

B. Cannas, A. Fanni, P. Sonato, M.K. Zedda and JET EFDA contributors

Real Time Prediction of Disruptions at JET

"This document is intended for publication in the open literature. It is made available on the understanding that it may not be further circulated and extracts or references may not be published prior to publication of the original when applicable, or without the consent of the Publications Officer, EFDA, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK."

"Enquiries about Copyright and reproduction should be addressed to the Publications Officer, EFDA, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK."

Real Time Prediction of Disruptions at JET

B. Cannas¹, A. Fanni¹, P. Sonato², M.K. Zedda¹ and JET EFDA contributors*

¹*Electrical and Electronic Engineering Dept.-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.*

** See annex of J. Pamela et al, "Overview of JET Results",*

(Proc. 20th IAEA Fusion Energy Conference, Vilamoura, Portugal (2004).

ABSTRACT

A disruption prediction system, based on neural networks, is presented in this paper. The system is presently running in real-time on the JET tokamak.

A Multi Layer Perceptron 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 Multi Layer Perceptron 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 behavior 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%.

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 tokamak like ITER. In fact, to meet its objectives of producing significant fusion power, ITER will have much higher stored magnetic and thermal energies, than JET. Therefore, avoiding plasma disruptions or predicting a forthcoming disruption in order to mitigate its effects, is a major issue for the operation of next tokamak devices.

Disruptions have different physical causes mostly 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 doesn't require to explicitly define the problem.

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

- the prediction success rate has to 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 has to sufficiently anticipate the starting of the disruption, in order to allow the mitigation and shut-down systems to safely intervene;
- the false alarm rate has to be limited;
- at the same time, an as low as possible missed alarm rate should be obtained;
- the prediction system has to be able to forecast different types of disruptions, characterized by different operational scenarios and dynamics;
- the prediction system has to be able to operate in real-time.

As in [8], in this paper the performance of the prediction system is evaluated in terms of Percentage of False Alarms (PFA) and 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_{pred} prior the disruption time, where t_{pred} depends on the tokamak machine considered. At JET, the prediction time t_{pred} has been conventionally set equal to 100ms before t_{D} , where t_{D} is the time of disruption. According to JET disruptions experts, that choice has been done since this time interval should be sufficient 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) [9-10].

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 are supposed to last less than 40 ms [11]. 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 $t_{\text{D}}-40\text{ms}$ to t_{D} , has not been monitored in the present work.

Defining t_{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_{prec} to $t_{\text{D}}-40\text{ms}$. Thus,

two consecutive phases can be identified before and after t_{prec} : non-disruptive phase and disruptive phase. Unfortunately, t_{prec} does not have a prefixed value, and the identification of the two different phases is often a very difficult task. Presently, indexes of the transition from a phase to the other are not available.

Nevertheless, as 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 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 to a null value of the network output, as well as all the samples of non disruptive pulses, whereas the samples of a disruptive pulse, which belong to the disruptive phase, will be associated to 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).

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 the proposed 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 [11-13].

In this paper a clustering procedure is used, based on Self Organising Maps (SOMs) [14-15], which allows 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 deteriorate its performance once it is on-line. In fact, a network, which is trained to discriminate between inputs coming from a set of distributions, could produce a not reliable output when input come from an entirely new distribution. This could be the case in JET, and it has been also experienced in other tokomaks, where new plasma configurations or improved operational boundaries can led to not recognized discharges.

An improvement can be made using Novelty Detection techniques [16]. Several novelty detection methods have been proposed in literature to determine the degree of novelty of a given input based both on statistical and neural network approaches [17-19].

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 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 Multi Layer Perceptron neural networks and Self Organizing Maps, used in the present 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 literature several papers 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 I_p - q_{95} 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.

In [12] 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 plasma which has been in 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 [11] 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 by 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 [13]. The concept of ‘stability level’, proposed in the paper is calculated from nine plasma parameters by

a 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 \leq -limit disruption prediction performance has been improved in [8] 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 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 base of a deeper analysis, some classes have been discarded or merged [9-10].

For example, it is worth noting that all but one the 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.

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

- Plasma current $I_{pla} > 1.5$ MA;
- X-point configuration;
- Flat-top plasma current profile.

Discharges with I_{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 [9, 11, 20].

The sampling time is 20 ms. This sampling time has been chosen in order to allow the synchronization among different acquisition systems. The neural predictor works indeed by gathering signals from various not synchronized sources; therefore, it has to wait for the slower system before the computation. Table I 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).

