PAPER • OPEN ACCESS

Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles

View the article online for updates and enhancements.

Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles

E. Aymerich[1],[*] [ORCID]**, G. Sias[1]** [ORCID]**, F. Pisano[1]** [ORCID]**, B. Cannas[1]** [ORCID]**, S. Carcangiu[1]** [ORCID]**, C. Sozzi[2]** [ORCID]**, C. Stuart[3], P.J. Carvalho[4]** [ORCID]**, A. Fanni[1]** [ORCID] **and JET Contributors[a]**

[1] Dept. of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

[2] Istituto per la Scienza e Tecnologia dei Plasmi, Consiglio Nazionale delle Ricerche, Via R. Cozzi, 53, 20125 Milano, Italy

[3] UK Atomic Energy Authority, Culham Science Centre, Abingdon OX14 3DB, United Kingdom of Great Britain and Northern Ireland

[4] Culham Centre for Fusion Energy, UK Atomic Energy Authority, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom of Great Britain and Northern Ireland

E-mail: enrico.aymerich@unica.it

Abstract

In view of the future high power nuclear fusion experiments, the early identification of disruptions is a mandatory requirement, and presently the main goal is moving from the disruption mitigation to disruption avoidance and control. In this work, a deep-convolutional neural network (CNN) is proposed to provide early detection of disruptive events at JET. The CNN ability to learn relevant features, avoiding hand-engineered feature extraction, has been exploited to extract the spatiotemporal information from 1D plasma profiles. The model is trained with regularly terminated discharges and automatically selected disruptive phase of disruptions, coming from the recent ITER-like-wall experiments. The prediction performance is evaluated using a set of discharges representative of different operating scenarios, and an in-depth analysis is made to evaluate the performance evolution with respect to the considered experimental conditions. Finally, as real-time triggers and termination schemes are being developed at JET, the proposed model has been tested on a set of recent experiments dedicated to plasma termination for disruption avoidance and mitigation. The CNN model demonstrates very high performance, and the exploitation of 1D plasma profiles as model input allows us to understand the underlying physical phenomena behind the predictor decision.

Keywords: disruption mitigation and avoidance, deep learning, spatiotemporal feature extraction, automatic pre-disruptive phase identification

(Some figures may appear in colour only in the online journal)

1. Introduction

Mitigating or, even better, avoiding plasma disruptions has become mandatory in view of future nuclear fusion devices, like ITER. The sudden loss of plasma confinement, followed by the quench of the plasma current, could result in the release of large amounts of energy and large thermal and electromagnetic loads, possibly causing severe damage to the plasma facing components and stressing the device with high mechanical forces [1]. In over two decades of investigations, artificial intelligence-based approaches demonstrated the great potential to predict disruptions in tokamak devices. Several machine learning methods such as multi-layer perceptron neural networks, support vector machines, self organizing maps and generative topographic mapping (GTM), classification and regression trees and random forests have been used to develop disruption prediction models for JET [2–8], ASDEX Upgrade [9–11], EAST [12], J-TEXT [13], DIIID [14] and Alcator C-Mod [12]. All these contributions have highlighted the importance of studying in depth the physical phenomena involved in disruptive processes in order to synthesize suitable disruption precursor signals to be used as inputs in the data-driven prediction models for avoidance or mitigation actions [7, 15–18]. In particular, invaluable information can be obtained from temperature, density and radiation plasma profiles due to their close connection with plasma stability and destabilization of MHD modes that may cause disruptions. Just as an example, in most cases, a disruption is the consequence of the development of tearing modes inside the plasma, which leads to the growing of the magnetic islands. Usually, well before the onset of the tearing modes, an increase of the radiation emission in the core, which leads to a hollow temperature profile, can be observed, whereas an increase in the radiation emission at the edge of the plasma leads to a cooling at the edge. In case of temperature hollowing, there is a broadening of the current density profile from inside, whereas a shrinking of the same profile from outside corresponds to the edge cooling. In both cases an unstable MHD scenario may arise due to a continuous increase of the current density gradient near the mode resonant surface [19]. On the other hand, the spatiotemporal information contained in the plasma profiles is crucial in describing destabilizing localized phenomena, such as the radiation emission in the core rather than at the edge, which cannot be enough described by zero-dimensional (0D) parameters, as the radiated fraction of the total input power.

To this end, 0D peaking factor signals have been introduced to encode the spatial information contained in the 1D profiles, through the ratio between the mean values of measurements over different regions of the plasma cross section. The peaking factor signals, constructed starting from temperature, density and plasma radiation profiles, and therefore well anchored to the plasma physics, demonstrated to increase the performance of the machine learning models predicting disruptions with enough warning time to more efficiently enable avoidance strategies [7, 8, 15, 17]. As an example, in [7, 15], the peaking factors of the radial profiles of temperature and density are defined as the ratio between the mean value around the magnetic axis and the mean value of the measurements over the entire radius. The radial interval to define the 'core' with respect to the magnetic axis was empirically set to a prefixed percentage of the radial coordinate. Regarding the radiated power, two distinct peaking factors were introduced to decouple the contribution of the plasma radiation from the core and from the divertor. In this case, the 'core' is defined as a certain percentage of the vertical semi-axis of the poloidal cross section. It is to be noted that, the definitions of such peaking factors are based on heuristics that arbitrarily assume the 'core' chords and the 'divertor' chords and can lose precious spatial information contained in the plasma profiles. Moreover, the peaking factor definitions must be changed depending on the different diagnostic systems available in the different devices. Additionally, in [20] peaking factor signals were defined for considering profile information, together with data from other diagnostics, with the aim of identifying the cause of the disruption, allowing the implementation of different responses.

In recent years, deep convolutional neural networks (CNNs) have proved capable of overcoming the most established machine learning techniques, especially in the field of image processing and computer vision, for their ability to learn relevant features from images at different scales, avoiding hand-engineered feature extraction [21].

The potential of CNNs has made them attractive also in the field of nuclear fusion in general [22–25], and for disruption prediction in particular [16, 26, 27]. In [16], the authors propose a disruption prediction model that combines recurrent and CNNs to extract spatiotemporal information coming from high-dimensional diagnostic data, such as the electron temperature and electron density profiles, in addition to 0D diagnostic signals. The proposed algorithm has been assessed using data from DIII-D and JET showing promising cross-machine generalization capability. In [26], an approach to disruption prediction using deep CNN is proposed, where raw data from only the electron cyclotron emission imaging diagnostic from the DIII-D tokamak is used. The initial results seem promising corroborating the idea of using CNNs to learn long-range, multi-scale plasma dynamics related with disruptions. In [27], a CNN is used to perform the bolometer tomographic reconstruction that provides a 2D image of the plasma radiation profile at JET. The bolometer data is then processed through a series of 1D convolutional layers whose output is the input of a recurrent neural network, which performs the disruption prediction. This approach led to the implementation of real-time tomography at JET [28]. Even if the disruption prediction performance is not comparable with

literature, the paper shows an interesting perspective of how deep learning architectures can manage multi-dimensional diagnostics.

In this paper, CNNs are proposed both to extract the spatiotemporal features from the plasma profiles of temperature, density and plasma radiation, overcoming the previously described limits of the 0D peaking factors, and to develop a quite simple deep neural network disruption prediction model that uses these features together with other diagnostic signals commonly used in the literature. The deep-CNN predictor has been trained using data from experimental campaigns performed at JET from 2011 to 2013. Its prediction performance has been evaluated using disrupted and regularly terminated discharges from a decade of JET experimental campaigns, from 2011 to 2020, showing the robustness of the algorithm, even though the operational spaces in the different campaigns may change significantly. Moreover, the performance of the proposed disruption predictor has been compared, on the same test sets, with the performance of the GTM predictor [7, 8] implemented in the Plasma Event TRiggering and Alarms (PETRA) system at JET [29], with even better results. Furthermore, monitoring the deep-CNN predictor output together with the provided input images, it is possible to interpret the physics mechanisms leading to the model response.

This paper is organized as follows: section 2 details the data base used to assess and optimize the algorithm and to validate it. Section 3 reports our implementation of the deep-CNN prediction model, whereas section 4 reports the testing results and the comparison with the GTM. In section 5, a detailed analysis of the plasma termination experiments performed at JET within the 2019 campaigns is reported. Finally, some conclusions are provided in section 6.

2. Database

In this work, in order to develop and test a disruption predictor based on a deep convolutional neural network model, a database has been built selecting disruptive and regularly terminated discharges from JET experimental campaigns from 2011 to 2020. The database for this work contains a total of 193 disrupted and 219 regularly terminated discharges having a flat-top plasma current higher than 1.5 MA, and a flat-top length greater than 200 ms (the minimum length necessary to create the input images for the model). The analysis of the pulses refers to the flat-top phase; the ramp-up and the ramp-down have not been considered. In particular, for each selected discharge, the flat-top starting time has been assumed as the first time instant where the plasma is in X-point configuration. For the disrupted pulses, the flat-top ending time (t_{end}) is assumed as the time of the valve activation for those terminated by massive gas injection (MGI), and as the disruption time (t_D), corresponding to the drop of the core temperature and the start of the plasma current spike, for the unmitigated ones. Disruptions caused by vertical displacement events have been excluded at all from the data set. These criteria are widely employed in disruption prediction and avoidance studies to select relevant experiments [7, 16]. The considered database covers a wide set of experimental conditions, starting from the

Table 1. Database composition.

Dataset	Campaigns	Composition	
		Disruptions	Regular pulses
I	2011 ÷ 2013	127	115
II	2016	29	41
III	2019 ÷ 2020	37	63

earlier campaigns with the ILW until the recent experiments where high power experiments were carried out. It has been grouped in three datasets, as detailed in table 1, following the different experimental campaigns.

The differences among the three datasets in terms of explored regions in the operational space are shown in figure 1 where, from top-left to bottom-right, the distributions of the plasma current (I_p), the toroidal field (B_T), the normalized beta (β_N), the total input power, the line integrated density, and the edge safety factor (q_{95}) are reported for the regularly terminated discharges. It can be noted that the datasets I and II share the same parameter ranges even if, for some parameters (such as I_p, B_T or q_{95}), their distributions slightly differ. Instead, the dataset III, which is related to experiments aiming to study the baseline scenario suitable for sustained high D–T fusion power, is characterized by higher currents, density and input power, also exceeding the range of the other two datasets. The model has been trained and validated using a part of the dataset I, by selecting the same 85 disrupted and 70 regularly terminated discharges used in [8]. The remaining pulses of dataset I and all the pulses of datasets II and III, resulting in 108 disruptive and 149 regularly terminated pulses, have been used for testing the model performance and studying its behaviour with unseen data.

The database has been featured by the 1D profiles of electron temperature, density, and radiated power. Together with such information, also the 0D signals of the internal inductance (l_i) and of the mode-lock signal normalized by the plasma current (ML_{norm}) are provided to the model. 1D profiles ought to be included because, as well known, their behaviour is often connected with the development of destabilizing physical mechanism, such as MHD precursors. In [7, 8, 17] 0D peaking factors were synthesized from these profiles to feed the predicting models. On the contrary, the proposed CNN approach allows us to provide the predicting model with the whole spatiotemporal information contained in the electron temperature and density, and radiated power profiles by converting the same set of 1D diagnostics used in [7, 8] in 2D images.

