

Disruption prediction with artificial intelligence techniques in tokamak plasmas

J. Vega¹, A. Murari², S. Dormido-Canto³, G. A. Rattá¹, M. Gelfusa⁴ and JET Contributors*

¹Laboratorio Nacional de Fusión, CIEMAT. Av. Complutense 40. 28040 Madrid. Spain.

²Consorzio RFX (CNR, ENEA, INFN, Universita' di Padova, Acciaierie Venete SpA), Corso Stati Uniti 4, 35127 Padova, Italy.

³Depto. De Informática y Automática. UNED. 28040 Madrid, Spain.

⁴Department of Industrial Engineering, University of Rome “Tor Vergata”, via del Politecnico 1, Roma, Italy.

* See the author list of ‘Overview of JET results for optimising ITER operation’ by J. Mailloux et al. to be published in Nuclear Fusion Special issue: Overview and Summary Papers from the 28th Fusion Energy Conference (Nice, France, 10-15 May 2021).

e-mail: jesus.vega@ciemat.es; andrea.murari@istp.cnr.it; sebas@dia.uned.es; giuseppe.ratta@ciemat.es; gelfusa@ing.uniroma2.it.

ORCID codes: 0000-0002-1622-3984; 0000-0002-3932-3865; 0000-0001-7652-5338; 0000-0002-5676-9631; 0000-0001-5158-7292.

In nuclear fusion reactors, plasmas are heated to very high temperatures of more than 100 million kelvin and in so-called tokamaks, they are confined by magnetic fields in the shape of a torus. Light nuclei, such as deuterium and tritium, undergo a fusion reaction, which releases energy and makes fusion a promising option for a sustainable and clean energy source. Tokamak plasmas, however, are prone to disruptions, which are a sudden collapse of the system terminating the fusion reactions. As disruptions lead to an abrupt loss of confinement, they can cause irreversible damage to present-day fusion devices and are expected to have a more devastating effect in future devices. Disruptions expected in the next-generation tokamak ITER, for example, could cause electromagnetic forces larger than the weight of an Airbus A380. Furthermore, the thermal loads in such an event could exceed the melting threshold of the most resistant state-of-the-art materials by more than an order of magnitude. To prevent disruptions or at least mitigate their detrimental effects, empirical models obtained with artificial intelligence methods, of which an overview is given here, are commonly employed to predict their occurrence — and ideally give enough time to introduce counteracting measures.

Tokamaks are currently the most promising configuration for a commercial fusion reactor but — contrary to stellarators — they are prone to disruptions. Because they are also very

complex devices, disruptions depend on many effects as well as on nonlinear interactions between them. Pulsed tokamak experiments consist of discharges of currents of the order of millions of amperes. The normal evolution of these discharges can be suddenly interrupted by various types of instabilities [1]. Particularly frequent and dangerous are instabilities related to excessive radiation (from the visible to the X-ray region of the spectrum), too high plasma density or anomalous current profiles.

Disruptions occur in two stages, namely the thermal quench and the current quench. During the thermal quench, most of the plasma's internal energy is lost on time scales of the order of one millisecond. This thermal quench is immediately followed by the current quench, during which the plasma current is extinguished in time intervals that can last from a few to hundreds of milliseconds in present-day tokamaks. The lead-up to a disruption is typically characterised by anomalies in several diagnostic signals, for example in the electron temperature (see Figure 1). These so-called precursor signals, however, can also be present in non-disruptive plasmas, making the prediction of disruptions a complex multi-objective problem. Because the mitigation of disruptions requires the immediate termination of the discharge, false alarms are costly in terms of resources as well as risking damage to the devices. For this reason, both false positives and false negatives need to be kept to a minimum.

The accurate prediction of disruptions will be even more important for next-generation tokamaks, which will operate with metallic plasma facing components. Metal offers several advantages. First, it can stand the loads with acceptable erosion, meaning that it has a smaller impact on the lifetime of the components facing the plasma and thus on the efficiency of the tokamak. Second, the retention of the plasma fuel is comparatively low. High retention, that is the accumulation of radioactive fuel within the wall, is a safety threat and would also affect the availability of the tokamak. However, recent experiments with metallic walls have shown that the challenge posed by disruptions is more severe for devices with metallic plasma facing components than expected. ITER requires less than 5% of pulses with disruptions at a maximum current of 15 MA [2]. Experiments on the Joint European Torus (JET) with the ITER-like wall made of tungsten and beryllium have demonstrated that the rate of disruptions can be unacceptably high [3]. The rate in the so-called baseline reference scenario for ITER reached 80% on JET [4], but was excessive in all the devices, on which the scenario has been tested under reactor relevant conditions. In the hybrid reference scenario that is developed on JET, the rate of disruptions was about 20%, which also does not meet the requirements for ITER.

