Likelihood Identification of High-Beta Disruption in JT-60U

Tatsuya YOKOYAMA[1,2)], Hiroshi YAMADA[1)], Akihiko ISAYAMA[3)], Ryoji HIWATARI[4)],
Shunsuke IDE[3)], Go MATSUNAGA[3)], Yuya MIYOSHI[4)], Naoyuki OYAMA[3)], Naoto IMAGAWA[1)],
Yasuhiko IGARASHI[5,6)], Masato OKADA[1)] and Yuichi OGAWA[1)]

[1)]*Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-8561, Japan*
[2)]*Research Fellow of Japan Society for the Promotion of Science, Tokyo 102-0083, Japan*
[3)]*Naka Fusion Institute, National Institutes for Quantum and Radiological Science and Technology,
Naka 311-0193, Japan*
[4)]*Rokkasho Fusion Institute, National Institutes for Quantum and Radiological Science and Technology,
Rokkasho 039-3212, Japan*
[5)]*Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan*
[6)]*Japan Science and Technology Agency, PRESTO, Kawaguchi 332-0012, Japan*

Prediction and likelihood identification of high-beta disruption in JT-60U has been discussed by means of feature extraction based on sparse modeling. In disruption prediction studies using machine learning, the selection of input parameters is an essential issue. A disruption predictor has been developed by using a linear support vector machine with input parameters selected through an exhaustive search, which is one idea of sparse modeling. The investigated dataset includes not only global plasma parameters but also local parameters such as ion temperature and plasma rotation. As a result of the exhaustive search, five physical parameters, i.e., normalized beta β_{N}, plasma elongation κ, ion temperature T_{i} and magnetic shear s at the $q = 2$ rational surface, have been extracted as key parameters of high-beta disruption. The boundary between the disruptive and the non-disruptive zones in multidimensional space has been defined as the power law expression with these key parameters. Consequently, the disruption likelihood has been quantified in terms of probability based on this boundary expression. Careful deliberation of the expression of the disruption likelihood, which is derived with machine learning, could lead to the elucidation of the underlying physics behind disruptions.

1. Introduction

Disruption is a serious threatening event in tokamak plasmas, resulting in large electromagnetic and thermal loads on structural components of a tokamak device. In order to avoid critical damage of a device and ensure reliable operation, prediction, avoidance, and mitigation of disruption are crucial issues, in particular, in ITER [1, 2] as well as a future tokamak fusion reactor. Disruption is caused by a variety of magnetohydrodynamics (MHD) instabilities [3, 4], where major driving forces are the gradients of plasma currents and pressure.

While the physical modeling of disruption is progressing, reliable prediction of disruption still remains an open issue. In parallel with the elaboration of physics modeling, data-driven approaches to the prediction of disruption is attracting much attention and a variety of methods have been applied to this issue [5–8]. The criteria called disruptivity was proposed to define the likelihood of a disruption in JET [9] and NSTX [10]. The disruptivity here is defined as the number of disruptions that occur in a spe-

cific parameter space divided by the duration time that the plasma is in that state. The relationships between disruptivity and operational boundaries has been discussed statistically. Recently, the onset of tearing instability, which is one of the major causes of disruption, has been characterized using hazard function based on the survival analysis method [11] in DIII-D [12]. It should be noted that the interpretable disruption prediction is thought to be important not only to understand disruption dynamics but also to secure extrapolatable capability [13]. Appropriate selection of the input parameters is an essential element in the development and interpretation of a disruption predictor independently of the method of machine learning such as neural network, support vector machine, or random forest. In the earlier study [14], selection of input parameters was assessed by exhaustive search (ES) on the subject of predicting high-beta disruption in JT-60U with support vector machine (SVM). ES is one method of sparse modeling, which exploits the inherent sparseness in all high-dimensional data to extract the maximum amount of information from the data [15], and has been applied to devel-

opment of a disruption predictor. This scheme of machine learning has improved the performance of the classifier for the prediction of disruption.