In order to generate the CNN input data, the following steps have been implemented:

(a) Firstly, data from the high-resolution Thompson scattering (HRTS) for the electron temperature (T_e) and density (n_e), the horizontal lines of sight of the bolometer for the radiated power (P_{rad}), together with l_i and ML_{norm} have been causally resampled with a sampling time of 2 ms. This means that the resampling is performed using only current and past inputs, which is the only option for real-time implementation. This operation allows the system to
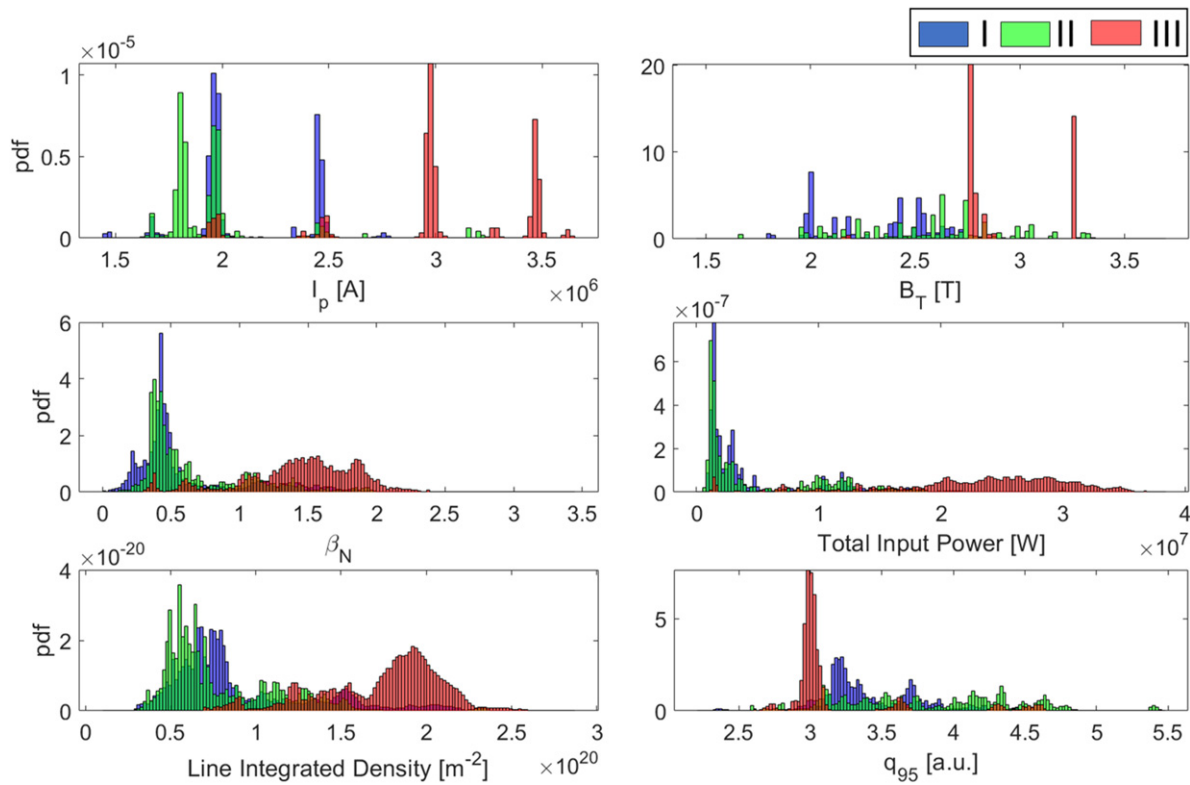
Figure 1. Distributions of the main parameters of the regularly terminated discharges in the dataset I (blue), dataset II (green) and dataset III (red) for (from top left to bottom right): plasma current, toroidal field, normalized beta, total input power, line integrated density and edge safety factor.

work with signals at the same time scale, as these diagnostics have different sampling times, which vary from 10^{-4} to 10^{-2} s. Note that, the raw measures from these diagnostics are used and no inversion procedure is necessary in the proposed approach.

(b) Once the resampling has been carried out, the 1D profile data is processed to provide the CNN with a set of input images, as detailed in the following and sketched in figures 2(a)–(c):

1. A pre-processing is applied to each diagnostic to remove outliers. In particular, for the HRTS diagnostic, the pre-processing consists in the comparison of the measurement with the diagnostic estimated error [30]. As some shots, both for HRTS and bolometer profiles, presented corrupted measures, a pre-processing procedure has been developed, based on the correlation between the measure of each line of sight and those of their neighbours. The corrupted measures are replaced by the interpolated values between the closest ones. From an inspection of the training dataset, the outer 9 lines of sight (from major radius greater than 3.78 m) of the HRTS are discarded as, at least on the selected dataset, they usually provide unreliable data. For the bolometer data, no estimation of the measurement error was available, so negative power values have been substituted with null values,

whereas unreliable positive ones are saturated to a fixed threshold empirically fixed to 1 MW m^{-2};

2. For the HRTS diagnostics, the lines of sight are ordered from the inner ($R = 2.96$ m) to the outer one ($R = 3.78$), where R is the major radius. For the bolometer diagnostic, the lines of sight are ordered as labelled in figure 3. Then a spatiotemporal matrix is built, whose elements assume the value of the measure in the corresponding line of sight and the corresponding time sample. The obtained images are shown in figure 2(b);

3. The three images are vertically stacked, and their ranges are normalized with respect to the signal ranges in the training set. After retrieving the maximum and the minimum values from each diagnostic in the training set, the value x is normalized between [−1, 1] by

$$x_{\mathrm{norm}} = \frac{(2x - x_{\max} - x_{\min})}{x_{\max} - x_{\min}}$$

obtaining the final image in figure 2(c);

4. The final image is segmented using an overlapping sliding window of 200 ms, as sketched by the dashed black line in figure 2(c). In this way, each segment corresponds to an image of 132 × 101 pixels.

(c) The 0D signals (l_i and ML_{norm}) are also sampled at the same time samples as the 1D data and at the same sampling frequency. The two arrays of 101 samples are fed to
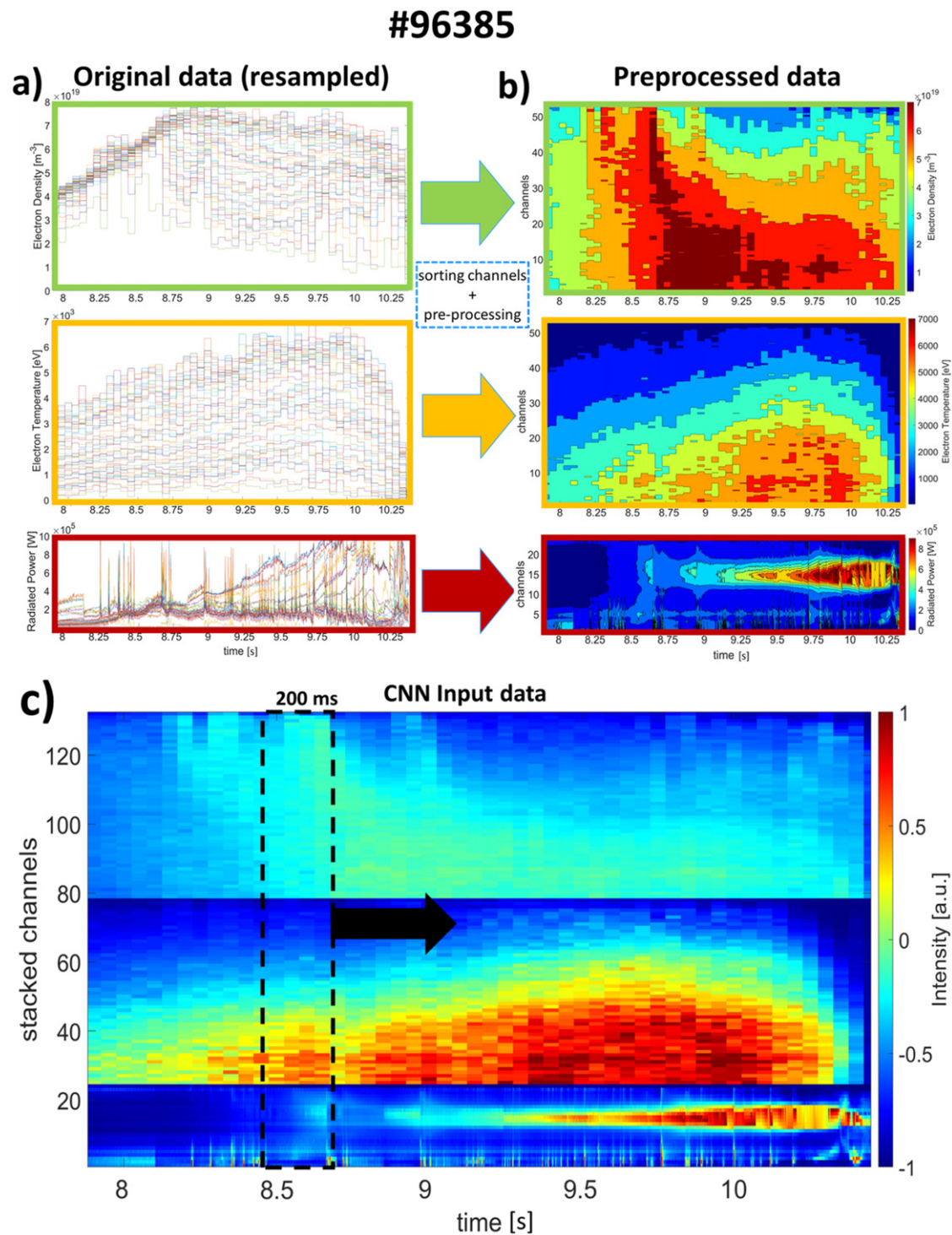
Figure 2. Sketch of the pre-processing steps applied to pulse #96385 to generate the input images for the CNN: (*a*) original data from the HRTS and bolometer diagnostics; (*b*) pre-processed data are converted into images; (*c*) input data for the CNN model, obtained by normalizing the data at point (*b*) with the training set ranges and by vertically stacking the lines of sight. An overlapping window of 200 ms produces the segmented images that feed the prediction model.

the prediction model together with the previous images. The internal inductance, which is intrinsically indicative of the peaking of the plasma current, comes from the EFIT equilibrium code, whereas the mode-lock signal is measured by a set of saddle flux loops, located at radial positions, above and below the middle plane, and mounted on

the outside of the vacuum vessel at the low-field side of the plasma [31].

As the CNN is a supervised algorithm, during the training, a label has to be explicitly assigned to the time windows (or time slices) in the dataset. All the segments belonging to the regularly terminated discharges have been labelled
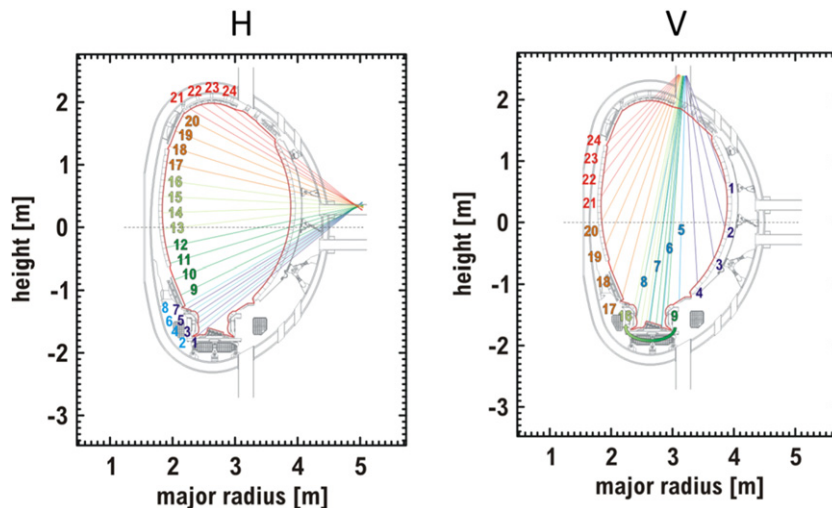
Figure 3. View of the JET bolometer camera system: horizontal camera (left side) and vertical camera (right side).

as 'stable'. For each disruptive discharge, the labelling of the 'unstable' has been carried out by automatically identifying the pre-disruptive phase by means the algorithm proposed in [8]. The algorithm is based on a statistical analysis of the following six dimensionless plasma parameters computed for a selection of disrupted and regularly terminated discharges performed at JET during the experimental campaigns from 2011 to 2013: peaking factor of temperature, peaking factor of electron density, peaking factor of the radiation (excluding the contribution of the X-point/divertor region), peaking factor of the radiation (excluding the contribution of the core region), internal inductance, fraction of the radiated power. These parameters are the same used in [7, 15], and have been defined in section 1. In the algorithm, it is assumed that, before the onset of the chain of events leading to disruption, the distributions of the selected parameters in the disruptive discharges are close to those of the regularly terminated ones, whereas they become more and more dissimilar while approaching the disruption time. In particular, a similarity measure between distributions is used, and the contribution of each input feature is weighted in order to construct a unique warning time indicator (WTI). The study of the WTI distribution in the regular discharges allowed us to optimize a coherent threshold value for the identification of the pre-disruption times $T_{\text{pre-disr}}$. The identification of a pre-disruption time instants $T_{\text{pre-disr}}$, different for the different disruptions, is indeed very beneficial, as demonstrated in literature [7, 8, 15]. Note that, for the disrupted pulses, only the segments belonging to the pre-disruptive phases are used during the training of the model.

In order to reduce the unbalance between the stable and unstable classes, caused by the different duration of the two phases, the overlap times of the sliding window for the regularly terminated and disrupted discharges have been differently chosen. Due to the low time resolution of the HRTS, only one segment every 24 ms has been extracted from the pre-disrupted phase of the disrupted discharges, whereas one segment every 150 ms has been retained from the regularly terminated discharges. Note that, during testing, a sliding window of 200 ms

Table 2. Number of pulses and time slices in the training, validation and test sets.