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 Multi Layer Perceptron has been trained to estimate the disruption risk from diagnostic data, whereas a Self Organising Map has been used both for clustering and novelty detection tasks.

4.1 THE MULTI LAYER PERCEPTRON

The Multi Layer Perceptron 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 i^{th} neuron in the l^{th} layer is a nonlinear function of the weighted sum of the previous layer outputs. The non linearity is introduced by the non linear, usually sigmoidal, activation function of the neurons [21]. Thus, a MLP with one hidden layer basically performs a linear combination of sigmoidal functions of the inputs. A linear combination of sigmoids is useful because:

- it can approximate any continuous function of one or more variables. This is useful to obtain a continuous function fitting a finite set of points when no underlying model is available [22];
- 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 classification performance [21].

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 to 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 validation set [21].

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 training set and validation set play a key role in selecting the model, their reliability as 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 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 ORGANISING MAP

A SOM is a type of ANN developed by Kohonen [14]. It projects high dimensional input space on a low-dimensional regular grid that can be effectively utilized to visualize and explore properties of the data.

Let us consider an I-dimensional input space X. The SOM defines a mapping from X onto a regular (usually two-dimensional) array of neurons, preserving the topological properties of the input. A key feature of SOMs is that the nodes in the output grid are arranged such that neighbour nodes represent similar patterns, and nodes that are well separated represent different patterns. This means that points close to each other in the input space are mapped on the same or neighbour neurons in the output space. Every neuron of the map corresponds to an I-dimensional weight vector that minimizes the distance from every input x_i associated to the neuron.

The map can be used as a convenient visualization surface for showing different features of the data, e.g., the cluster structure. However, the visualization can only be used to obtain qualitative information. To produce quantitative description of data properties other methods to give good candidates for map unit clusters are required [15].

In the first step of the training procedure, the weights are randomly initialised. Then, for each point x_i , the closest neuron in terms of Euclidean distance (or another measure of similarity) is determined and the winning neuron, and those within a certain neighbourhood around the winner, are updated. This process is repeated over several iterations until a stopping condition is reached. 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 organisation are refined. When training is completed, the weight vectors associated with each neuron define the partitioning of the multidimensional data.

5. DISRUPTION PREDICTOR

In literature, the disruption prediction problem has been investigated for different tokamak machines [11-13], and several methods have been proposed to solve it, most of which 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 [11] 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 to optimize the real time performances. The real time performances have 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 [11], are presented to the predictor.

Here, the disruption predictor is a traditional MLP. It takes as input the 9 values of the diagnostic signals at the generic time instant t . Hence, the MLP has 9 input nodes, while it has a single output node, with a logistic activation function.

Training and the validation set 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 to the output node for all the samples belonging to the non disruptive phase of a disruptive pulse. On the contrary, a value equal to one is associated to the network output for all the samples of the disruptive phase of a disruptive pulse. In the following, details on MLP architecture and on the training procedures used to obtain the prediction system will be reported.

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 $[t_{\text{pred}} \div t_{\text{D}}]$. 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 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.

On the contrary, 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.

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

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

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 $w_{FA} = 4$, because in experimental machines, like 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} .

As previously mentioned, the most critical phase in developing a supervised neural network predictor, like the MLP predictor here considered, is the learning phase. During this phase a suitable set of input-output couples (\underline{x}^i, 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 i^{th} input sample \underline{x}^i , is characterized by the 9 diagnostic signals (see Table I). Input data have been normalized in the [0; 1] interval using the following expression:

$$x_{jnorm} = \frac{x_j - m_j}{M_j - m_j}, j = 1, \dots, 9$$

where M_j and m_j are the maximum and the minimum absolute value of the j^{th} training signal respectively. The corresponding output training pattern y^i will assume a value equal to 0 or 1 depending on the belonging of the sample 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 a MLP, as shown in Fig. 1.

In order to build the training set for the MLP predictor, 86 SOMs have been constructed, one for each pulse in the training set and in the validation set.

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 2-dimensional map of clusters. As a SOM preserves the input topology, the map will be arranged such that samples close in the 9-dimensional input space are mapped in the same or in adjacent clusters in the 2-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 ($t_D - 40\text{ms}$), 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, a region of the map, corresponding to the transition between the non disruptive

phase and the disruptive phase, is identified on the basis of empirical considerations on the map topology. The samples belonging to this region, are considered transition samples and they are excluded from the training phase since they cannot be classified with a sufficient confidence either as disrupted or non disrupted samples.

