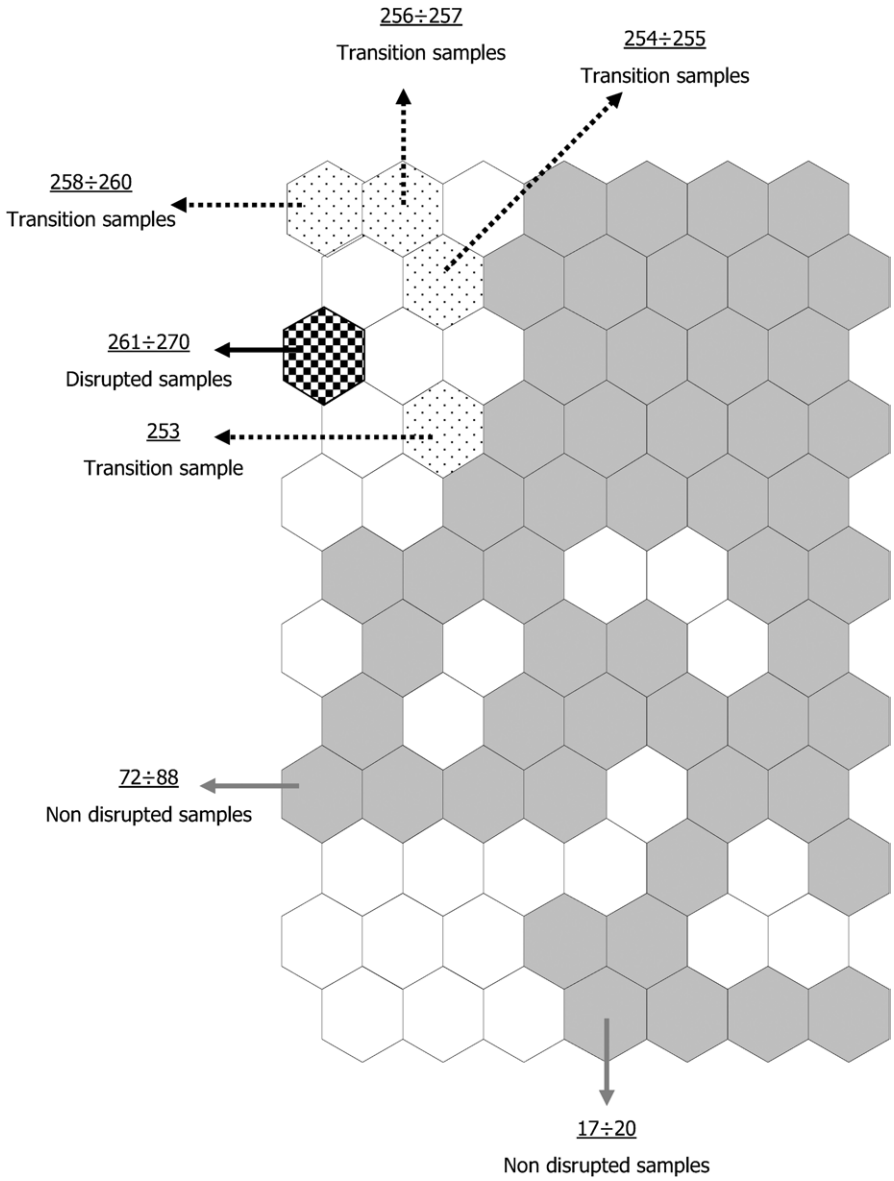

Figure 4. SOM of the disruptive pulse no 53041: the chequered map unit contains the sample at t_D-40 ms; the dotted map units contain transition samples; the white map units do not contain any sample; the grey map units contain safe samples.

contains other nine samples, which have been all included in the training set. The dotted clusters in figure 4 are clusters belonging to the transition phase. They have been excluded from the training set. The white clusters do not contain any sample, while only one sample for each grey cluster, which contains samples supposed to be non-disrupted samples, has been selected for the training set.