In light of these recent findings, the abrupt termination of the discharges is a major issue for tokamaks and disruption prediction will be a crucial real-time requirement in even larger ones. Because theoretical models are often insufficient to reliably describe disruptions, empirical models based on machine learning are a common approach for understanding and predicting disruptions.

Models based on traditional machine learning

Disruption predictors based on machine learning are usually conceptualised as binary classifiers: the training process splits the operational space into two regions — the disruptive and the non-disruptive region — and determines the boundary in between. Ideally, classifications are meant to be carried out during discharges with a typical time resolution on the order of milliseconds. The signals are typically available in the form of time series, which are sequences of data points indexed in time (usually equally spaced). For disruption prediction, the signals have to be processed and then suitable predictors have to be developed. So far, the variety of existing real-time signal processing methods implemented have explored practically all known data analysis techniques for time series in the time domain [5-15]. These techniques have been complemented with tools in the frequency domain [16], based on Fourier transforms. Approaches relying on a mixture of time/frequency domains, including wavelet decompositions, have also been pursued [17-19]. With regard to classifier technologies, real-time compatible predictors have typically been based on artificial neural networks, support vector machines, fuzzy logic, generative topographic mapping and deep learning and have been studied on a broad range of tokamaks, including ADITYA (India) [20], ASDEX Upgrade (Germany) [21], DIII-D (US) [22-24], J-TEXT (China) [25], NSTX (US) [26], ALCATOR C-MOD (US) [27], JT-60U (Japan) [28], EAST (China) [29-31], HL-2A (China) [32] and JET (UK) [33-35]. Out of the three machine-learning based predictors that were implemented in JET's real-time network, APODIS [17], SPAD [36], and Centroid [37], the former correctly identified disruptions in more than 98% of cases, and had a false alarm rate, that is wrongly classified non-disruptive data, in less than 2% of cases with an average warning time of hundreds of milliseconds.

Despite encouraging results, this disruption predictor and others based on traditional machine-learning technologies suffer from inherent fundamental limitations. First, they are not derived from first principles but are empirical. This means that their results are difficult to interpret in terms of plasma dynamics and whether they can be extrapolated

to future, larger devices [38-39] remains unclear. Second, traditional machine-learning predictors require very large amounts of data for the training. Given the potential damage caused by disruptions, collecting many examples is not a viable option for large-scale devices such as ITER. Finally, the predictors lack generality. Even when large data sets are available, the performance of these predictors degrades quickly when the discharge present characteristics different from those in the training data. Moreover, the use of predictors is typically restricted to the specific tokamak for which they were derived and it has proven challenging to transfer machine-learning classifiers from one device to another.

The classifiers employed in the studies on tokamaks mentioned above [20–35] were developed using real-time valid solutions, which guarantee response times within a specified time window. The predictors discussed in the remainder of this work have been tested offline with real-time compatible technologies and using only real-time available signals. After having been trained, these predictors can provide their output on a millisecond time scale, which is sufficiently fast not to affect the overall reaction time of the actuators for disruption mitigation and avoidance in tokamaks.

From closed- to open-world learning

The main drawbacks of traditional machine-learning predictors can be attributed to the assumptions adopted for their training. Most examples discussed above followed a closed-world approach to learning. This means that the information required for training the classifier has to be available prior to the first prediction. Moreover, the performance of traditional classifiers hinges on the assumption that the data are sampled independently from an identical distribution function. This assumption, however, implies that the plasmas are stationary in the sense that their data distribution function does not change significantly because the predictors have no capability of adapting to new regimes or new physics. In practice, these assumptions are systematically violated due to the rapid evolution of experimental programmes.

This situation is particularly unsatisfactory because humans can learn from few examples, can adapt to changing situations and can also transfer knowledge from one problem to similar ones.