The samples belonging to the remaining clusters are considered non disrupted samples. In order to reduce the data dimensionality, only one sample for each of these clusters is considered for the training phase, and the corresponding network output has been set equal to 0.

In Fig.2, an example is reported for the Pulse No: 53041, composed by 270 samples. The cluster that contains the 270th sample, corresponding to the time instant $t_D-40\text{ms}$, is easily identified. This cluster is chequered in Fig.2. This cluster contains other nine samples, which have been all included in the training set. The dotted clusters in Fig.2 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 Fig.2, 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.

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, bypassing the SOM, as shown in Fig.1.

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, until the network reaches the desired performance, in terms of mean squared error. In this way the training procedure avoids the 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 II 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 percentage of missed alarms 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 performance of the MLP predictor in successfully predict 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 as correctly predicted because the system triggers the alarm too much in advance with respect to the disruption time t_D .

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 Fig.3 the network output for the disruptive Pulse No: 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 Fig.3), the output of the network seems to follow the dynamic of the plasma, showing a growing trend from the non disruptive state to the disruptive state (output close to 1).

In Fig.4, 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 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 III 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) [12].

6.1 NOVELTY DETECTION TECHNIQUES

In the last ten years novelty detection acquired an increasing attention, and a number of techniques have been proposed and investigated to address it. In [18], the authors provide state-of-the-art review in the area of novelty detection based on statistical approaches, whereas in [19] novelty detection using neural networks is detailed.

In [18] and [19] 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 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 2-dimensional SOM map. Each cluster i of this map is represented by a prototype vector. Let D_{max-i} be the maximum Euclidian distance of the samples in the cluster i^{th} 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 Best Matching Unit (BMU) is the nearest cluster, and the test sample is associated to 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:

- 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 $D_{max-BMU}$, 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 $D_{max-BMU}$, the answer of the MLP is confirmed. On the contrary, if the distance of the test sample from its BMU is greater than $D_{max-BMU}$, but 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 becoming 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 5 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 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, realised by the Helsinki University [23].

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 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 equals to 7.14%, and a PFA equals 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 t_{met_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 of MAs considerably decreases, the discrimination capability of the system in the on-line application slightly reduces, with a prediction success rate, calculated on the entire test set, decreasing from 87.23% to 84.57%.

CONCLUSIONS

A Real time 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 robustness of the tool has been quantified on a test set of pulses at 100ms before the disruption occurrence with a percentage of False Alarms less than 1% whereas the percentage of Missed Alarms is lower than 24%.

The capability to predict 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 percentage of Missed Alarms 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 belonging 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 as novel also a subset of disruptive pulses correctly predicted by the MLP predictor. This result was also partially expected due to the choice done on the creation of the test set of pulses.