Hence, for the pulse considered in figure 4, only 72 samples have been selected for the training phase: 63 samples are non-disrupted samples; 9 samples are disrupted ones; 8 samples have been discarded, as they are transition samples.

Using the proposed procedure, the amount of training data has been considerably reduced with respect to the whole flat-top phase of the pulses, without losing any useful information for the network training. In particular, after the clustering, the 86 selected disruptive pulses, which contain 36316 samples, resulted in 7070 samples.

During the on-line running of the predictor, when the sequence of the samples is presented to the neural predictor, for each sample the MLP will return a real number between 0 and 1. This output represents a measure of the probability of the sample at the generic time instant t to belong to the disruptive phase, i.e. the probability of the time instant t to fall within the time windows $[t_{prec} \div t_D]$. Hence, if the MLP returns a value close to zero, it means that the sample belongs to the non-disruptive phase of a disruptive pulse or that it belongs to a non-disruptive pulse.

For a disruptive pulse, if the MLP output is greater than a prefixed alarm level (al) for at least two consecutive samples up to t_{pred} seconds prior to the disruption time, the predictor will correctly predict the disruption, and it will trigger the alarm. At JET, the prediction time t_{pred} is set equal to 100 ms, in order to allow a mitigation system to intervene.

In contrast, if the MLP output is less than al for all the samples of a disruptive pulse, the predictor will miss the alarm.

For a non-disruptive pulse, the predictor will trigger a false alarm if the network output is greater than the prefixed alarm level *al* for at least two consecutive samples of the pulse.

Note that the number of consecutive samples has been chosen by means of a sensitivity analysis.

The choice of the alarm level *al* is performed by minimizing a prediction error expressed as

$$e(al) = PFA(al) \cdot w_{\mathrm{FA}} + PMA(al),$$

where PFA (*al*) is the false alarm rate (or percentage of false alarms), PMA (*al*) is the missed alarm rate (or percentage of missed alarms) and w_{FA} is a false alarms weight factor.

The misclassification of non-disruptive pulses has been penalized by a weight factor $w_{\mathrm{FA}} = 4$, because in experimental machines, such as the JET tokamak, the minimization of the false alarm rate is mandatory. It is worth noting that, in order to minimize the missed alarm rate, rather than the false alarm rate, as it will be crucial in future fusion reactors, it will be sufficient to use a different alarm threshold modifying w_{FA}.

The test set consists of 86 disruptive pulses (34407 samples) and 102 non-disruptive pulses (61185 samples). During the test phase, all the samples of the test pulses directly feed the MLP predictor, by-passing the SOM, as shown in figure 3.

The MLP network topology has been selected by a trial and error procedure. For this purpose, several MLPs have been trained varying the number of hidden layers and the number of the hidden layer neurons. In particular, the growing method has been adopted. It consists of training a network having few neurons and then evaluating its performance. If such performance is satisfactory, the procedure ends, otherwise a network having more hidden neurons is trained, and so on, adding new neurons as long as the root mean squared error on the validation set is decreasing. In this way the training procedure avoids over-fitting, which derives from the excessive number of degree of freedom.

5.1. Results

The best network configuration, chosen considering the best performance in the validation set, is composed of 9 input neurons, 1 hidden layer with 15 neurons and 1 output neuron, resulting in 166 network parameters.

Table 2 shows the performance of the MLP predictor fed by the pulses selected for the training, the validation and test sets, respectively. Note that, as for the test set, for each pulse of the training and validation sets, the errors have been calculated considering the whole sequences of the samples rather than the samples selected by the SOM during the training phase. This demonstrates the success of the proposed samples selection and reduction procedure.

As can be noted, the network performance is excellent in terms of false alarms, while the PMA is quite high. This is partially due to the choice of a high value of the false alarms weight factor, and it is not an inherent property of the neural networks.

The performances of the MLP predictor in successfully predicting both disruptive and non-disruptive pulses are quite good. In particular, the system is able to correctly predict 63 pulses over the 86 disruptive pulses of the test set, and it does not trigger any alarm for 101 over the 102 non-disruptive pulses of the test set. It has to be pointed out that 3 disruptive pulses have not been considered to be correctly predicted because the system triggers the alarm too much in advance with respect to the disruption time t_{D}.