Set	Disruptions		Regular pulses	
	Pulses	Time slices	Pulses	Time slices
Training	63	3698	54	4239
Validation	22	1191	16	1381
Test	108	313 392	149	588 143

with a stride of 2 ms, for all discharges (regularly terminated and disrupted), has been used.

Table 2 reports the total number of pulses and time slices sampled for the train, validation and test sets. The validation set was used to monitor the training performance during the training and to perform an early stop if the performance on the validation data would not improve. The validation set discharges were randomly sampled among the 85 disrupted discharges and the 70 regularly terminated ones from dataset I.

3. Disruption predictor architecture

In recent years, the use of deep learning in research has increased significantly, due to the improved capability of computers in processing huge amounts of data and to the ability of deep neural networks in producing high accuracy performances even without a feature extraction procedure. Among the architectures in deep learning able to process images, CNNs are the most used [32, 33]. The deep architecture of a CNN normally consists of a cascade of blocks of different layers which performs a filtering of an input image to extract significant features from it [21]. The features are produced by a cascade of filtering blocks, interconnected through nonlinear activation functions (typically a rectified linear unit), and a multi-layer perceptron combines them to produce the output of the network. A dropout layer is usually inserted before the multi-layer perceptron in order to reduce overfitting on the training set and improve generalization.

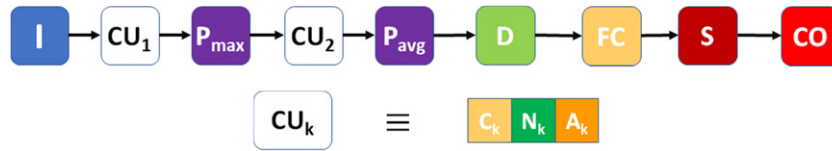The architecture of the proposed CNN is shown in figure 4.

Figure 4. CNN architecture, where: I is the image input; CU_k is the kth convolutional unit, composed by the cascade of a convolutional layer (C_k), a batch-normalization layer (N_k) and a nonlinear activation layer with ReLU functions (A_k); P_{max} and P_{avg} are the max-pooling and average-pooling layers, respectively; D is a dropout layer; FC is a fully-connected layer; S and CO are the SoftMax and classification output layers, respectively.
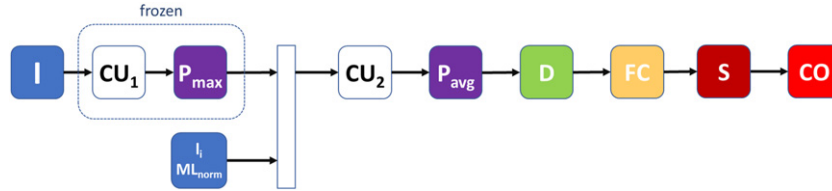
Figure 5. Modified CNN architecture, where the internal inductance (l_i) and the normalized locked mode (ML_{norm}) are added as input to the second convolutional unit and concatenated with the output image produced by the max pooling layer.

Table 3. Training parameters of the CNN model.

| Parameters	1st training (images only)	2nd training (images + signals)
Optimizer	Stochastic gradient descent with momentum	Stochastic gradient descent with momentum
Initial learning rate	2.5×10^{-4}	1×10^{-4}
Learning rate drop factor	0.1	0.15
Learning rate drop period (epochs)	20	10
Momentum	0.9	0.9
MiniBatch size	16	16
Validation frequency (iterations)	50	50
Validation stop (consecutive evaluations)	100	100
Weight decay (L2 regularization)	1×10^{-4}	1×10^{-4}
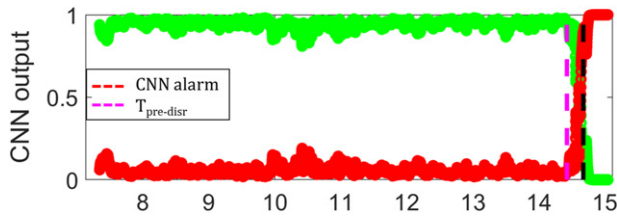
Figure 6. CNN likelihood curves for a disrupted pulse, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood. The dashed black line indicates the CNN alarm time, whereas the dashed magenta line indicates $T_{pre-disr}$.

Table 4. CNN predictor performance.

Dataset	SP	MA	FA
Train	98%	0%	4.28%
Test	93%	3.7%	9.40%

A first convolutional unit (CU_1) followed by a max pooling layer (P_{max}), with pool size and stride 8×1, filters out vertically (along the 'spatial' dimension) the input image by reducing the size from 132×101 to 16×101. A second convolutional unit (CU_2) followed by an average pooling layer (P_{avg}), with pool size and stride 1×12, filters out horizontally (along the 'time' dimension) the resulting image by reducing the image size to 16×20. The two convolutional units (CU_1 and CU_2) are made out of three layers: a convolutional layer (C_k), a batch normalization layer (N_k) and a rectified linear unit (ReLU) activation layer (A_k). The two convolutional layers have one single filter (one-channel kernel) of size 5×1 and 1×11, respectively. The output of the 2nd convolutional layer is then a 16×20 image, which is flattened and provided as input to a fully connected layer (FC). Finally, the FC layer processes the 320 features and feeds a SoftMax layer (S) for classification (CO). A dropout layer with dropout probability of 20% has been included before the fully connected layer in order to reduce overfitting on the training set and improve generalization.

In order to include also the information given by the two 0D signals, i.e., l_i and ML_{norm}, two segments of size 1×101 have been added as input to the second convolutional unit and concatenated with the output image produced by the max pooling layer (see figure 5). As a result, the output of the average pooling layer has size 18×20 and 40 additional features coming from the two signals are processed by the fully connected layer, combined with the remaining 320 features and used for the classification.

The soft max (S) layer produces the likelihood of the input segment to belong to a regularly terminated or a disrupted discharge. As an example, figure 6 shows the soft max output for
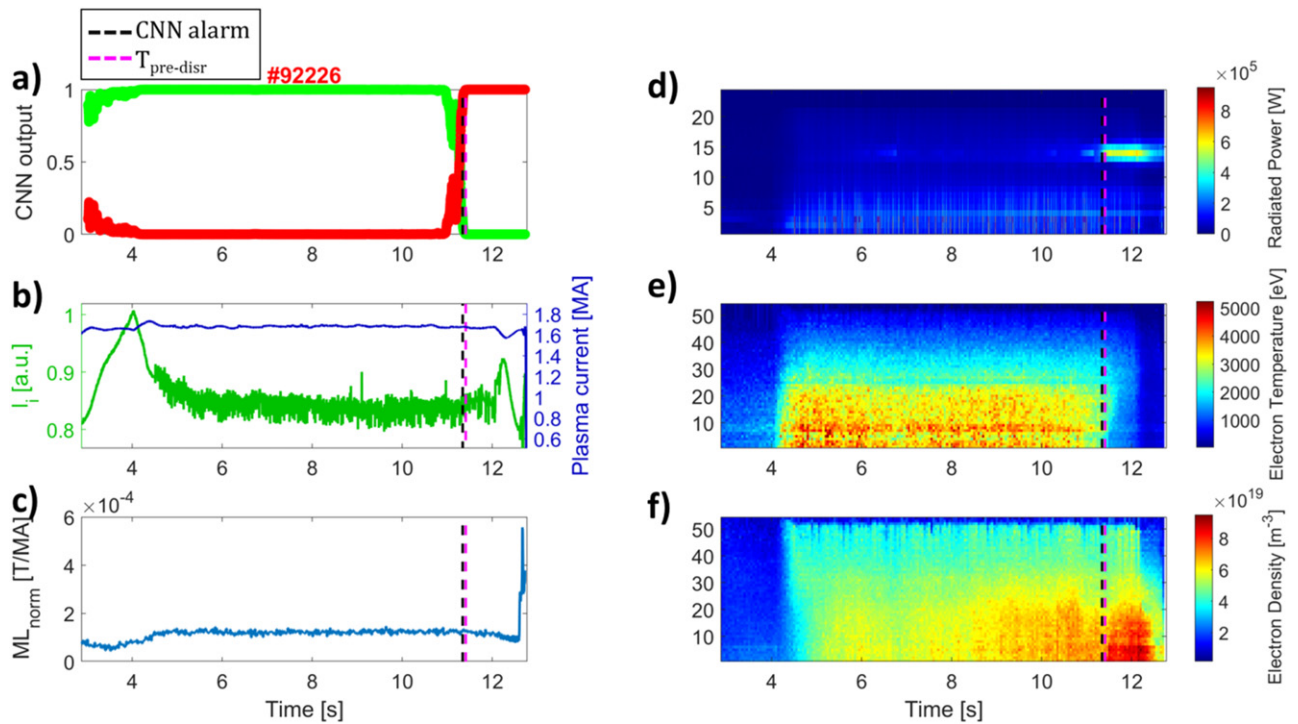
Figure 7. JET disrupted discharge #92226. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood. The dashed black line indicates the CNN alarm time, while the dashed magenta line indicates the $T_{\text{pre-disr}}$; (*b*) internal inductance, in green, and plasma current in blue; (*c*) mode lock normalized by the plasma current, in blue; (*d*) radiated power from the bolometer; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS.

a JET disrupted pulse, where the green line refers to the likelihood of a segment to belong to a regularly terminated pulse and the red line refers to a disrupted one. The two curves add up to one, and it can be noticed that the red line starts to rise in correspondence of the $T_{\text{pre-disr}}$ (dashed magenta line) and then it straightforwardly reaches the value 1. The last classification layer (CO) implements a threshold on the disrupted likelihood to perform the final classification. Such alarm threshold has been optimized by means of a heuristic procedure, maximizing the number of correct predictions on the training and validation discharges. The optimal threshold is found to be 0.89 and the alarm time is triggered when the disrupted likelihood overcomes such threshold. In figure 6 the alarm time is identified by the black vertical dashed line.

In order to limit the complexity of the training procedure a first training step has been performed using only the 1D diagnostics. Then, the CU_1 and the P_{max} blocks were frozen. In a second training procedure, only the second convolutional block and the fully connected layer were trained, using both the 1D and 0D diagnostics. This approach greatly reduces the computation time of the training procedure because it reduces the number of parameter updates being made without having a major impact on the network's accuracy [34]. Table 3 shows the hyperparameters of the two training procedures.

Note that, the network architecture allows us to uncorrelate the two dimensions, spatial and temporal: in fact, the first two blocks (CU_1 and P_{max}) filter only across the spatial direction, while the second two (CU_2 and P_{avg}) filter only across time. This allows to easily concatenate the signals

(l_i and ML_{norm}) to the image features processed by the first convolutional and pooling blocks, so that the temporal synchronization is preserved.

The vertical kernel size for the convolutional and pooling blocks was designed considering a few constraints: a kernel size equal or larger than 24 would have been larger than the bolometer number of lines of sight, and a small size kernel would reduce the effect of the discontinuity between the stacked diagnostic images. The small kernel size (5×1) allows the network to still identify changes in the spatial dimension of the HRTS scattering profile. Regarding the time filtering, a similar operation was performed: due to the different time resolution of the diagnostics employed, the filter size has been chosen to mainly process the higher frequency signals, such as the bolometer data.

The pooling type was optimized: a network with only average pooling was trained, as well as one with only the max-pooling. Analysing the performances on the training and the validation set, the average pooling had lower performance than the max-pooling, but the max-pooling response was too sensitive to transient changes in the data time traces. Hence, the max-pooling layer was left in the spatial processing block (vertical pooling), while the average pooling was selected for the temporal pooling.