ACKNOWLEDGEMENT

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 the 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., Berry L.A. 1976 Some observation on maximum Density in Tokamaks Experiments *Nucl. Fusion* **16** 347-348.
- [3]. Fielding S.J., Hugill J, McCracken G.M., Paul J.W.M., Pretince R., Stott P.E. 1977 High-Density discharges with gettered Torus walls in DITE *Nucl. Fusion* **17** 1382-1385.
- [4]. Wars D.J., Wesson L.A. 1992 Impurity influx model of fast Tokamaks disruptions *Nucl. Fusion* **32** 1117-1123.
- [5]. Nave M.F.F., Wesson L.A. 1990 Mode locking in Tokamaks *Nucl. Fusion* **30** 2575-2583.
- [6]. Wesson J. 2004 Tokamak 4th edition Clarendon Press-Oxford.
- [7]. Schuller F.C. 1995 Disruptions in TOKAMAK *Plasma Phys. Control. Fusion* **37** 135-162.
- [8]. Yoshino R. 2005 Neural-net predictor for beta limit disruptions in JT-60U *Nucl. Fusion* **45** 1232-1246.
- [9]. 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 Conference on Contr. Fusion and Plasma Phys., St. Petersburg, **27** A 2.93.
- [10]. 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.
- [11]. Cannas B. et al. 2004 Disruptions forecasting at JET using Neural Networks Nucl. Fusion **44** 68-76.
- [12]. Pautasso G. et al. 2002 On-line prediction and mitigation of disruption in Asdex Upgrade Nucl. Fusion **42** 100-108.
- [13]. Yoshino R. 2003 Neural-net disruption predictor in JT-60U Nucl. Fusion **43** 1171-1786.
- [14]. Kohonen T. 1982 Self-organized formation of topologically correct feature maps Biological cybernetics **43** 59-69.
- [15]. Vesanto J., Alhoniemi E. 2000 Clustering of the Self Organising Map, IEEE Trans. on Neural Networks **11** 586-600.
- [16]. Cannas B., Fanni A., Sonato P., Zedda M.K., and JET EFDA contributors 2005 Novelty Detection for On-Line Disruption Prediction Systems 32th EPS Conf. on Plasma Phys., Tarragona, Spain, **29C** P 5.058
- [17]. Bishop C. M. 1994 Novelty detection and neural network validation Proc. of the IEE Conf. on Vision and Image Processing 217-222.
- [18]. Markou M., Singh S. 2003 Novelty Detection: a review – part 1: statistical approaches Signal Processing **83** 2481-2497.
- [19]. Markou M., Singh S. 2003 Novelty Detection: a review – part 2: neural network based approaches Signal Processing **83** 2499-2521.
- [20]. Cannas B., Fanni A., Sias G., Sonato P., Zedda M.K., and JET EFDA contributors 2004 Neural approaches to disruption prediction at JET 31th EPS Conf. on Plasma Phys. London, UK **28** G 1.167.
- [21]. Haykin S. 2000 Neural Networks Wiley & Sons.
- [22]. Cybenko G. 1989 Approximation by superposition of a sigmoidal function Mathematics of Control, Signal and Systems **2** 304.
- [23]. Vesanto J. 1999 SOM-based data visualization methods Intelligent Data Analysis **3** 111-126.

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]

Table I: Diagnostic signals.

	Training Set	Validation Set	Test Set
PMA	0.00%	17.65%	23.26%
PFA	–	–	0.98%
Prediction Success Rate	100%	82.35%	87.23%

Table II: Network performance in terms of Percentage of False Alarms (PFA), Percentage of Missed Alarms (PMA), and Prediction Success Rate.

	MLP	MLI
PMA	23.26%	51.16%

Table III: Comparison between MLP and SS

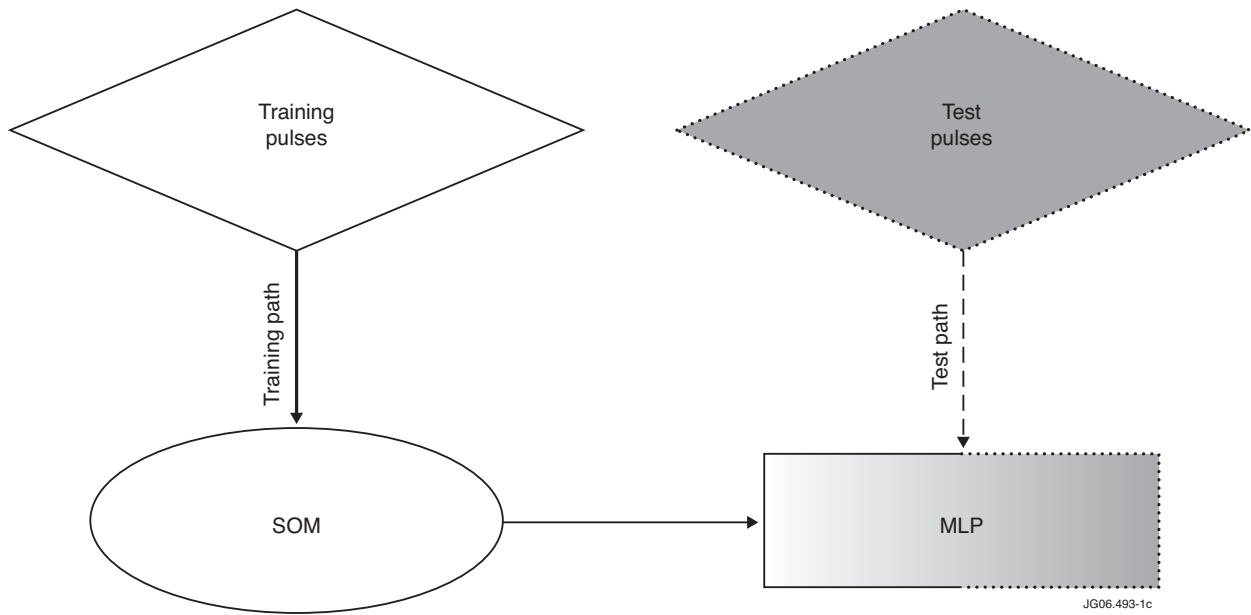


Figure 1: Architecture of the prediction system. Continuous lines: training path; dashed line: test path