In recent years, more attention has been devoted to providing deep learning, and interpretable solutions for disruption prediction across tokamaks and in particular for next-generation of devices such as ITER [40]. Efforts towards implementing an open-

world approach to learning have also become more popular, as evidenced by a series of adaptive strategies that have been developed to maximise the performance of disruption predictors in non-stationary conditions. The training of adaptive predictors is done from scratch, meaning that only a single example of each class (discharge with or without a disruption) is needed for a prediction [18, 19, 21, 34]. The predictors are updated between discharges by refining the training sets and by implementing trajectory learning during the shots. The cases causing the predictors to err still contain a lot of useful information. Retraining them with these failed examples is an effective form of adaptive learning. Because fusion plasmas exhibit memory effects, taking into account the evolution of their properties — their trajectories as opposed to values at specific times — is beneficial for the improvement of the predictor performance. The most advanced versions of open-world strategies also include various forms of de-learning that allows predictors to discard or reduce the effects of training data that do not apply anymore.

For the JET tokamak, fully general and automatic adaptive predictors have been developed. These are based on ensemble classifiers, which consist of a high number of specific predictors — each trained on slightly different datasets [41]. The outputs of the individual classifiers are then evaluated with a suitable decision function to determine whether a disruption is likely to occur. The overall performance of these adaptive predictors is promising: false alarms are less than 1% and the rate of correctly identified disruptions is higher than 99%. This is shown in Figure 2, which reports results for thousands of discharges at the beginning of the operation of JET with the ITER-like wall to simulate the initial operation of a new device [41]. In this case, the first prediction of the ensemble classifier was based on a single disruptive and four non-disruptive discharges [34, 41]. Because adaptive predictors based on open-world learning have been successfully transferred from one device to another [42, 43], they provide much better flexibility and generality than more traditional machine-learning techniques. Therefore, open-world learning is a very promising approach for implementation in next-generation tokamaks.

Interpretable models with symbolic regression

The classifiers discussed above provide mathematical models, which have no relation to the actual plasma dynamics and are difficult to interpret in terms of our understanding of plasma physics. For this reason, methodologies for steering the machine-learning process toward interpretable models, reflecting the actual physics and dynamics of the phenomena

involved, are under development. This is more ambitious than ‘traditional’ explainable artificial intelligence [44], because the goal is to obtain mathematical equations describing the underlying physics [26, 45-46].

Physics interpretability of models obtained with support vector machines was achieved by applying symbolic regression methods [47], which make use of genetic algorithms [48]. The deployment of symbolic regression allows the exploration of a large set of mathematical equations — hundreds of thousands of models — describing the boundary between non-disruptive and disruptive regions of the operational space. Each generation of models is evaluated based on a fitness function and those demonstrating better performance in terms of this metric are retained and used as starting point for the next iteration, from which new models are derived by using traditional genetic operators such as mutation, copy and cross-over.

The equation of the boundary, between disruptive and non-disruptive regions in JET’s operational space with an ITER-like wall displayed in Figure 3, was obtained with symbolic regression and revealed information on factors that are likely to trigger disruptions [49].

Warning time prediction

Even predictors with a good performance in terms of success and false alarm rates suffer from a major limitation: on JET, the range of their warning times can be of the order of one second. Thus, the control system lacks the information about the time remaining before the beginning of the current quench— it could be in a few milliseconds or in a second. Given the importance of predicting the time remaining before the occurrence of a disruption [37] or at least providing a robust estimate of the minimum time still available to introduce remedial actions [50-51], accurate estimates of the warning time are indispensable. A recent study on JET combining Support Vector Machines and genetic programming [52] allowed integrating three classes of predictors for avoidance, prevention and mitigation of disruptions. As shown in Figure 4, the warning times of the overall system obtained with the three predictors present negligible overlap, providing a clear lower bound of the time intervals available to introduce remedial actions.

Outlook

For the prediction of disruptions on next-generation tokamaks like ITER, these results need to be transferred to these devices. Achieving the same performance after adequate

modifications of the algorithms would be a major breakthrough — even if the overall system requires large amounts of data for the training. In this regard, a coherent strategy is emerging for ITER, suggesting to deploy adaptive predictors in the first operational stages, providing inputs to the genetic programming for the training of the integrated system that would provide accurate warning times for avoidance, prevention and mitigation of disruptions.

Apart from the prediction of disruptions in tokamaks, most of the discussed data-driven techniques are fully general and could be adapted for forecasting and understanding any form of collapse, crash and catastrophe in other fields of science. Indeed, some of these techniques are already being deployed in various disciplines ranging from earth science to epidemiology [53-55].