It has to be highlighted that, as the input image in figure 2(*c*) concatenates the 1D signals coming from three diagnostics. Following Baltrusaitis *et al* [35], the proposed approach can be classified as a joint multimodal representation (also named early fusion as opposed to late fusion). Joint representations
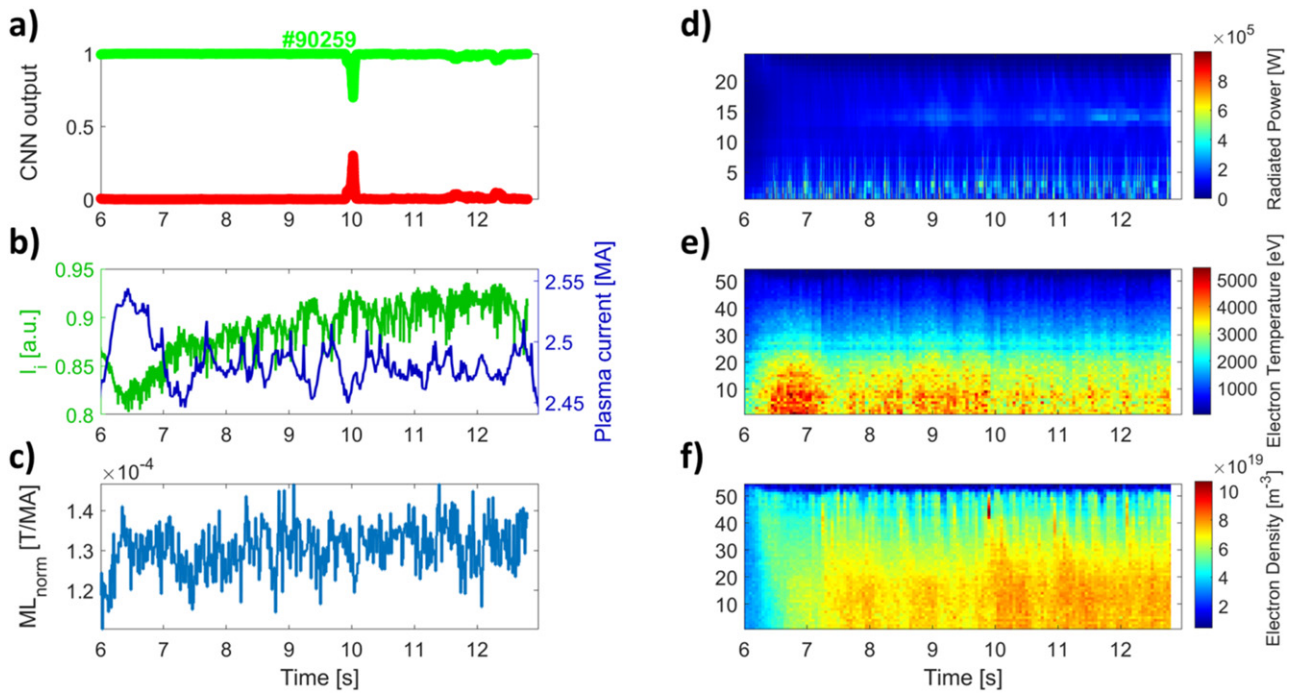
Figure 8. CNN output of the JET regularly terminated discharge #90259. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood; (*b*) internal inductance, in green and plasma current in blue; (*c*) mode lock normalized by the plasma current, in blue; (*d*) radiated power from the bolometer; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS.
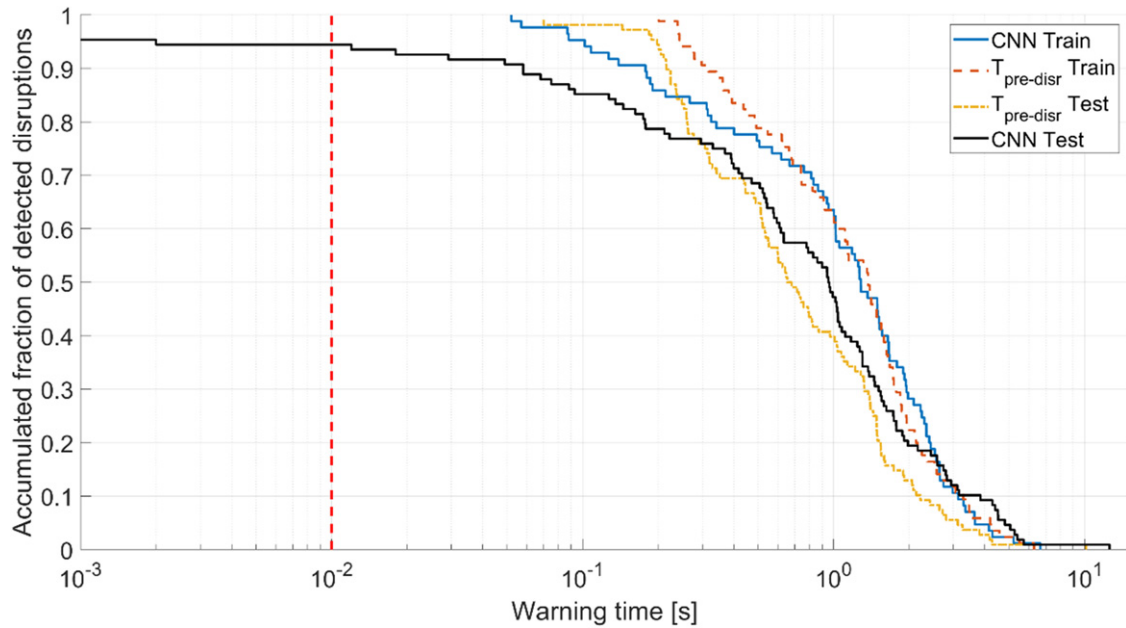

Figure 9. Cumulative fraction of detected disruptions by the CNN model versus the warning time in the training and in the test set. The vertical red dashed line allows us to identify tardy detections.

combine the unimodal signals into the same representation space. Mathematically, it is expressed as: $x_m = f(x_1, \ldots, x_n)$, where the multimodal representation x_m is computed using the CNN that relies on unimodal representations $x_1, \ldots, x_n$.

Conversely, the late fusion representation processes unimodal signals separately and then the results are merged. Early fusion has the advantage of merging data sources in the beginning of the processing (sometimes after a first convolution). If

Table 5. Performance of the CNN on the test discharges across the three datasets.

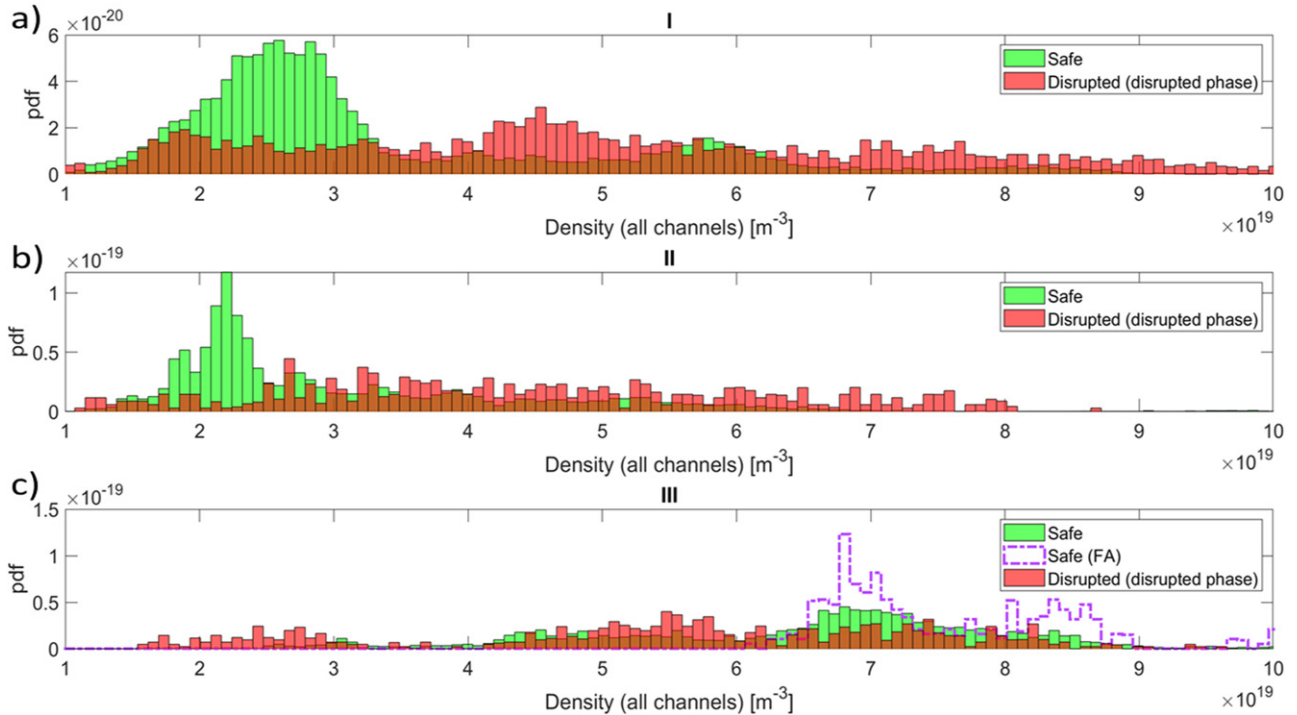| Dataset I			Dataset II			Dataset III		
SP	MA	FA	SP	MA	FA	SP	MA	FA
96.55%	2.38%	4.44%	91.42%	10.34%	7.31%	91%	0%	14.28%

Figure 10. (*a*)–(*c*) Comparison of the probability density functions of the electron density computed as the mean value across the lines of sight of the HRTS diagnostic for the regularly terminated (green) and the disrupted discharges (red) in the three datasets: (*a*) dataset I, (*b*) dataset II, and (*c*) dataset III. For the dataset III, the distribution of false alarm values is identified by a magenta dashed line.

the data is properly aligned, cross-correlations between data items may be exploited, thereby providing an opportunity to increase the performance of the system. In [36], the authors argue that those fused low-level features might be irrelevant for the task, thus decreasing the fusion power.

When signals from different modalities do not complement each other, i.e., input modalities separately inform the final prediction and do not have any inherent interdependency, then trying an another fusion approach is preferred [37]. Moreover, late fusion retains the ability to make predictions in case of missing or incomplete data, because it employs separate models for separate modalities, and aggregation functions can be applied even when predictions from a modality is missing. The major drawback is the limited potential for the exploitation of cross correlations between the different unimodal data.

The optimum fusion strategies for many applications have yet to be determined [38]. In a recent review on fusion techniques for deep learning applications in medicine [37], the authors report that in most applications early fusion is used as the first attempt, a straightforward approach that does not necessarily require training multiple models.

Note that, in our case, convolutions involving the boundaries are not really senseful and the network discards the corresponding features. On the contrary, the model can learn the associations across the informative regions of the multiple images. Moreover, the idea to train only one CNN, instead of several, allows a simpler implementation of the model.

4. Predictor performance

In this section, the potentialities of the CNN model to detect a disruptive behavior early enough to enable avoidance actions are presented. In the disruption prediction literature, the performance of a predictive model is evaluated in terms of:

- Successful predictions (SP): pulses correctly predicted by the system (alarm for disruptions and no alarm for regularly terminated discharges);
- Missed alarms (MAs): disruptions for which the system does not provide any alarm;
- False alarms (FAs): regularly terminated discharges for which the system provides an alarm

The performance in terms of SPs, MAs, and FAs rate of the proposed predictor is reported in table 4, for a training set

composed by 63 disruptive and 54 regularly terminated pulses and a test set of 108 disruptive and 149 regularly terminated pulses.

Table 4 highlights that the CNN has very good predictive performance with a successful prediction rate of 93% and a FA rate below 10% on the test set.

An example of the CNN capability in predicting disruptions can be seen in figure 7 which refers to the test pulse #92226 (outside the training range). The CNN output in figure 7(*a*) identifies a rising of the disruptive likelihood at about 11.20 s, accordingly with the visible change of the plasma behavior across the input profiles shown in figures 7(*d*)–(*f*). It can be noted that, at the plasma core, the electron temperature collapses (figure 7(*e*)) while the electron density peaks (figure 7(*f*)). This phenomenon is accompanied by strong radiation (figure 7(*d*)). This pattern is well-known as the impurity accumulation disruptive mechanism [15, 19], and it is typical of the JET ILW disruptions due to the penetration of high-Z impurities (such as the tungsten of the JET divertor) into the plasma core [18].

A typical regularly terminated discharge, as the one shown in figure 8, is usually characterized by very regular radiated power profiles with low radiation from the central chords of the bolometer horizontal camera (lines of sight 13–16, figure 3, as visible in figure 8(*d*). At the plasma core, the electron temperature profile peaks (figure 8(*e*)), while the electron density, distributed across the profile, presents slightly higher values (figure 8(*f*)). In addition, l_i and ML_{norm} do not reveal any approaching disruption. In agreement, the regularly terminated likelihood of the CNN, green curve in figure 8(*a*), stays high for the whole pulse and the disrupted likelihood never reaches 0.5.