Table 2. Network performance in terms of percentage of false alarms (PFA), percentage of missed alarms (PMA) and prediction success rate.

	Training set (%)	Validation set (%)	Test set (%)
PMA	0.00	17.65	23.26
PFA	—	—	0.98
Prediction success rate	100	82.35	87.23

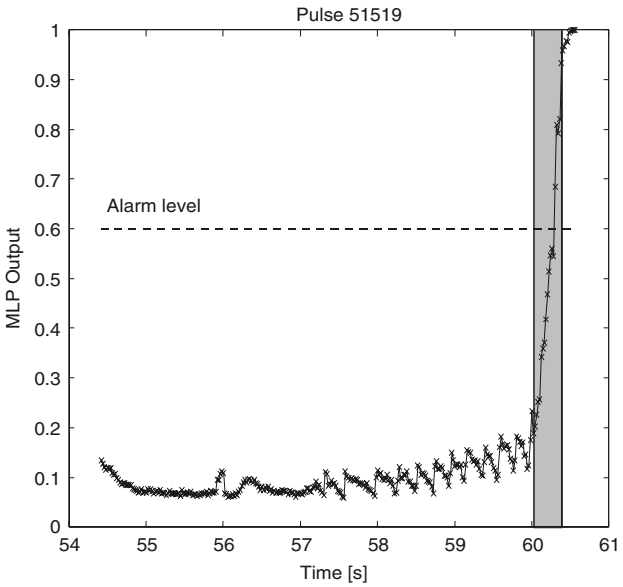

Figure 5. Output of the MLP predictor for the disruptive training pulse no 51519 that disrupts at 60.6 s.

Note that 40 test pulses belong to experimental campaigns performed 15 months before the campaigns considered for the training set. This choice has been done to verify the so-called ageing of the neural networks, which cause a deterioration of the performance of the predictor. In the next section, a novelty detection procedure will be presented to limit the ageing effect.

In figure 5 the network output for the disruptive pulse 51519, belonging to the training set, is reported. As can be noted, in this case the network is able to correctly predict all the samples of the non-disruptive phase, returning an output value close to zero. Moreover, concerning the samples in the transition region (shaded region in figure 5), the output of the network seems to follow the dynamics of the plasma, showing a growing trend from the non-disruptive state to the disruptive state (output close to 1).

In figure 6, the same behaviour can be observed for a disruptive pulse of the test set (pulse number 50166).

5.2. Prediction system versus mode lock indicator

The prediction capability of the MLP predictor has been compared with the performance of the mode lock indicator. The mode lock indicator (MLI) triggers a shut-down

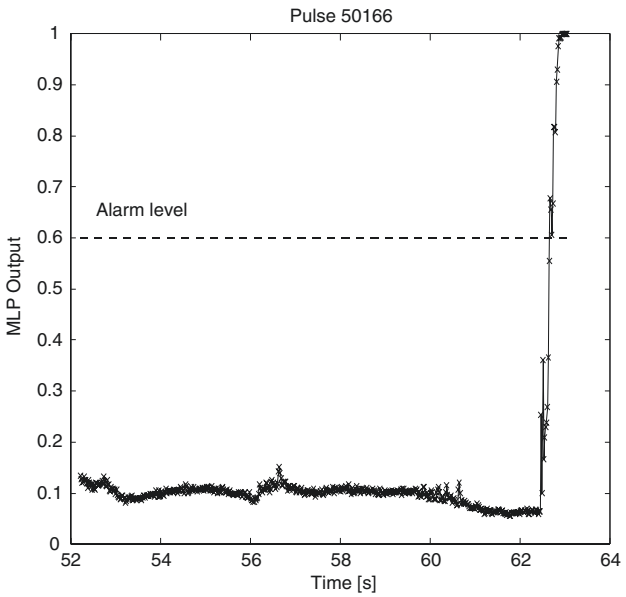

Figure 6. Output of the MLP predictor for the disruptive test pulse no 50166 that disrupts at 63.1 s.