References

- [1] Boozer, A. H. Theory of tokamak disruptions. *Physics of plasmas* **19**, 058101 (2012).
- [2] de Vries, P. C. et al. Requirements for Triggering the ITER Disruption Mitigation System. *Fusion Science and Technology* **69**, 471-484 (2016).
- [3] Wenninger, R., et al. Power handling and plasma protection aspects that affect the design of the DEMO divertor and first wall. *Proceedings of the of 26th IAEA Fusion Energy Conference, Kyoto, Japan. 2016. FIP /P7-14. (<https://nucleus.iaea.org/sites/fusionportal/Shared%20Documents/FEC%202016/fec2016-preprints/preprint0322.pdf>).*
- [4] Sozzi, C. et al. Termination of discharges in high performance scenarios in JET. *Proc. of the 28th IAEA Fusion Energy Conference. (<https://conferences.iaea.org/event/214/contributions/17328/>).*
- [5] Strait, E. J. et al. Progress in disruption prevention for ITER. *Nuclear Fusion* **59**, 112012 (2019).
- [6] Hollmann, E. M. et al. Status of research toward the ITER disruption mitigation system. *Physics of Plasmas* **22**, 021802 (2015).
- [7] Esposito, B. et al. Disruption avoidance in the Frascati Tokamak Upgrade by means of magnetohydrodynamic mode stabilization using electron-cyclotron-resonance heating. *Physical Review Letters* **100**, 045006 (2008).
- [8] Maraschek, M. et al. Path-oriented early reaction to approaching disruptions in ASDEX Upgrade and TCV in view of the future needs for ITER and DEMO. *Plasma Physics and Controlled Fusion* **60**, 014047 (2017).