Nowadays, disruption prediction systems are being developed especially for avoidance purposes; for a disruption, the goal of an avoidance system is to associate the alarm to the presence of a destabilizing mechanism in the plasma, regardless of the distance of such event to the ending time t_{end}. Hence, metrics such as premature alarms, defined as the alarms triggered at a prefixed time before the disruption, have become less significant in the definition of the system performance. Instead, the prediction capability within the scope of avoidance and/or disruption control can be evaluated in terms of warning time, which represents the distance of the model alarm from t_{end}. A well-timed warning time allows the control system to react to the presence of an instability, while with a short warning time the disruption is generally mitigated by MGI. Thus, the premature alarm rate is replaced by the cumulative warning time distribution (see figure 9). The blue and black lines in figure 9 show the CNN warning times in the training set and test set respectively. Moreover, to evaluate the suitability of the alarm triggered by the CNN predictor, the predicted warning time is compared with respect to the one defined by $T_{pre-disr}$. In this regard, in the same figure 9, the dark red and yellow dashed lines report the $T_{pre-disr}$ warning time distribution for the training and test pulses, respectively. The vertical red dashed line indicates the minimum warning time (10 ms) necessary at JET to adopt mitigation actions. Detections made after this line can be considered as tardy alarms.

From figure 9, it is possible to see that the CNN warning times and $T_{pre-disr}$ ones are quite close, both for the training and the test set. This means that, in most of the cases, the CNN detections are coherent with the instability mechanisms automatically detected with the $T_{pre-disr}$.

As described in section 2, the database employed in this work includes discharges from several experimental campaigns and different experimental conditions. This motivated an in-depth analysis of the results, to investigate a possible degradation of the CNN performances with the changing of the operating conditions. Table 5 reports the SPs, MAs and FAs rate for the test discharges, among the three datasets. As expected, the best performances are reached on the dataset I, which covers the same pulse range of the training data. Note that, the CNN predictor tested on dataset II still performs quite well, with a FA rate lower than 7%, whereas an increase in the MA is observed. The degradation of the MA rate is due to the presence of some discharges which disrupt very abruptly because of a sudden locking mode, as already observed in [8] for the same data set. Conversely, in the dataset III the errors on the disrupted pulses are extremely low, whereas the false alarm rate is the highest. An explanation for this difference should be sought in the new region of the operational space covered by the regularly terminated pulses belonging to this dataset, which are characterized by higher input power and electron density (as highlighted in figure 1). Indeed, by comparing the distributions of the features provided to the model in the three datasets, the regularly terminated pulses in the dataset III are characterized by higher n_e and radiated power values and lower l_i values. In figures 10(*a*)–(*c*)–12(*a*)–(*c*), the probability density functions of the average density across the plasma radius, l_i and the average radiated power are reported for the three considered datasets, for regularly terminated (green) and disruptive (red) pulses, respectively. In addition, for the dataset III, the related distributions of the false alarms are added to the regularly terminated and disruptive ones (see figures 10(*c*)–12(*c*)), and identified by a magenta dashed line. It can be noted that the distribution values of the three features for the regularly terminated pulses in dataset III are shifted with respect to the ones of the previous datasets. Considering that the full training set of the model is contained in dataset I, the distribution of the values in the dataset III is then covering ranges poorly represented by the non-disruptive behavior of the training set. This trend is confirmed or accentuated by the distribution of the feature values related to FAs. Moreover, during high power experiments the presence of localized radiation in the outer half of the plasma, not necessarily correlated to the onset of a disruptive mechanism, has been observed [39]. This phenomenon can play a crucial role in the erroneously detection of a disruptive behavior in a regularly terminated pulse, as in the pulse #94785 reported in figure 13. Indeed, at around 11.5 s, despite a non-disruptive behavior shown by the HRTS profiles and the 0D signals, a high radiation seen from the central lines of sight of the bolometer horizontal camera triggers a FA (see figure 13(*d*)). From the radiation profile recorded by the bolometer vertical camera (not provided as input to the CNN) it is possible to localize the radiation blob in the outer half of the plasma, thus not related to any impurity accumulation.
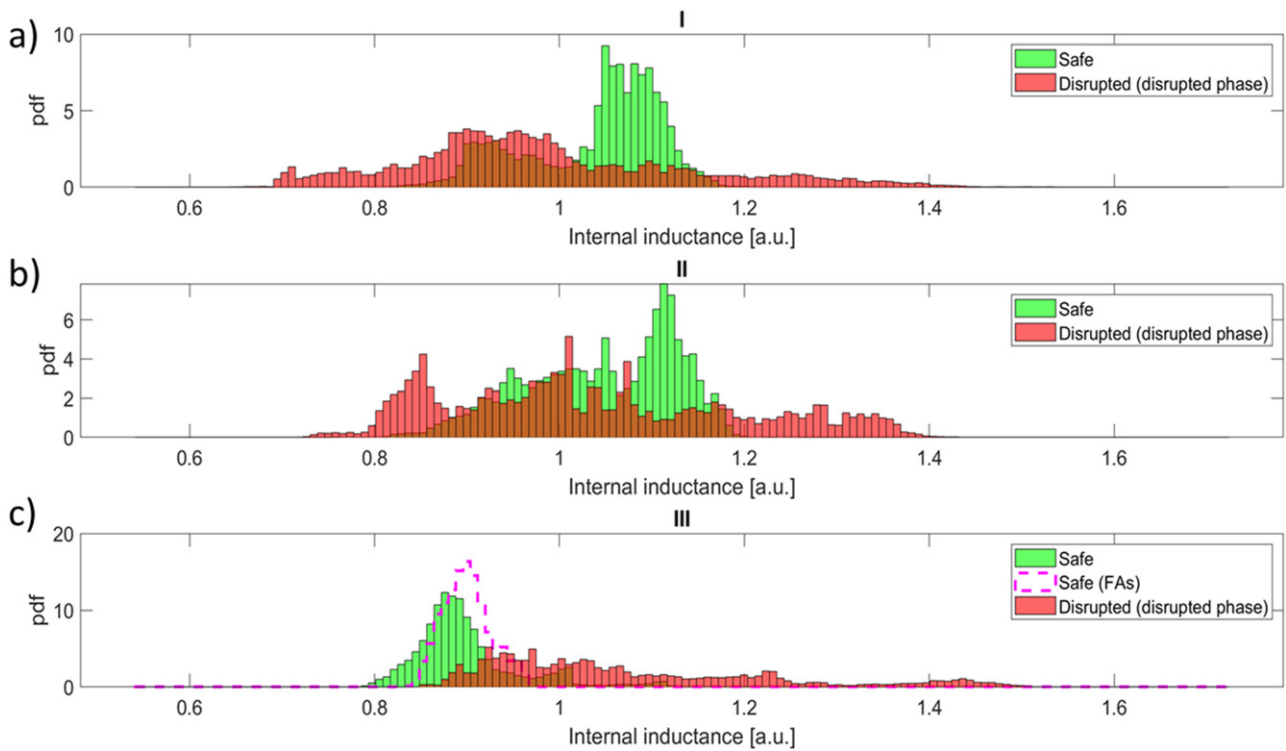
Figure 11. (*a*)–(*c*) Comparison of the probability density functions of the internal inductance for the regularly terminated (green) and the disrupted discharges (red) in the three datasets: (*a*) dataset I, (*b*) dataset II, and (*c*) dataset III. For the dataset III, the distribution of false alarms values is identified by a magenta dashed line.
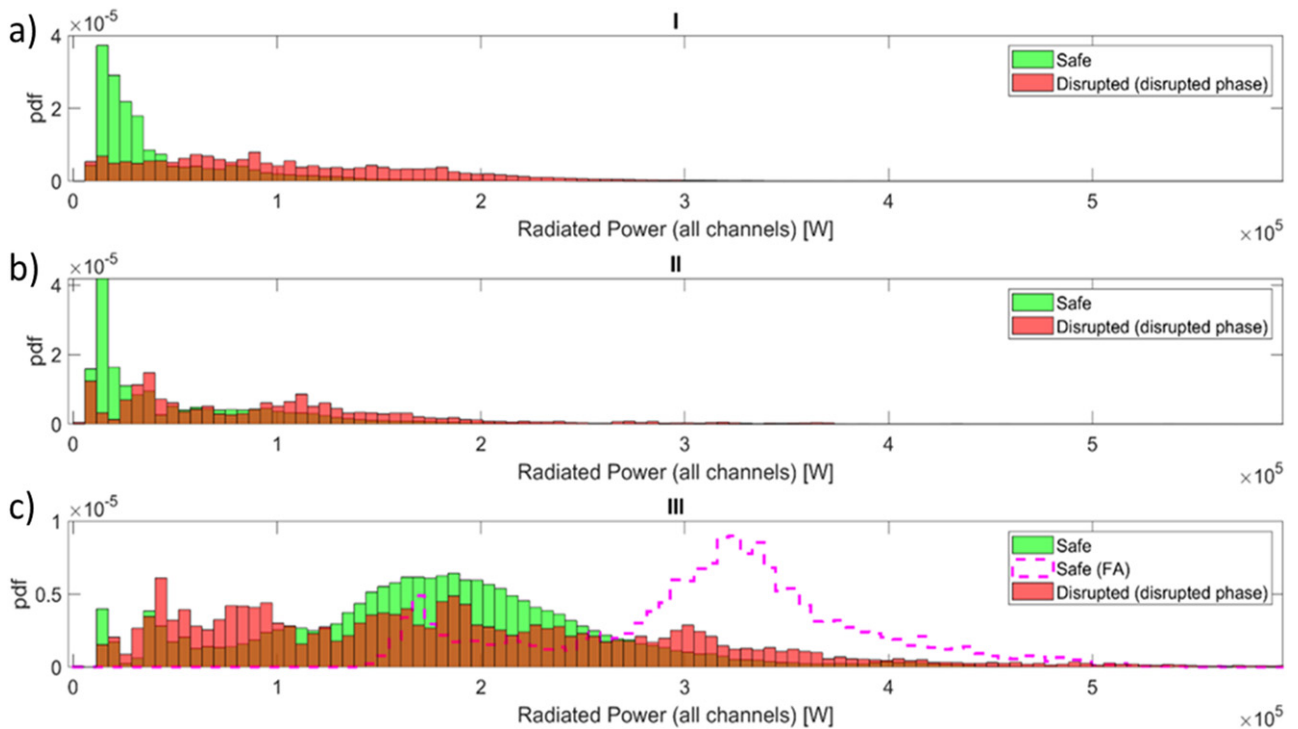