In the present study, the SVM predictor has been developed and the input plasma parameters have been selected by ES using a new dataset in which local parameters are added to the dataset in a previous study [14]. The SVM provides the expression of the boundary to classify the disruption zone and the non-disruption zone. The distance from the boundary can be a meaningful scale to discuss the likelihood of disruption which is expected to prevent disruptions.

This article is organized as follows: Section 2 describes the high-beta experiments and disruptions in JT-60U and the result of feature extraction by ES. Section 3 presents discussions of the disruption likelihood. Section 4 concludes the paper.

2. Beta Limit and High-Beta Disruption

2.1 High-beta disruption in JT-60U

High-beta disruption in JT-60U has been investigated in this study. It is known that high-beta disruption has less obvious precursors than other types of disruptions such as density-limit disruptions and vertical displacement event (VDE) disruptions [16]. The difficulty of detection of precursors is catenated with difficulties in predicting disruption and identifying beta limit.

In this study, 36 plasma discharges were identified as "disruptive" discharge with the criterion that plasma currents shut down with their decay times shorter than 40 ms. Here, only current quenches during the flattop phase with an I_p of 0.9 MA have been taken into account. The occurrence of current quench is defined as the time at which plasma current falls below 95% of the flattop current. The decay time is defined as the time divided by 0.6 from the occurrence of current quench to the time plasma current falls below 40% of flattop current [17]. The discharges taken into the dataset are 75% of available discharges in which plasma current shut down. On the other hand, sixty-one plasma discharges in which the plasma current is controlled to be stationary at 0.9 MA, are defined as "non-disruptive" discharges.

Figure 1 shows toroidal beta β_t plotted against the Troyon limit I_p/aB_t for discharges that are compiled in the dataset. According to Fig. 1, disruptive discharges are distributed in a relatively higher beta regime than non-disruptive ones. Here, the dashed lines in Fig. 1 correspond to the typical values of normalized beta β_N, which is defined as follows:

$$\beta_N = \frac{\beta_t[\%]}{I_p[MA]/a[m]B_t[T]}. \tag{1}$$

In the present experiment, suppression of the resistive wall mode (RWM) with a growth time corresponding to the skin time of the resistive wall was attempted by driving toroidal
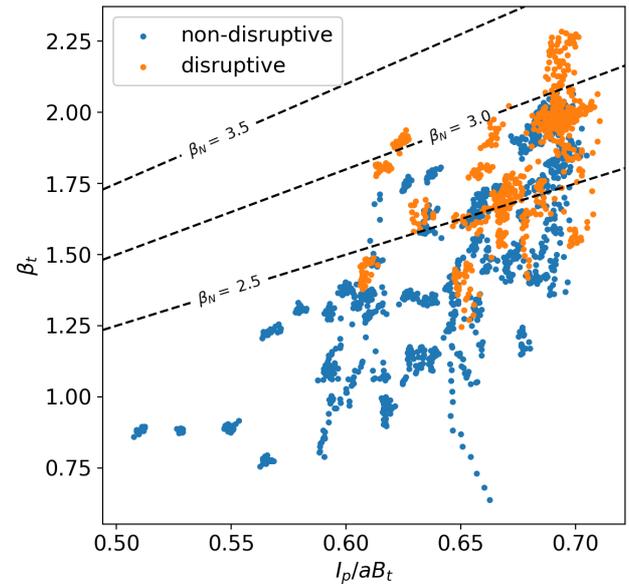
Fig. 1   Toroidal beta β_t against Troyon parameter I_p/aB_t in the analyzed dataset. Orange and blue dots show data points from disruptive and non-disruptive discharges, respectively. The dashed lines correspond with normalized beta β_N.

rotation of plasma with neutral beam injection (NBI). Consequently, the beta value was close or above the no-wall beta limit which is given as three times the internal inductance l_i by stability analysis using MARG2D [18–20].