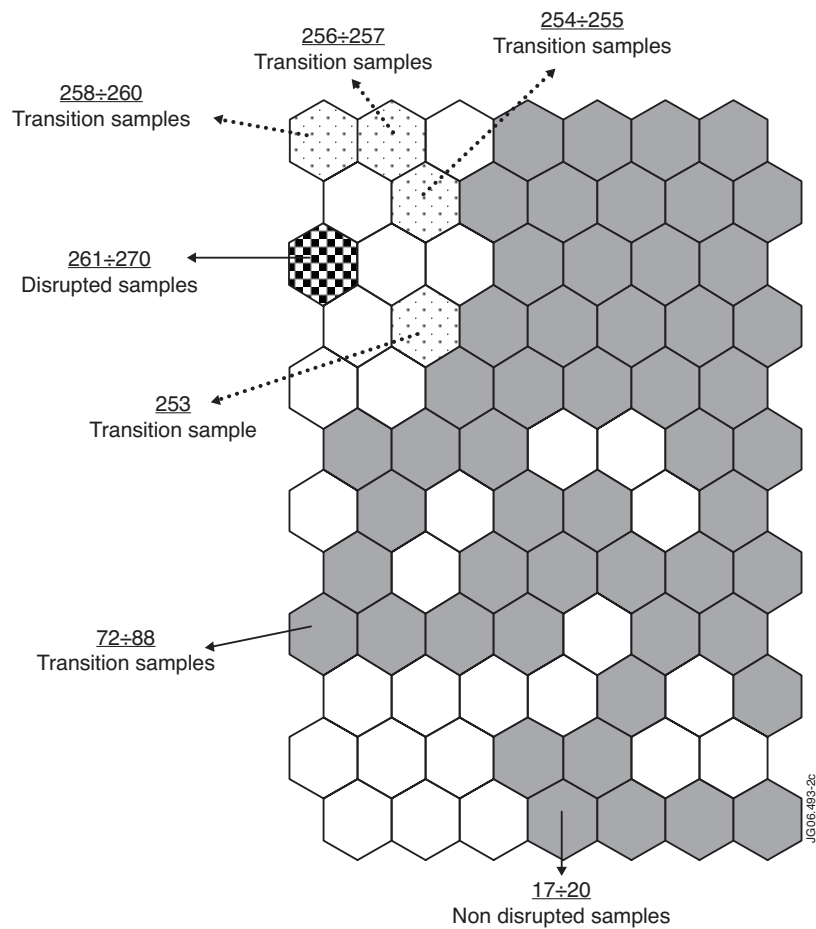


Figure 2: SOM of the disruptive Pulse No: 53041: the chequered map unit contains the sample at t_D ; the dotted map units contain transition samples; the white map units do not contain any sample; the grey map units contain safe samples.

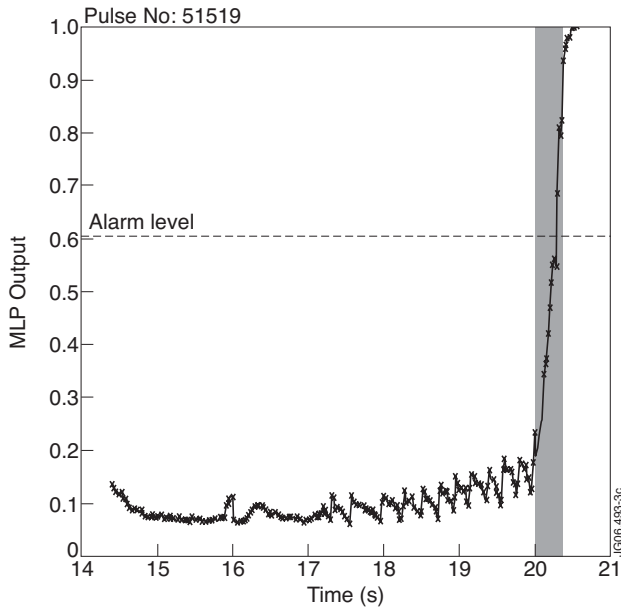


Figure 3: Output of the MLP predictor for the disruptive training pulse Pulse No: 51519.