Figure 12. Comparison of the probability density functions of the radiated power computed as the mean value across the lines of sight of the HRTS diagnostic for the regularly terminated (green) and the disrupted discharges (red) in the three datasets (from (*a*) to (*c*)). For dataset III, the distribution of false alarms values is identified by a magenta dashed line.
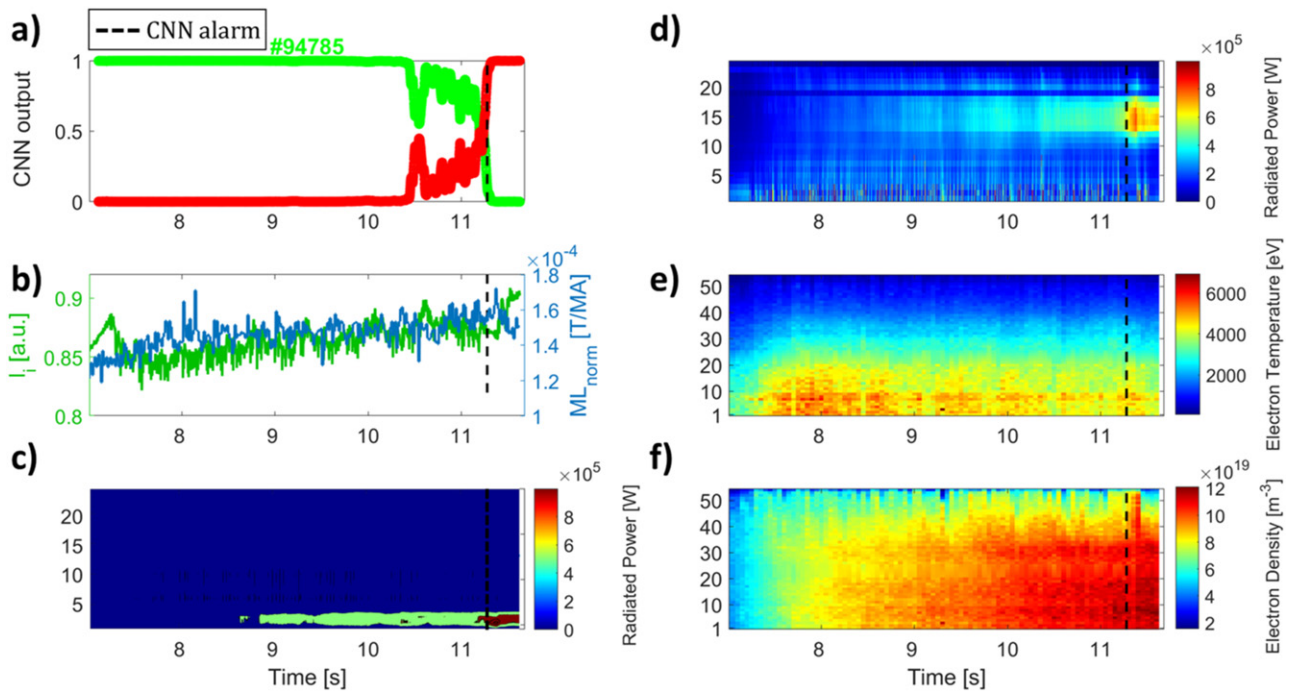
Figure 13. CNN output on the regularly terminated discharge #94785. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood; (*b*) internal inductance in green and mode lock normalized by the plasma current in blue; (*c*) radiated power from the bolometer vertical camera; (*d*) radiated power from the bolometer horizontal camera; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS. The dashed black line indicates the CNN alarm time.
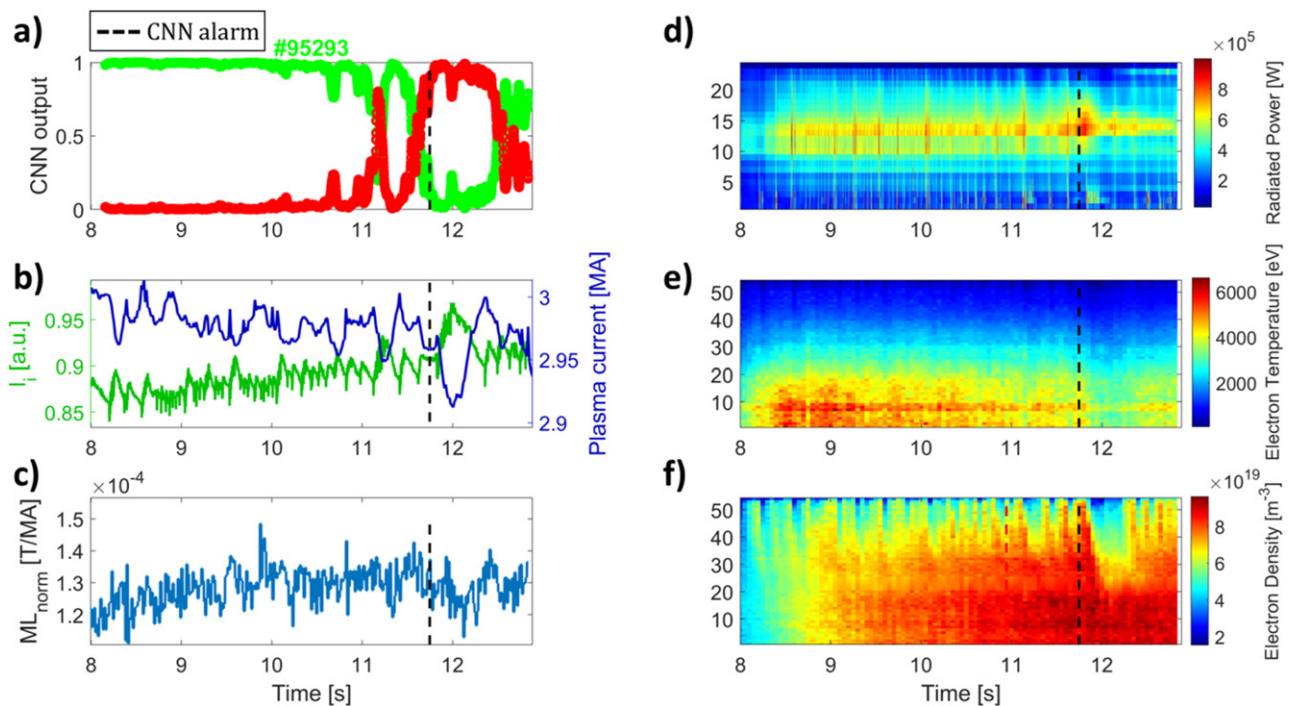

Figure 14. CNN output on the regularly terminated discharge #95293. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood; (*b*) internal inductance in green and plasma current in blue; (*c*) mode lock normalized by the plasma current, in blue. (*d*) Radiated power from the bolometer horizontal camera; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS. The dashed black line indicates the CNN alarm time.

Figure 14 reports the response of the CNN to the regularly terminated pulse #95293 from dataset III. As it can be seen, the CNN triggers a FA nearby a high radiation from the central lines of sight of the bolometer horizontal camera, together with the decrease of both the electron temperature and the peaking of the electron density at the core. On the other hand, this pattern causes a high number of the false alarms observed in this dataset; in fact, it is very similar to the disruptive mechanism
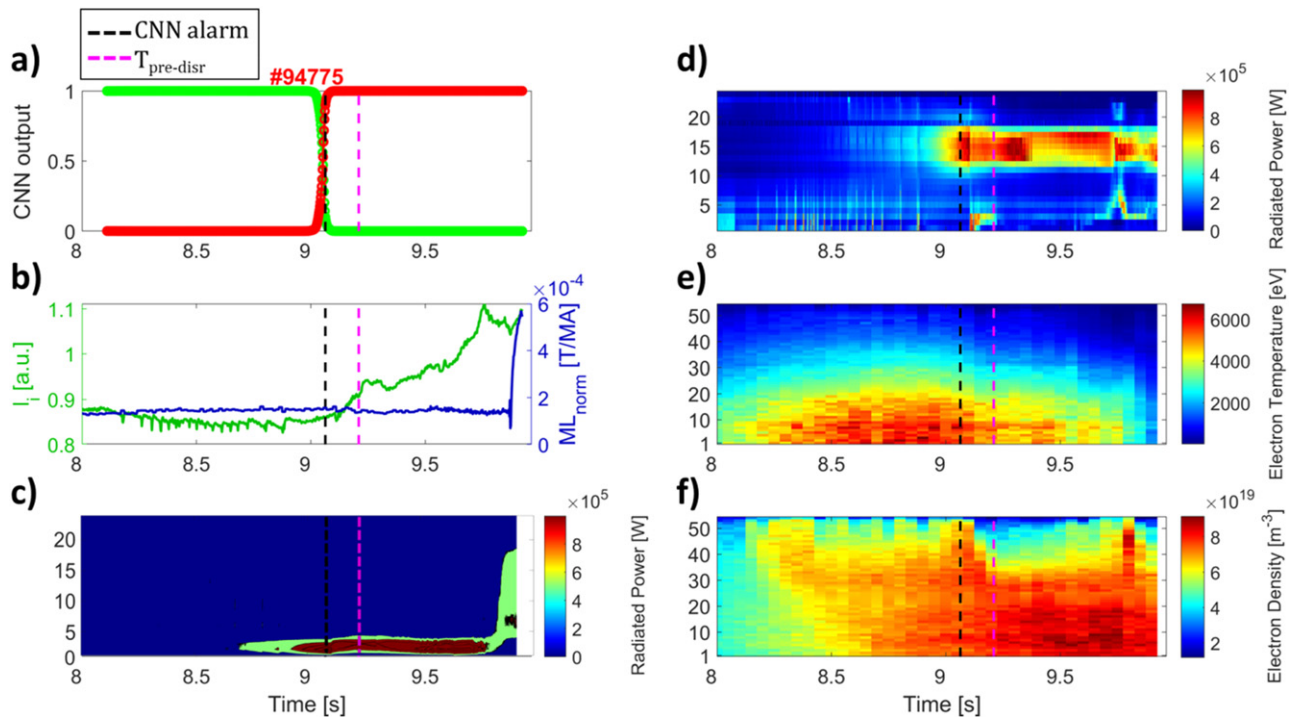
Figure 15. JET disrupted discharge #94775. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood. The dashed black line indicates the CNN alarm time, while the dashed magenta line indicates the $T_{\text{pre-disr}}$; (*b*) internal inductance, in green, and mode lock normalized by the plasma current in blue; (*c*) radiated power from the bolometer vertical camera; (*d*) radiated power from the bolometer horizontal camera (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS.

Table 6. Performances of the GTM.

Set	SP	MA	FA
Train	98.53%	0%	2.61%
Test	84.11%	1.83%	26.17%

represented in figure 7. It has to be highlighted how, at about 12.5 s, following the temperature and the density flattening over the plasma profile, the CNN predictor reports a gradual reduction of disruption likelihood.

Instead, figure 15 shows a disrupted pulse belonging to the dataset III where a disruption due to an edge collapse is detected (#94775). Differently from the impurity accumulation, the edge collapse is characterized by the presence of a blob of radiation in the outer part of the plasma. The radiation causes a localized cooling of the plasma temperature which in turns induces the peaking of the plasma current profiles [19]. This mechanism, visible in figure 15, triggers the alarm at around 9 s (black vertical dashed line). In figure 15(*e*) it is possible to see the cooling of the plasma between the HRTS lines of sight 12 and 30 (which corresponds to a radial position from 3.13 m to 3.46 m), together with a high plasma radiation at the central lines of sight of the bolometer horizontal camera (figure 15(*d*)) and the rise of the plasma internal inductance (figure 15(*b*)). The further analysis of the bolometer vertical camera data allows to localize the radiation blob in the outboard of the plasma (between chords 1–5, see figure 15(*c*)).

Despite it is not possible to distinguish between the core radiation and the outer low field radiation only from the bolometer horizontal camera lines of sight, by combining the spatiotemporal information of the HRTS and the bolometer horizontal camera with the l_i signal, the network is able to detect two different 'off-normal' patterns: one characterized by a strong radiation due to an impurity accumulation process (see figure 7) and another one where the radiation leads to a cooling at the edge (see figure 15).

4.1. Comparison with GTM predictor

In literature, the GTM [40] approach has been successfully proposed to solve disruption classification and prediction problems [7, 41]. For the sake of comparison, the CNN prediction model presented in this work has been compared with the GTM disruption predictor, developed by the same authors and presented in [8]. Both predictors have been trained on the same training set and using the automatically identified $T_{\text{pre-disr}}$. Table 6 reports the GTM performance evaluated on the same training and test set referred in tables 1 and 2. It can be noted that, despite the very high number of correctly predicted disruptions, the GTM provides a quite higher number of FAs with respect to the CNN.