Table 3. Comparison between MLP and SS.

	MLP	MLI
PMA	23.26%	51.16%

procedure, used at JET in the on-line disruption protection system.

As data on missed alarms are, obviously, the only data available at JET for the mode lock indicator, table 3 shows a comparison between PMA for the MLP and for the MLI.

The MLI missed the alarm for 44 pulses, while the MLP predictor missed the alarm for 20 pulses. It has to be pointed out that these 20 pulses are the same for the two systems. Moreover, a comparison of the prediction time for the 41 disruptions correctly predicted by both systems shows that the MLP triggers the alarm before the MLI for 28 disruptions.

6. Novelty detection

Novelty detection (ND) consists of identifying new or unknown data that a machine learning system is not aware of during the training phase. Thus, novelty detection is one of the fundamental requirements of a good classification or prediction system. In fact, actual data may contain patterns belonging to operational regions not explored when the learning system was developed. This could be the case of the disruption predictor presented in this paper, where new plasma configurations might present features completely different from those observed in the experiments selected for the training set. This 'novelty' can lead to incorrect behaviour of the MLP predictor (ageing of the neural network) [13].

6.1. Novelty detection techniques

In the last ten years novelty detection has attracted increasing attention, and a number of techniques have been proposed and investigated to address it. In [19], the authors provide a state-of-the-art review in the area of novelty detection based on statistical approaches, whereas in [20] novelty detection using neural networks is detailed.

In [19] and [20] the authors highlighted that it is not possible to *a priori* identify a single best model, and the success of a novelty detection technique mainly depends on the statistical properties of the data handled. Both statistic and neural clustering methods can be used for novelty detection tasks.

In this paper, novelty detection has been performed using a SOM for data clustering.

All the samples of the training and validation sets are considered to obtain a unique two-dimensional SOM map. Each cluster i of this map is represented by a prototype vector. Let $D_{\text{max-i}}$ be the maximum Euclidian distance of the samples in the cluster ith from its prototype vector.

Then, for each sample of the test set, the values of the distances from each prototype are evaluated. The winning cluster, called the best matching unit (BMU), is the nearest cluster, and the test sample is associated with this BMU.

During the test phase the answer of the MLP predictor is validated by the novelty detector using the information provided by that SOM. In particular, the following holds.

- When the output of the MLP predictor is close to zero, (i.e. MLP claims that the sample is a non-disrupted sample), if the distance of the test sample from its BMU is lower than $D_{\text{max-BMU}}$, the MLP answer is confirmed, and the sample is definitively considered as non-disrupted. Conversely, the test sample is labelled as novel.
- When the output of the MLP predictor is close to one, (i.e. MLP claims that the sample is a disrupted sample), if the distance of the test sample from its BMU is lower than $D_{\text{max-BMU}}$, the answer of the MLP is confirmed. In contrast, if the distance of the test sample from its BMU is greater than $D_{\text{max-BMU}}$, but the BMU contains only training disrupted samples, the novelty detector accepts the answer of the MLP, and the sample is definitively considered as disrupted, while if BMU does not contain only training disrupted samples, the test sample is labelled as novel.

In this paper, the proposed novelty detection technique is used to assess the network reliability. In particular, samples coming from unexplored operational spaces can be reliably rejected by the SOM novelty detector and used to update the training of the MLP neural predictor.

Figure 7 shows the architecture of the disruption prediction system integrated with the novelty detection block.

The presence of the ND block influences the predictor behaviour only in the case of disruption alarm. In particular, if the MLP triggers the alarm for a sample considered 'novel' by the novelty detector, the alarm will be rejected.

6.2. Results

The ND has been trained by means of the SOM Toolbox for Matlab, realized by the Helsinki University [26].