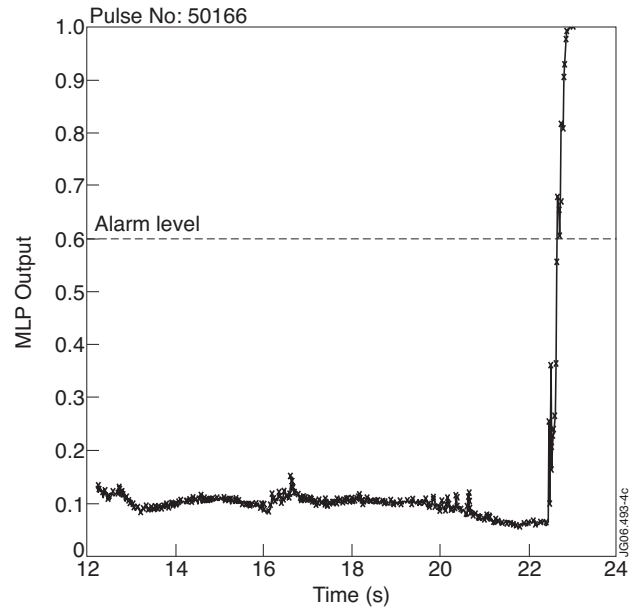


Figure 4: Output of the MLP predictor for the disruptive test pulse Pulse No: 50166.

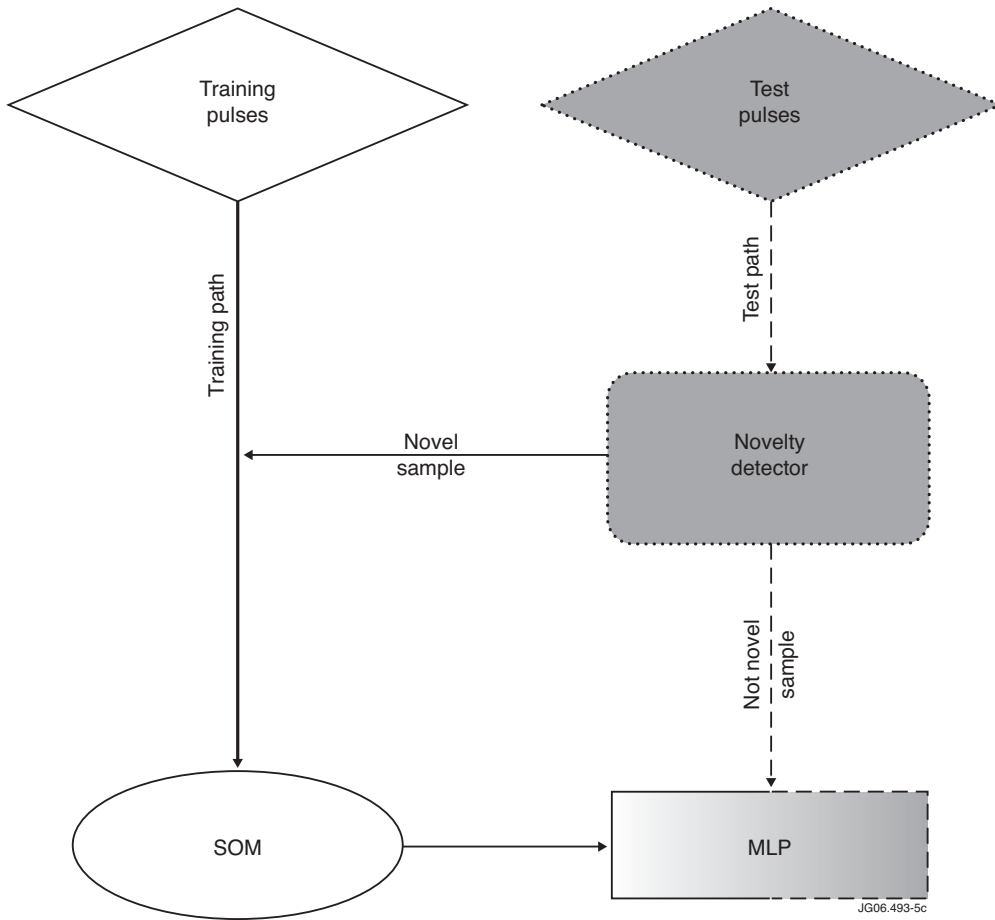


Figure 5: Architecture of the prediction system with the Novelty Detector.