Figure 16 reports the GTM output and inputs for the regularly terminated pulse #90259 already shown in figure 8. The GTM output, shown in figure 16(*a*), provides the class membership function as the likelihood of the sample to belong to the regularly terminated (in green) or to the disruptive class
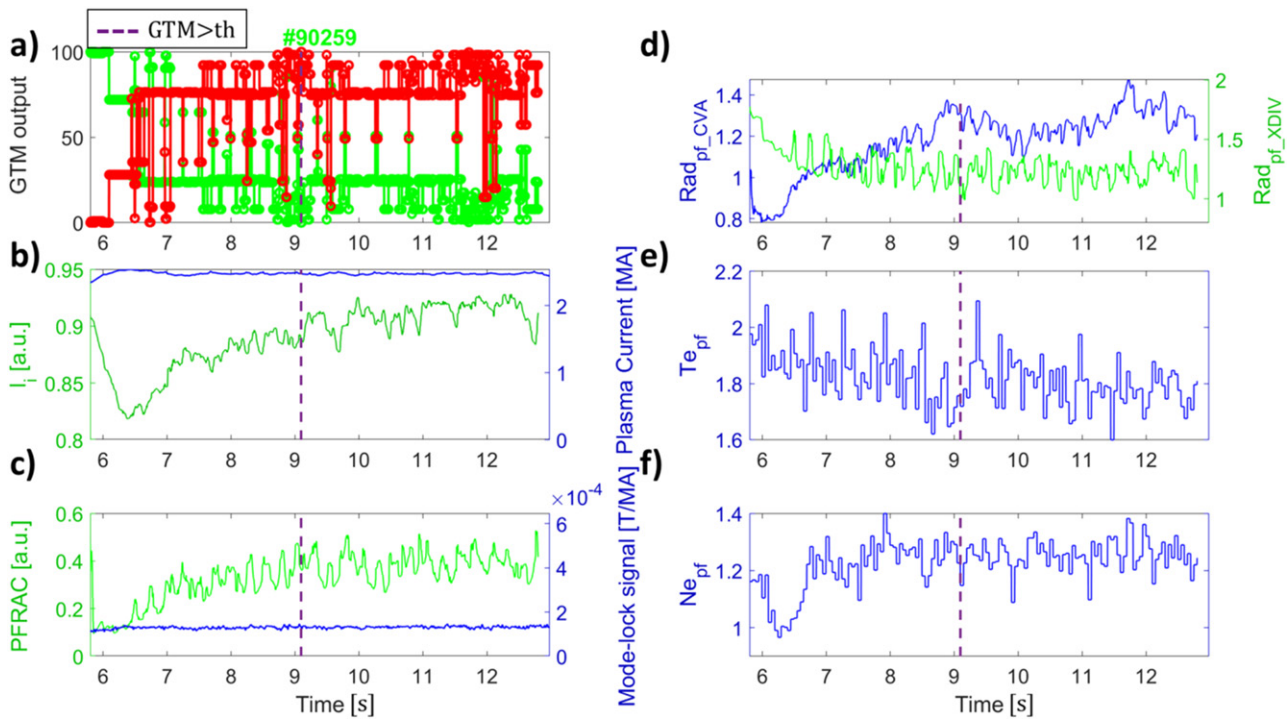
Figure 16. GTM output on the same regularly terminated discharge #90259 of figure 8. (*a*) GTM membership functions, where the green line is the regularly terminated membership and the red line is the disruptive membership; (*b*) plasma current, in blue, and internal inductance in green; (*c*) fraction of radiated power, in green, and mode-lock signal normalized by the plasma current; (*d*) peaking factor of the radiation at the central lines of sight of the bolometer horizontal camera, in blue and peaking factor of the radiation at the divertor, in green; (*e*) peaking factor of the electron temperature; (*f*) peaking factor of the electron density. The vertical purple dashed line indicates when the GTM output rises over the threshold. An assertion time of 60 ms prevents the GTM from triggering a false alarm.
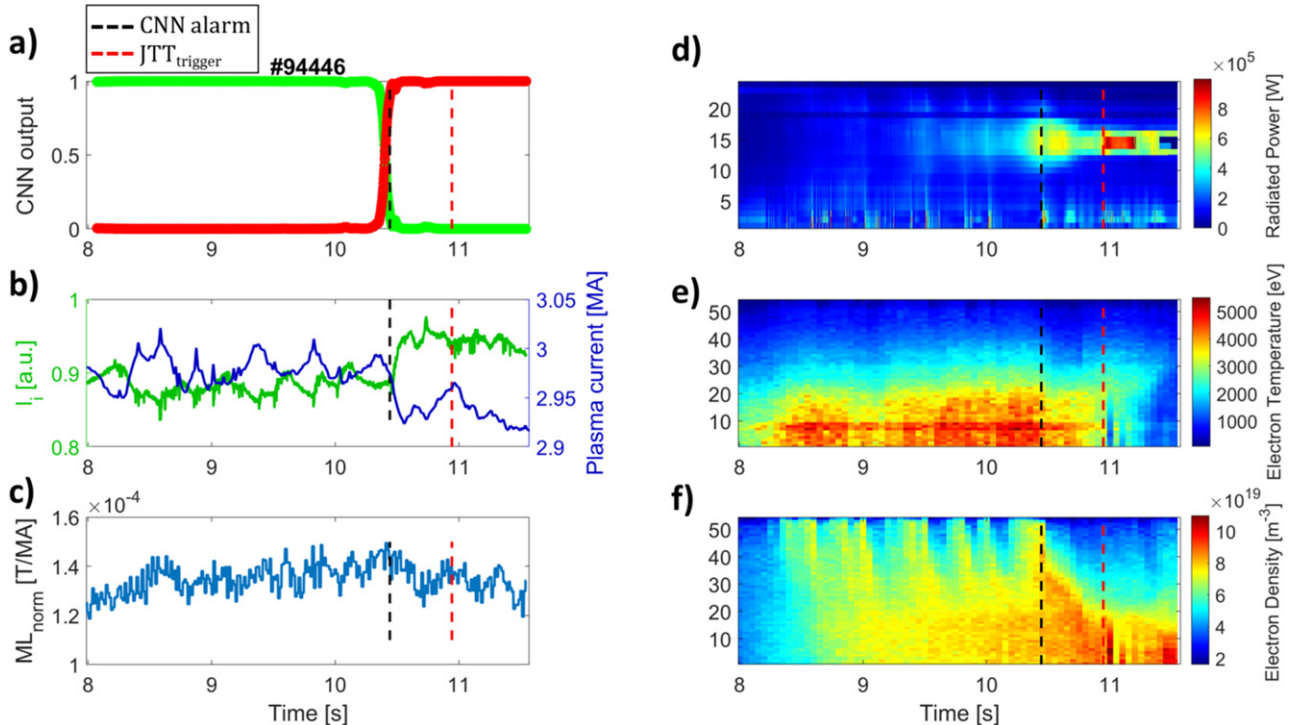

Figure 17. JET discharge #94446. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive one. The dashed black line indicates the CNN alarm time, while the dashed red line indicates $JTT_{trigger}$; (*b*) internal inductance, in green, and plasma current in blue; (*c*) mode lock normalized by the plasma current, in blue; (*d*) radiated power from the bolometer; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS.
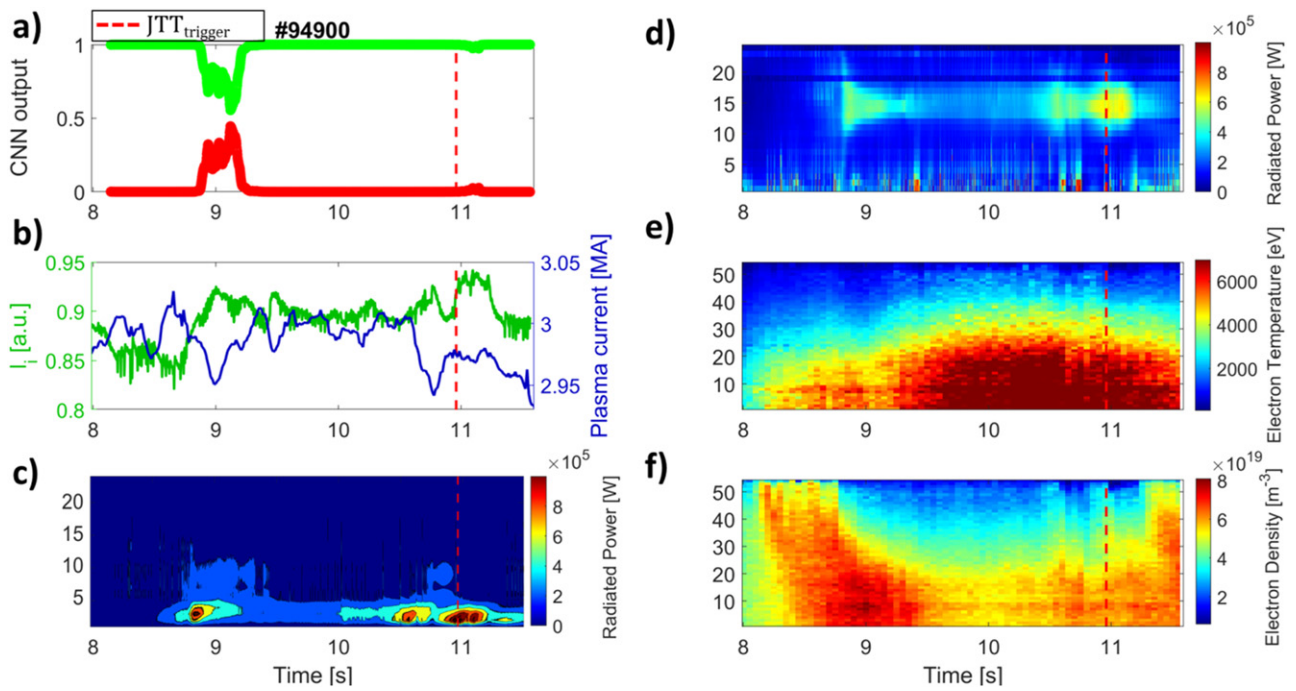
Figure 18. JET discharge #94900. (*a*) CNN likelihood curves, where the green line is the regularly terminated likelihood and the red line is the disruptive likelihood. The dashed red line indicates the time of the JTT$_{trigger}$; (*b*) internal inductance, in green, and mode lock normalized by plasma current in blue (*c*) radiated power from the bolometer vertical camera; (*d*) radiated power from the bolometer horizontal camera; (*e*) electron temperature from the HRTS; (*f*) electron density from the HRTS.

(in red). From figure 16(*a*) it can be seen that the GTM output oscillates, with the disruptive membership function eventually reaching 100% several times. As consequence, in the GTM alarm scheme presented in [7, 8], an assertion time was optimized to avoid false or inappropriate alarms due to a temporary change in the sample classification. In fact, GTM output rises over the alarm threshold at around 9.10 s, as highlighted with a vertical purple dashed line, and without an assertion time of 60 ms [8] a false alarm would have been triggered. Conversely, the CNN output is smoother, and the disruptive likelihood never goes over the threshold; as visible in figures 7(*a*) and 8(*a*), both for a disruption and a regularly terminated pulse respectively, the likelihoods are smooth, and no assertion time is needed for the CNN alarm criterion. Note that, this different behavior of the two models has been observed on the whole database, and it also includes the stable phase of disruptions.

5. Comparison with discharge termination from RTPP

In view of the next step fusion experiments, in the JET real time plasma protection (RTPP) system new algorithms for detecting off-normal events or pre-disruptive states have been implemented. In particular, triggers related to the peaking of the temperature, density and radiated power have been implemented for the detection of impurity accumulation, cooling edge or radiation peaking events [42]. These detectors have been developed with the aim of triggering an early plasma termination of the disruptive discharges.

In 2019, experiments dedicated to plasma termination for disruption avoidance and mitigation tasks have been performed in the same shot range of dataset III. In these experiments, specific user-defined protection schemes were adopted to identify anomalous plasma conditions and to try to recover the pulse through a jump-to-termination (JTT) procedure. In this paper, 31 pulses, fitting the requirements discussed in section 2, have been considered to further test the CNN predictor model. Overall, the CNN triggers an alarm for 22 discharges, for 80% of which the trigger is within 500 ms from the one of the RTPP.

As an example, figure 17 reports the pulse #94446, for which the CNN, as well as the RTPP system, triggers the alarm (JTT$_{trigger}$, red dashed line). Here, the plasma termination is triggered at 10.944 s because of the increase of the plasma radiation at the central lines of sight of the bolometer horizontal camera with respect to the outer ones, measured by the bolometer horizontal camera. The impurity accumulation event [34] can be observed in figures 17(*d*)–(*f*), by the collapse of the core electron temperature (figure 17(*e*)) supported by an increase of the core electron density (figure 17(*f*)) and a radiation peaking at the plasma central lines of sight of the bolometer horizontal camera (figure 17(*d*)). This pattern is well recognized by the CNN predictor. In fact, the disruptive likelihood rises at around 10.35 s and the alarm is triggered at 10.442 s, about 500 ms in advance with respect to JTT$_{trigger}$ time.

Instead, figure 18 reports an example (#94900) where the RTPP system triggers an alarm at 10.96 s, whereas the CNN does not trigger any alarm. The protection system alarm was

triggered because of the increase of the radiated fraction, computed as the total radiated power divided by the total auxiliary heating power. From figure 18(*a*), it is possible to see as, at the JTT$_{\text{trigger}}$ time, the CNN output is indicating a regular plasma condition. A weak change in the CNN likelihood functions is detectable shortly after 11 s, where a high radiation is measured by central lines of sight of the bolometer horizontal camera, as visible in figure 18(*d*). However, the temperature and the density profiles (figures 18(*e*) and (*f*) respectively) do not follow any disruptive pattern. Thus, no alarm is triggered by the CNN predictor. This case shows how the network takes advantage from combining the information from the HRTS and the bolometer horizontal camera profiles to distinguish between disruptive and not disruptive radiation patterns even without having additional information from the vertical bolometer camera. In fact, looking at the bolometer vertical camera data, the radiation blob is not located in the core, but on the outer poloidal region of the plasma (vertical lines of sight 1–5, see figure 18(*c*)), and the high radiation was not connected to a disruptive mechanism, as properly identified by the CNN.

6. Conclusions

In this paper, a deep-CNN model for disruption prediction has been trained and tested on a dataset representative of different experimental plasma conditions, and an in-depth analysis of the performance evolution over the different datasets has been done. The peculiarity of the proposed method consists in the processing of the spatiotemporal information coming from 1D plasma profiles through a deep-CNN, and the automatic detection of the pre-disruptive phase of disruptions for the training selection.