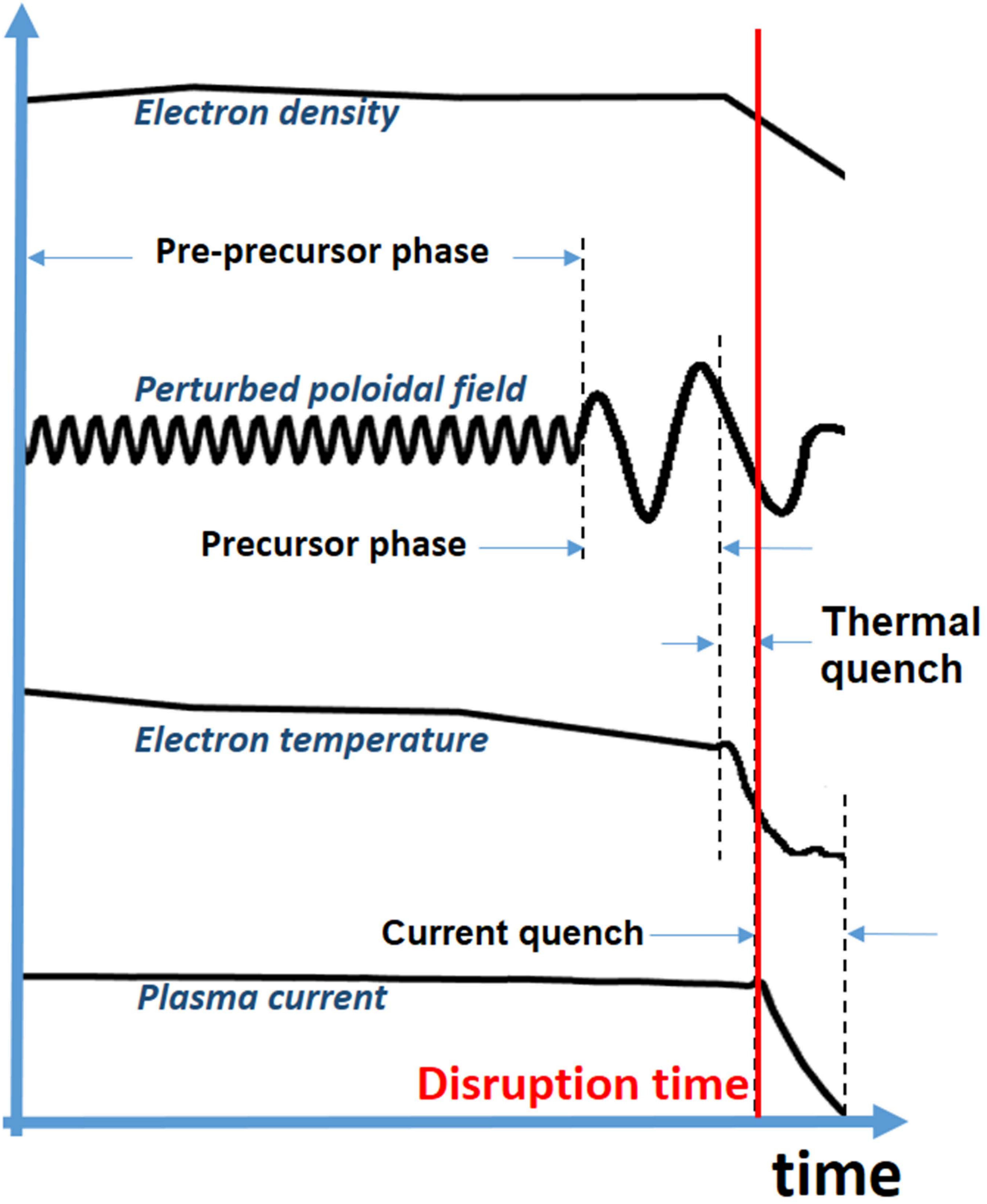
- [9] Baylor, L. R. et al. Disruption mitigation system developments and design for ITER. *Fusion Science and Technology* **68**, 211-215 (2015).
- [10] Pautasso, G. et al. On-line prediction and mitigation of disruptions in ASDEX Upgrade. *Nuclear Fusion* **42**, 100–109 (2002).
- [11] Cannas, B., Fanni, A., Pautasso, G., Sias, G., & Sonato, P. An adaptive real-time disruption predictor for ASDEX Upgrade. *Nuclear Fusion* **50**, 075004 (2010).
- [12] Kates-Harbeck, J., Svyatkovskiy, A., & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526-531 (2019).
- [13] Churchill, R. M., et al. Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data. *Physics of Plasmas* **27**, 062510 (2020).
- [14] Akçay, Cihan, et al. Machine learning methods for probabilistic locked-mode predictors in tokamak plasmas. *Physics of Plasmas* **28**, 082106 (2021).
- [15] Ferreira, D. R., Carvalho, P. J., & Fernandes, H. Deep learning for plasma tomography and disruption prediction from bolometer data. *IEEE Transactions on Plasma Science* **48**, 36-45 (2019).
- [16] Rattá, G. A. et al. An advanced disruption predictor for JET tested in a simulated real-time environment. *Nuclear Fusion* **50**, 025005 (2010).
- [17] Vega, J. et al. Results of the JET real-time disruption predictor in the ITER-like wall campaigns. *Fusion Engineering and Design* **88**, 1228-1231 (2013).
- [18] Dormido-Canto, S., et al. Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER. *Nuclear Fusion* **53**, 113001 (2013).
- [19] Vega, J. et al. Adaptive high learning rate probabilistic disruption predictors from scratch for the next generation of tokamaks. *Nuclear Fusion* **54**, 123001 (2014).
- [20] Agarwal, A., et al. Deep sequence to sequence learning-based prediction of major disruptions in ADITYA tokamak. *Plasma Physics and Controlled Fusion* **63**, 115004 (2021).
- [21] Zhang, Y., Pautasso, G., Kardaun, O., Tardini, G., & Zhang, X. D. Prediction of disruptions on ASDEX Upgrade using discriminant analysis. *Nuclear Fusion* **51**, 063039 (2011).
- [22] Rea, C. et al. Disruption prediction investigations using machine learning tools on DIII-D and Alcator C-Mod. *Plasma Physics and Controlled Fusion* **60**, 084004 (2018).
- [23] Rea, C., Montes, K. J., Erickson, K. G., Granetz, R. S., & Tinguely, R. A. A real-time machine learning-based disruption predictor in DIII-D. *Nuclear Fusion* **59**, 096016 (2019).