The performance of the MLP predictor presented in the previous sections, integrated with the ND block, is reported in terms of PMA and PFA calculated on a reduced test set obtained

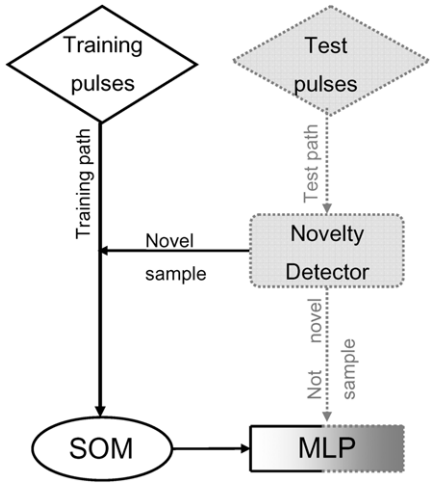

Figure 7. Architecture of the prediction system with the novelty detector.

by the previous test set by discarding the pulses labelled as novel by the novelty detector.

Note that, to emphasize the ageing of the network, the testing as been performed on discharges belonging to different campaigns.

The system returns a PMA equal to 7.14%, and a PFA equal to 0.00%.

In particular, 16 of the 20 missed alarms triggered by the MLP predictor are now reported as novel by the ND. The unique false alarm predicted by the MLP is reported as novel by the ND; thus the number of FAs in the on-line application is equal to zero. Furthermore, the 3 disruptive pulses, which have not been considered as correctly predicted, as the system triggered the alarm too much in advance with respect to the disruption time t_D, are reported as novel too.

It is worth noting that, as expected, some disruptions (11), correctly predicted by the MLP, are reported as novel by the ND. Hence, although the number of FAs and MAs considerably decreases, the discrimination capability of the system in the on-line application slightly reduces, with the prediction success rate, calculated on the entire test set, decreasing from 87.23% to 84.57%.

7. Conclusions

A disruption prediction tool, based on a MLP neural network, has been successfully implemented and tested over the whole flat-top phase of JET discharges.

The proposed neural predictor is ideally suitable for real-time application in the disruption mitigation/avoidance scheme at JET. Moreover, owing to the relatively low computational cost required by the neural network evaluation, the tool can be easily implemented on-line within the JET control system.

The robustness of the tool has been quantified on a test set of pulses at 100 ms before the disruption occurrence with the PFA being less than 1% whereas the PMA being lower than 24%.

The capability of predicting the disruption 100 ms before the occurrence of the phenomenon is very promising in order to apply mitigation procedures or soft landing.

The MLP predictor has been also compared with the existing deterministic locked mode indicator based on a threshold applied on a magnetic signal. The MLP predictor is able to reduce the PMA by more than 50% with respect to the locked mode indicator.

Moreover, in 66% of the disruptions correctly predicted by both systems, the MLP predictor is able to trigger the alarm before the locked mode indicator.

The test set of pulses has been created including disruptive pulses belonging to experimental campaigns temporarily far from those used to train the network. This has been intentionally done to highlight the ageing effect of the proposed approach, as recorded also in other similar tools applied to different tokamaks.

An integrated system, based on a novelty detector tool, able to detect the ageing of the input data, because it belongs to an operational space different from those used to train the network, has been implemented for the first time. This is able to reduce the missed alarm rate and the false alarm rate on 'not novel' pulses, whereas, as reasonably expected, the prediction success rate on the entire data set decreases because the novelty detector reports also a subset of disruptive pulses correctly predicted by the MLP predictor as novel. This result was also partially expected due to the choice done on the creation of the test set of pulses.

Acknowledgment

This work, supported by the Euratom Communities under the contract of Association between EURATOM/ENEA, 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.

The authors would like to thank Mike Johnson and David Howell for providing the manual classification of the disruptions and Tim Hender, Richard Buttery and Simon Pinches for supporting the work and for useful discussions.