2.2 Feature extraction of high-beta disruption

The combination of plasma parameters that describes the feature of high-beta disruption has been extracted by the K-sparse exhaustive search (ES-K) connected with linear SVM (ES-K-SVM for short). ES-K is one of the sparse modeling techniques which exploits the inherent sparseness in all high-dimensional data to extract the maximum amount of information from the data [15]. In ES-K, all possible combinations of input parameters are compared to each other to find out optimal one, assuming the optimal combination of explanatory variables is K-sparse. An SVM is a supervised machine learning technique [21] and has been used as a basic two-class classifier, which classifies the discharges into non-disruptive or disruptive in the present study. The boundary between the two classes is given by linear SVM as $f(\boldsymbol{x}) = 0$, where the decision function $f(\boldsymbol{x})$ is as follows:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b. \tag{2}$$

The details of the ES-K-SVM are described in a previous study [14].

In the present study, 14 plasma parameters are used in the dataset (see Table 1). It is shown that the RWM is suppressed by plasma rotation in DIII-D [22] and by rotation at the $q = 2$ surface in JT-60U [23]. Therefore, four

Table 1   Plasma parameters used in the machine learning model.

| Name of parameters | Expression | Range of training data | |
		Min.	Max.
Plasma current [MA]	I_p	0.70	1.01
Normalized beta	β_N	0.58	3.29
Plasma internal inductance	l_i	0.80	1.26
Safety factor at 95% of poloidal flux	q_{95}	3.07	4.33
Plasma triangularity	δ	0.28	0.46
Plasma elongation	κ	1.30	1.49
Magnetic perturbation amplitude ($n = 1$) [mT]	$|B_r^{n=1}|$	0.00	1.44
Ratio of electron density to the Greenwald density limit	$f_{GW} = \bar{n}_e / n_{GW}$	0.31	1.66
Ratio of radiated power to total input power	$f_{rad} = P_{rad} / P_{input}$	0	0.47
Speed of toroidal plasma rotation[$\times 10^5$ m/s] *	$|V_t|$	0.00	0.60
Direction of toroidal plasma rotation* **	$\exp(V_t/|V_t|)$	$1/e$	e
Ion temperature[keV] *	T_i	0.59	3.21
Radial location of $q = 2$ rational surface normalized by the minor radius of plasma surface a	$\rho_{q=2}/a$	0.46	0.80
Magnetic shear*	s	0.56	2.10

* For those parameters, values on rational surface where safety factor equal two, which is obtained by equilibrium calculation assuming $q = 1$ at plasma center, are used.

** For taking logarithms of this parameter in training process, exponential of the value has been taken.

radial profile parameters at around the $q = 2$ rational surface consist of $|V_t|$, T_i, $\rho_{q=2}/a$, and s, where q is the safety factor, and are added to the parameters used in the previous study [14]. For these parameters, the volume-averaged minor radius ρ is used as a radial index and the volume-averaged minor radius ρ is calculated as follows,

$$\rho = \sqrt{\frac{V(\Psi)}{2\pi^2 R}}. \qquad (3)$$

Here, $V(\Psi)$ is plasma volume surrounded by magnetic surface Ψ and R is the major radius. Ion temperature T_i at the $q = 2$ rational surface is around 2 keV and is distributed from 0.59 keV to 3.21 keV. In order to take logarithms of the velocity of toroidal rotation V_t, V_t has been separated into two parameters, that is, the speed of rotation $|V_t|$ and the direction of rotation $V_t/|V_t|$. Although there are some data with a high (over 1) f_{GW} value or high (over 1 mT) $|B_r^{n=1}|$ value, those data will be ignored in the training of machine learning models and these discharges will be judged as extraordinary discharges in the evaluation.

For each discharge, the data are taken every 5 ms within 200 ms before the reference time. The reference time is defined as the occurrence of current quench for disruptive discharges, while it is selected randomly from the range during which the plasma current is kept steady for non-disruptive discharges. The data values have been taken in logarithms and min-max normalization has been applied before the dataset was used to train and test the machine learning model. In the min-max normalization, the range of data was scaled into [0,1] for each parameter. By this preprocessing, the decision boundary between dis-

ruptive and non-disruptive classes obtained by the linear SVM is turned into an exponential form like a power law, which is frequently seen in nature. A model in the power law would be more relevant to physical approach than the simple linear combination.