- [24] Rea, C., & Granetz, R. S.. Exploratory machine learning studies for disruption prediction using large databases on DIII-D. *Fusion Science and Technology* **74**, 89-100 (2018).
- [25] Zheng, W., et al. Disruption predictor based on neural network and anomaly detection on J-TEXT. *Plasma Physics and Controlled Fusion* **62**, 045012 (2020).
- [26] Gerhardt, S. P. et al. Detection of disruptions in the high- β spherical torus NSTX. *Nuclear Fusion* **53**, 063021 (2013).
- [27] Montes, K. J. et al. Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST. *Nuclear Fusion* **59**, 096015 (2019).
- [28] Yokoyama, T. et al. Likelihood Identification of High-Beta Disruption in JT-60U. *Plasma and Fusion Research* **16**, 1402073 (2021).
- [29] Guo, B. H., et al. Disruption prediction on EAST tokamak using a deep learning algorithm. *Plasma Physics and Controlled Fusion* **63**, 115007 (2021).
- [30] Guo, B. H. et al. Disruption prediction using a full convolutional neural network on EAST. *Plasma Physics and Controlled Fusion* **63**, 025008 (2020).
- [31] Hu, W. H. et al. Real-time prediction of high-density EAST disruptions using random forest. *Nuclear Fusion* **61** 066034 (2021).
- [32] Zhong, Y., et al. Disruption prediction and model analysis using LightGBM on J-TEXT and HL-2A. *Plasma Physics and Controlled Fusion* **63**, 075008 (2021).
- [33] Moreno, R. et al. Robustness and increased time resolution of JET Advanced Predictor of Disruptions. *Plasma Physics and Controlled Fusion* **56**, 114003 (2014).
- [34] Murari, A. et al. Adaptive predictors based on probabilistic SVM for real-time disruption mitigation on JET. *Nuclear Fusion* **58**, 056002 (2018).
- [35] López, J. M. et al. Implementation of the disruption predictor APODIS in JET's real-time network using the MARTe framework. *IEEE Transactions on Nuclear Science* **61**, 741-744 (2014).
- [36] Esquembri, S. et al. Real-Time Implementation in JET of the SPAD Disruption Predictor Using MARTe. *IEEE Transactions on Nuclear Science* **65**, 836-842 (2018).
- [37] Vega, J. et al. A linear equation based on signal increments to predict disruptive behaviours and the time to disruption on JET. *Nuclear Fusion* **60**, 026001 (2019).
- [38] Rea, C. et al. Progress Toward Interpretable Machine Learning–Based Disruption Predictors Across Tokamaks. *Fusion Science and Technology* **76**, 912-924 (2020).
- [39] Zhu, J. X. et al. Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks. *Nuclear Fusion* **61**, 026007 (2020).

- [40] Zhu, J. et al. Scenario adaptive disruption prediction study for next generation burning-plasma tokamaks. *Nuclear Fusion* **61**, 114005 (2021).
- [41] Murari, A., Lungaroni, M., Gelfusa, M., Peluso, E., Vega, J. & JET Contributors. Adaptive learning for disruption prediction in non-stationary conditions. *Nuclear Fusion* **59**, 086037 (2019).
- [42] Murari, A. et al. On the transfer of adaptive predictors between different devices for both mitigation and prevention of disruptions. *Nuclear Fusion* **60**, 056003 (2020).
- [43] Murari, A., Rossi, R., Lungaroni, M., Baruzzo, M., & Gelfusa, M. Stacking of predictors for the automatic classification of disruption types to optimize the control logic. *Nuclear Fusion* **61**, 036027 (2021).
- [44] Gunning, D. et al. XAI-Explainable artificial intelligence. *Science Robotics* **4**, eaay7120 (2019).
- [45] Yokoyama, T. et al. Prediction of high-beta disruptions in JT-60U based on sparse modeling using exhaustive search. *Fusion Engineering and Design* **140**, 67-80 (2019).
- [46] Piccione, A. et al. Physics-guided machine learning approaches to predict the ideal stability properties of fusion plasmas. *Nuclear Fusion* **60**, 046033 (2020).
- [47] Li, H. X., Yang, J. L., Zhang, G., & Fan, B. Probabilistic support vector machines for classification of noise affected data. *Information Sciences* **221**, 60-71 (2013).
- [48] Rattá, G. A., Vega, J., Murari, A. & JET-EFDA Contributors. Improved feature selection based on genetic algorithms for real-time disruption prediction on JET. *Fusion Engineering and Design* **87**, 1670-1678 (2012).
- [49] Murari, A. et al. Investigating the physics of tokamak global stability with interpretable machine learning tools. *Appl. Sci.* **10**, 6683 (2020).
- [50] Fu, Y. et al. Machine learning control for disruption and tearing mode avoidance. *Physics of Plasmas* **27**, 022501 (2020).
- [51] Boyer, M. D., Rea, C. & Clement, M. D. Toward active disruption avoidance via real-time estimation of the safe operating region and disruption proximity in tokamaks. *Nuclear Fusion* **62**, 026005 (2021).
- [52] Rattá, G. A., Vega, J., Murari, A., Gadariya, D. & JET Contributors. PHAD: a phase-oriented disruption prediction strategy for avoidance, prevention, and mitigation in JET. *Nuclear Fusion* **61**, 116055 (2021).
- [53] Rossi, R., Gelfusa, M., Malizia, A., Gaudio, P. Adaptive quasi-unsupervised detection of smoke plume by LIDAR. *Sensors* **20**, 6602 (2020).
- [54] Murari, A., Peluso, E., Gelfusa, M., Lungaroni, M. & Gaudio, P. How to handle error bars in symbolic regression for data mining in scientific applications. *Proceedings of the*

□□□□□.