References

[1] Greenwald M. 2002 Density limits in toroidal plasmas *Plasma Phys. Control Fusion* **44** 27–80
[2] Murakami M., Callen J.D. and Berry L.A. 1976 Some observation on maximum density in tokamaks experiments *Nucl. Fusion* **16** 347–8
[3] Fielding S.J., Hugill J., McCracken G.M., Paul J.W.M., Pretince R. and Stott P.E. 1977 High-density discharges with gettered torus walls in DITE *Nucl. Fusion* **17** 1382–5
[4] Wars D.J. and Wesson L.A. 1992 Impurity influx model of fast tokamaks disruptions *Nucl. Fusion* **32** 1117–23
[5] Nave M.F.F. and Wesson L.A. 1990 Mode locking in tokamaks *Nucl. Fusion* **30** 2575–83
[6] Wesson J. 2004 *Tokamak* 4th edn (Oxford: Clarendon)
[7] Schuller F.C. 1995 Disruptions in tokamak *Plasma Phys. Control. Fusion* **37** 135–62
[8] Wroblewski D., Jahns G.L. and Leuer J.A. 1997 Tokamak disruption alarm based on a neural network model of the high-beta limit *Nucl. Fusion* **37** 725–41
[9] Yoshino R. 2005 Neural-net predictor for beta limit disruptions in JT-60U *Nucl. Fusion* **45** 1232–46

[10] Zedda M.K., Bolzonella T., Cannas B., Fanni A., Howell D., Johnson M.F., Sonato P. and JET EFDA contributors 2003 Disruption classification at JET with neural techniques *30th EPS Conf. on Controlled Fusion and Plasma Physics (St Petersburg, Russia)* vol 27 A 2.93

[11] Cannas B., Cau F., Fanni A., Sonato P., Zedda M.K. and JET-EFDA contributors 2006 Automatic disruption classification at JET: comparison of different pattern recognition techniques *Nucl. Fusion* **46** 699–708

[12] Cannas B. *et al* 2004 Disruptions forecasting at JET using neural networks *Nucl. Fusion* **44**68–76

[13] Pautasso G. *et al* 2002 On-line prediction and mitigation of disruption in Asdex Upgrade *Nucl. Fusion* **42** 100–08

[14] Yoshino R. 2003 Neural-net disruption predictor in JT-60U *Nucl. Fusion* **43** 1171–786

[15] Kohonen T. 1982 Self-organized formation of topologically correct feature maps *Biol. Cybern.* **43** 59–69

[16] Vesanto J. and Alhoniemi E. 2000 Clustering of the self organising map *IEEE Trans. on Neural Netw.* **11** 586–600

[17] Cannas B., Fanni A., Sonato P., Zedda M.K. and JET EFDA contributors 2005 Novelty detection for On-line disruption prediction systems *32nd EPS Conf. on Plasma Physics (Tarragona, Spain*) vol 29C P 5.058

[18] Bishop C.M. 1994 Novelty detection and neural network validation *IEE Proc. Vision, Image and Signal Processing* **141** pp 217–22

[19] Markou M. and Singh S. 2003 Novelty detection: a review: 1. Statistical approaches *Signal Process.* **83** 2481–97

[20] Markou M. and Singh S. 2003 Novelty Detection: a review: 2. Neural network based approaches *Signal Process.* **83** 2499–521

[21] Hernandez J.V., Vannucci A., Tajima T., Lin Z., Horton W. and McCool S.C. 1996 Neural network prediction of some classes of tokamak disruptions *Nucl. Fusion* **36** 1009–17

[22] Vannucci A., Oliveira K.A. and Tajima T. 1999 Forecast of TEXT plasma disruptions using soft x-rays as input signal in a neural network *Nucl. Fusion* **39** 255–62

[23] Cannas B., Fanni A., Sias G., Sonato P., Zedda M.K. and JET EFDA contributors 2004 Neural approaches to disruption prediction at JET *31st EPS Conf. on Plasma Physics (London, UK)* vol 28 G 1.167

[24] Haykin S. 1999 *Neural Networks: A Comprehensive Foundation* (New Jersey: Prentice-Hall)

[25] Cybenko G. 1989 Approximation by superposition of a sigmoidal function *Math. Control, Signal Syst.* **2** 304

[26] Vesanto J. 1999 SOM-based data visualization methods *Intell. Data Anal.* **3** 111–26