2.3   Result of feature extraction

In order to assess the predictor model, two different evaluation values called prediction success rate (PSR) and false alarm rate (FAR) have been calculated for each combination. PSR is the ratio of the number of shots correctly judged as disruptive by the specific time of interest to the total number of disruptive shots. On the other hand, FAR is the ratio of the number of shots incorrectly judged as disruptive by the specific time of interest to the total number of non-disruptive shots. When the predictor shows the ideal performance, $PSR = 100\%$ and $FAR = 0\%$. Therefore, the distance from the ideal performance is defined as follows.

$$distance \equiv \sqrt{(100 - PSR)^2 + FAR^2}. \qquad (4)$$

In ES-K, the distance decreases from the beginning of the increase in K. Then, the reduction of the distance is saturated around $K = 5$. Figure 2 shows the result of ES-6-SVM at 30 ms before the reference time. Figure 2 (a) is referred to as a 2D density of state (2D-DoS) diagram, which is a two-dimensional (2D) histogram with PSR as the vertical axis and FAR as the horizontal axis. The color of each square corresponds to the number of combinations within the square. Figure 2 (b) shows the parameters included in
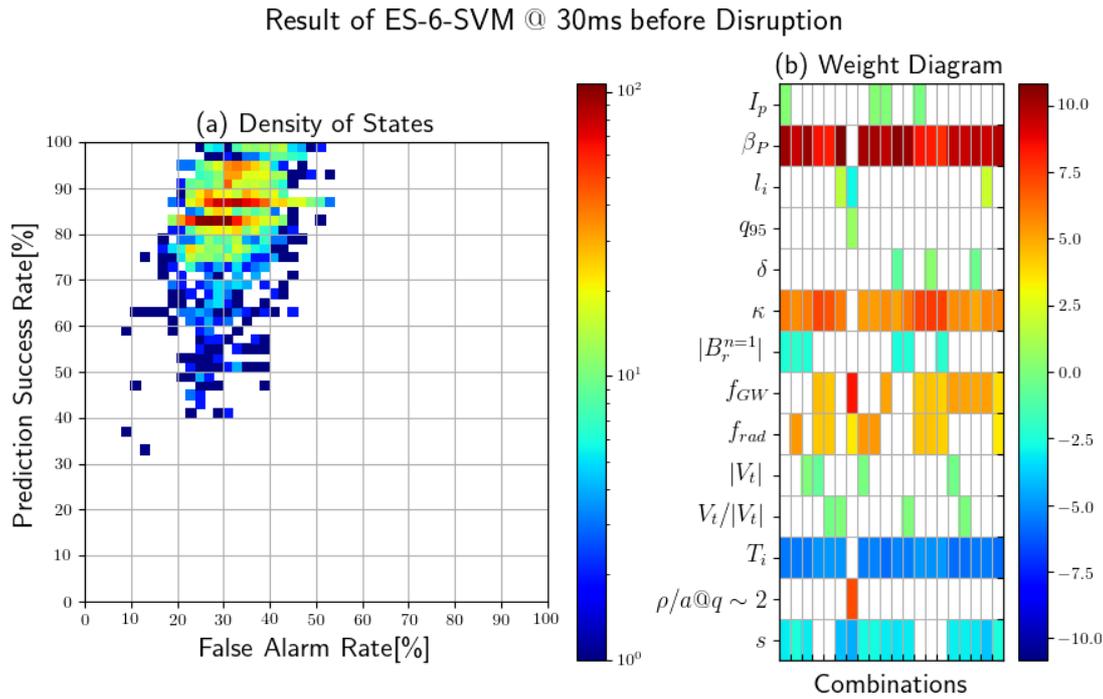
Fig. 2   (a) 2D-DoS diagram and (b) corresponding weight diagram of ES-6-SVM at 30 ms before disruption occurs.

the top 10 combinations in Fig. 2 (a) and is referred to as a weight diagram. The color of each square in the weight diagram represents the averaged weight $\boldsymbol{w}$ of the obtained decision function (shown in Eq. (2)), whose absolute value is basically paraphrased as the importance of each parameter in each combination.