The predictors were optimized for avoidance, prevention and mitigation of disruptions. Avoidance actions keep the plasma within stability boundaries, prevention methods terminate the discharge in a controlled way and mitigation techniques alleviate the consequences of unavoidable disruptions. More than one thousand discharges of JET with the ILW were analysed. The warning times present negligible overlap, providing a clear estimate of the minimum intervals remaining to introduce remedial actions — from a minimum of 400 milliseconds in the case of the avoidance predictor to tens of milliseconds in the case of the one for mitigation. The vertical purple line indicates the disruption time, i.e. the beginning of the current quench, and the blue solid line represents the plasma current.



Electron density

Pre-precursor phase

Perturbed poloidal field

Precursor phase

Thermal quench

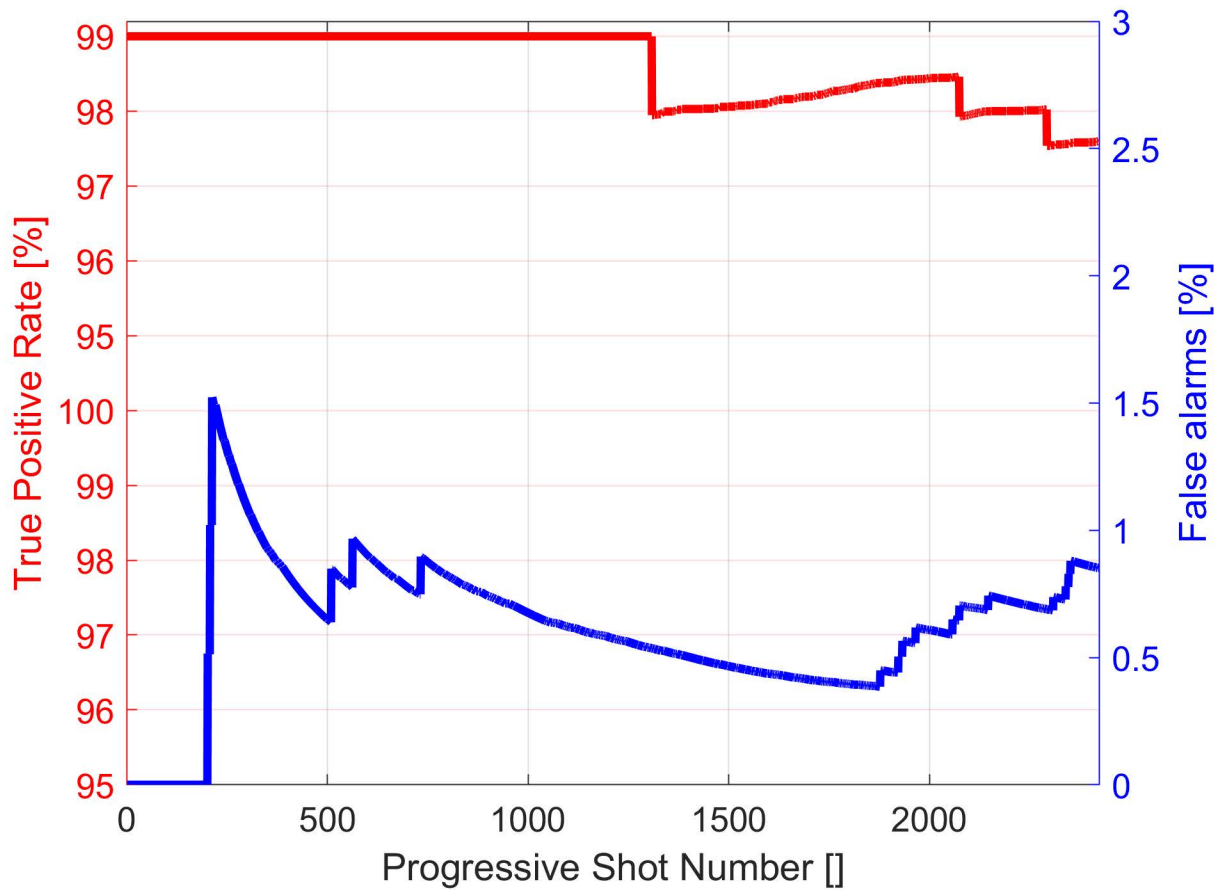
Electron temperature

Current quench

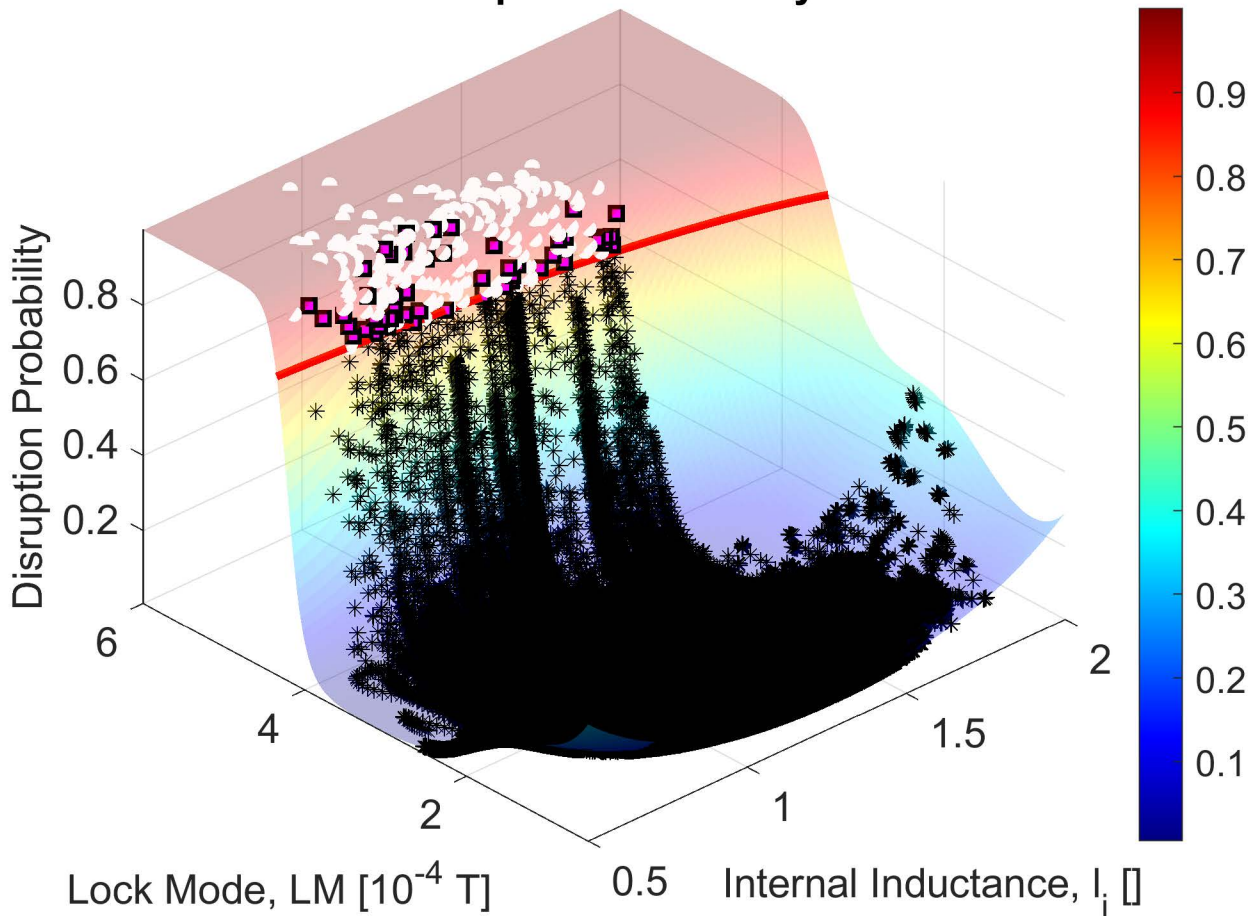
Plasma current

Disruption time

time



Disruption Probability



Abnormal profiles

(avoidance)

Radiative
instabilities

(prevention)

DISRUPTION

Plasma current

-1000

-800

-600

-400

-200

0

Warning time [ms]

Mitigation