The synthesis of appropriated features allowed the predictive model to correctly identify the localized destabilizations of the temperature, density, and radiation 1D profiles, leading to about, 93% of SPs, 9% of FAs and 4% of MAs, for a test set consisting in 108 disruptive and 149 regularly terminated discharges. The achieved performance is proved to be better than the one obtained with GTM model trained with the 0D peaking factors, reaching on the same test set to about 84% of SPs, 26% FAs and 2% of MAs. Moreover, the use of the 1D plasma profiles as model input allows the straightforward connection between the disruption chain of events and the predictors decisions. The analysis of the model output in the considered discharges highlighted that the combination of the profiles from the HRTS and the bolometer horizontal camera with the ML_{norm} and l_{i} signals allowed the identification of two different disruptive mechanisms: the impurity accumulation and the edge cooling. On the other hand, these two mechanisms are characterized by a different localization of the radiated power on the poloidal plane. Hence, in a future work, either the information from the bolometer vertical camera will be integrated as an input, or the tomography reconstruction will be provided as input to the CNN. This could help to improve the performance by avoiding FAs in high power

experiments. Note that, the CNNs are well suited to process images such as the bolometry ones.

Despite the general very good results, the performance analysis highlights as the operating range of the training data significatively affects the model performances through the later experimental campaigns. This fact is also confirmed by the well-known ageing issue which affects data-driven models. The continuous retraining performed by adding experimental data far from the target experimental conditions, can be a solution for limiting the predictor ageing. In this view, the automating data labelling steps adopted in this work can help in developing a hand-free retraining procedure. Therefore, the proposed approach offers a new prospective for the synthesis of general features which can detect the general physical mechanisms causing the disruptions and the development of more complex models, trained with very large datasets.

Another crucial aspect concerns the extrapolation of the proposed predictor, as well as any other kind of data-based predictors, to ITER, which is still an open issue. In literature, the extrapolation power of the neural network approach for disruption prediction has been investigated in [43] by a cross-tokamak method. The disruptions prediction performance in JET, based on neural networks trained on ASDEX Upgrade, is shown. The study concludes that, the weakness of the neural network approach to disruption prediction in a newly constructed tokamak and the need for a large training set of disruptions might be overcome by using training data from a pre-existing tokamak (or tokamaks), but refining some parameters on a reduced dataset from the test one. In fact, the knowledge of the scaling factor for the output threshold and of operating parameter ranges from the test data, are needed to have a reasonable prediction performance.

Presently, the predictor proposed in this work is not directly useable for disruption prediction in ITER plasmas because the input parameters are not dimensionless. But, once proved the suitability of the predictive algorithm, in view of overcoming the need of a large training set of disruptions, which is the main criticism of ITER disruption predictor, a pre-trained CNN model on pre-existing tokamaks data can be proposed. The model pre-trained to discriminate between safe and disruptive configurations, only with JET data or involving several other existing tokamaks, can be subsequently fine-tuned for the disruption prediction over a limited number of ITER disruptions. The reliability of this approach will be investigated in future works considering JET and ASDEX Upgrade data, as in [43].

Acknowledgments

ORCID iDs

E. Aymerich https://orcid.org/0000-0003-3787-7685
G. Sias https://orcid.org/0000-0002-2289-301X
F. Pisano https://orcid.org/0000-0003-0162-0562
B. Cannas https://orcid.org/0000-0002-2766-0557
S. Carcangiu https://orcid.org/0000-0002-0270-6432
C. Sozzi https://orcid.org/0000-0001-8951-0071
C. Stuart https://orcid.org/0000-0002-6790-1706
P.J. Carvalho https://orcid.org/0000-0002-8480-0499
A. Fanni https://orcid.org/0000-0001-8604-5282

References

[1] Ikeda K. 2009 ITER on the road to fusion energy *Nucl. Fusion* **50** 014002

[2] Cannas B., Fanni A., Marongiu E. and Sonato P. 2003 Disruption forecasting at JET using neural networks *Nucl. Fusion* **44** 68–76

[3] Cannas B., Fanni A., Sonato P. and Zedda M.K. 2007 A prediction tool for real-time application in the disruption protection system at JET *Nucl. Fusion* **47** 1559–69

[4] Cannas B., Delogu R.S., Fanni A., Sonato P. and Zedda M.K. 2007 Support vector machines for disruption prediction and novelty detection at JET *Fusion Eng. Des.* **82** 1124–30

[5] Dormido-Canto S., Vega J., Ramírez J.M., Murari A., Moreno R., López J.M. and Pereira A. 2013 Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER *Nucl. Fusion* **53** 113001

[6] Murari A., Lungaroni M., Peluso E., Gaudio P., Vega J., Dormido-Canto S., Baruzzo M. and Gelfusa M. 2018 Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET *Nucl. Fusion* **58** 056002

[7] Pau A., Fanni A., Carcangiu S., Cannas B., Sias G., Murari A. and Rimini F. 2019 A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET *Nucl. Fusion* **59** 106017

[8] Aymerich E., Fanni A., Sias G., Carcangiu S., Cannas B., Murari A. and Pau A. (JET Contributors) 2021 A statistical approach for the automatic identification of the start of the chain of events leading to the disruptions at JET *Nucl. Fusion* **61** 036013

[9] Cannas B., Fanni A., Pautasso G., Sias G. and Sonato P. 2010 An adaptive real-time disruption predictor for ASDEX Upgrade *Nucl. Fusion* **50** 075004

[10] Cannas B., Fanni A., Pautasso G. and Sias G. 2011 Disruption prediction with adaptive neural networks for ASDEX Upgrade *Fusion Eng. Des.* **86** 1039–44

[11] Aledda R., Cannas B., Fanni A., Pau A. and Sias G. 2015 Improvements in disruption prediction at ASDEX Upgrade *Fusion Eng. Des.* **96–97** 698–702

[12] Montes K.J. *et al* 2019 Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST *Nucl. Fusion* **59** 096015

[13] Zheng W. *et al* 2018 Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak *Nucl. Fusion* **58** 056016

[14] Rea C., Granetz R.S., Montes K., Tinguely R.A., Eidietis N., Hanson J.M. and Sammuli B. 2018 Disruption prediction investigations using machine learning tools on DIII-D and Alcator C-Mod *Plasma Phys. Control. Fusion* **60** 084004

[15] Pau A. *et al* 2018 A first analysis of JET plasma profile-based indicators for disruption prediction and avoidance *IEEE Trans. Plasma Sci.* **46** 2691–8

[16] Kates-Harbeck J., Svyatkovskiy A. and Tang W. 2019 Predicting disruptive instabilities in controlled fusion plasmas through deep learning *Nature* **568** 526–31

[17] Rea C., Montes K.J., Pau A., Granetz R.S. and Sauter O. 2020 Progress toward interpretable machine learning-based disruption predictors across tokamaks *Fusion Sci. Technol.* **76** 912–24

[18] de Vries P.C. *et al* 2014 The influence of an ITER-like wall on disruptions at JET *Phys. Plasmas* **21** 056101

[19] Pucella G. *et al* 2021 Onset of tearing modes in plasma termination on JET: the role of temperature hollowing and edge cooling *Nucl. Fusion* **61** 046020

[20] Rattá G.A., Vega J., Murari A. and Gadariya D. (JET Contributors) 2021 PHAD: a phase-oriented disruption prediction strategy for avoidance, prevention, and mitigation in JET *Nucl. Fusion* **61** 116055

[21] Bengio Y. 2009 Learning deep architectures for AI *Found. Trends Mach. Learn.* **2** 1–127

[22] Matos F.A., Ferreira D.R. and Carvalho P.J. 2017 Deep learning for plasma tomography using the bolometer system at JET *Fusion Eng. Des.* **114** 18–25

[23] Ferreira D.R., Carvalho P.J. and Fernandes H. (JET Contributors) 2018 Full-pulse tomographic reconstruction with deep neural networks *Fusion Sci. Technol.* **74** 47–56

[24] Cannas B., Carcangiu S., Fanni A., Farley T., Militello F., Montisci A., Pisano F., Sias G. and Walkden N. 2019 Towards an automatic filament detector with a Faster R-CNN on MAST-U *Fusion Eng. Des.* **146** 374–7

[25] Matos F., Menkovski V., Felici F., Pau A. and Jenko F. 2020 Classification of tokamak plasma confinement states with convolutional recurrent neural networks *Nucl. Fusion* **60** 036022

[26] Churchill R.M., Tobias B. and Zhu Y. 2020 Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data *Phys. Plasmas* **27** 062510

[27] Ferreira D.R., Carvalho P.J. and Fernandes H. 2020 Deep learning for plasma tomography and disruption prediction from bolometer data *IEEE Trans. Plasma Sci.* **48** 36–45

[28] Ferreira D.R., Carvalho P.J., Carvalho I.S., Stuart C. and Lomas P.J. 2021 Monitoring the plasma radiation profile with real-time bolometer tomography at JET *Fusion Eng. Des.* **164** 112179

[29] Stuart C.I. *et al* 2021 PETRA: a generalised real-time event detection platform at JET for disruption prediction, avoidance and mitigation *Fusion Eng. Des.* **168** 112412

[30] Leyland M.J., Beurskens M.N.A., Flanagan J.C., Frassinetti L., Gibson K.J., Kempenaars M., Maslov M. and Scannell R. (JET Contributors) 2016 Edge profile analysis of Joint European Torus (JET) Thomson scattering data: quantifying the systematic error due to edge localised mode synchronisation *Rev. Sci. Instrum.* **87** 013507

[31] Sias G., Cannas B., Fanni A., Murari A. and Pau A. 2019 A locked mode indicator for disruption prediction on JET and ASDEX Upgrade *Fusion Eng. Des.* **138** 254–66

[32] Pouyanfar S., Sadiq S., Yan Y., Tian H., Tao Y., Reyes M.P., Shyu M.-L., Chen S.-C. and Iyengar S.S. 2018 A survey on deep learning: algorithms, techniques, and applications *ACM Comput. Surv.* **51** 92

[33] Alom M.Z. *et al* 2019 A state-of-the-art survey on deep learning theory and architectures *Electronics* **8** 292

[34] Brock A., Lim T., Ritchie J.M. and Weston N. 2017 Freeze-Out: accelerate training by progressively freezing layers (arXiv:1706.04983 [cs, stat])

[35] Baltrusaitis T., Ahuja C. and Morency L.-P. 2019 Multimodal machine learning: a survey and taxonomy *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 423–43

[36] Gadzicki K., Khamsehashari R. and Zetzsche C. 2020 Early vs late fusion in multimodal convolutional neural networks *IEEE 23rd Int. Conf. Information Fusion (FUSION)* (Virtual Event 6 - 9 July) pp 1–6 (https://ieeexplore.ieee.org/xpl/conhome/9183728/proceeding)

[37] Huang S.-C., Pareek A., Seyyedi S., Banerjee I. and Lungren M.P. 2020 Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines *npj Digit. Med.* **3** 136

[38] Snoek C.G.M., Worring M. and Smeulders A.W.M. 2005 Early versus late fusion in semantic video analysis *Proc. 13th Annual ACM Int. Conf. Multimedia (MULTIMEDIA '05)* (Hilton Singapore, 6-11 November) (New York: Association for Computing Machinery) pp 399–402 (https://dl.acm.org/doi/proceedings/10.1145/1101149)

[39] Garcia J. *et al* 2021 Integrated scenario development at JET for DT operation and ITER risk mitigation *28th IAEA Fusion Energy Conf. (FEC 2020)* (Virtual Event. 10–15 May) (https://conferences.iaea.org/event/214/)

[40] Bishop C.M., Svensén M. and Williams C.K.I. 1998 GTM: the generative topographic mapping *Neural Comput.* **10** 215–34

[41] Cannas B., de Vries P.C., Fanni A., Murari A., Pau A. and Sias G. (JET Contributors) 2015 Automatic disruption classification in JET with the ITER-like wall *Plasma Phys. Control. Fusion* **57** 125003

[42] Piron L. *et al* 2021 Progress in preparing real-time control schemes for deuterium–tritium operation in JET *Fusion Eng. Des.* **166** 112305

[43] Windsor C.G., Pautasso G., Tichmann C., Buttery R.J. and Hender T.C. (JET EFDA Contributors and the ASDEX Upgrade Team) 2005 A cross-tokamak neural network disruption predictor for the JET and ASDEX Upgrade tokamaks *Nucl. Fusion* **45** 337–50