According to Fig. 2, a combination of four parameters, that is, β_N, κ, T_i, and s, is dominant in the in the top 10 combinations. This means these parameters are the key parameters to predict disruption and it is implied that these parameters could characterize the condition of disruption. Distributions of these four parameters in disruptive and non-disruptive discharges are shown in Fig. 3. While the distribution β_N could be a relevant parameter to separate non-disruptive and disruptive discharges, other distributions are not well separated into two classes. Even for β_N, overlap of two classes implies that involvement of hidden parameters is the key for reliable prediction of disruption.

In the previously performed study [14], $|B_r^{n=1}|$ and its time derivative seemed to be the most important parameter in disruption prediction, because it was the parameter most frequently included in the superior combinations and the performance of the predictor seemed to depend on whether it is included or not in the combination. However, $|B_r^{n=1}|$ did not appear as the commonly included parameter in the present result. In the technical aspect, this different result was caused by the reconstruction of the dataset. More non-disruptive discharges with relatively large mode lock amplitude ($|B_r^{n=1}| > 0.5\,\text{mT}$) are included in the dataset used in the present study compared to the previous study.

Using those four parameters, the equation of decision
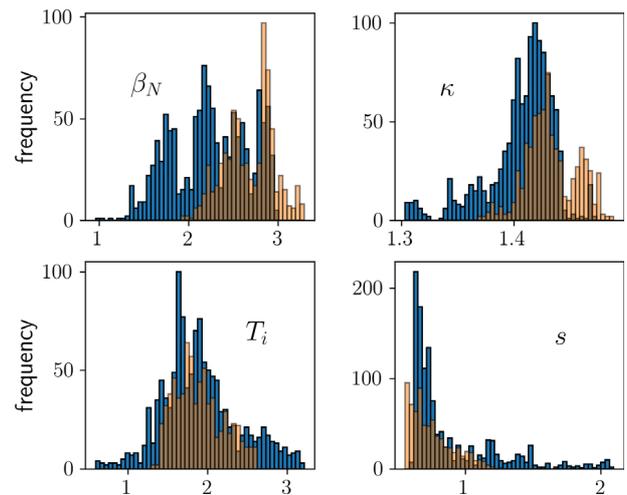
Fig. 3   Histograms of the representative parameters, i.e., β_N, κ, T_i, and s. The plots show the data distributions in the training time range in the non-disruptive (blue bars) and disruptive (orange bars) discharges.

boundary between disruptive and non-disruptive classes is obtained as $f_{\exp}(\boldsymbol{x}) = 1$, where $f_{\exp}$ is as follows:

$$f_{\exp} = \exp(-19.1)\,\beta_N^{6.32}\,\kappa^{39.6}\,T_i^{-2.48}\,s^{-2.43}. \qquad (5)$$

Note that the boundary equation shown in Section 2.2, $f(\boldsymbol{x}) = 0$, has been deformed into $f_{\exp}(\boldsymbol{x}) = 1$ by taking logarithms of the input data as preprocessing.

The decision function Eq. (5) indicates that the exponent of κ is much larger than those of other extracted parameters. This is mainly because the range of κ in the

dataset is much smaller than other extracted parameters. By min-max normalization, the range of κ is expanded to [0,1]. Therefore, the exponent of κ is amplified compared with the case where the range of κ is reduced into the original range.

Since the RWM is supposed to limit β_N in the present experiment, the plasma rotation to suppress the RWM would be the key parameters. However, neither the direction nor speed of plasma rotation has been extracted by the ES-K-SVM. In the present study, those local values at the $q = 2$ rational surface have been used since the mode structure of precursor oscillation is localized at the $q = 2$ rational surface. This mode is the energetic particle driven wall mode (EWM), which is thought to be one cause of RWM and lead to disruption in these experiments [19]. It should be noted here that the position of the $q = 2$ rational surface is provisionally evaluated by the equilibrium analysis assuming $q(\rho = 0) = 1$ while direct measurement of the poloidal magnetic field by the motional Stark effect shows that $q(\rho = 0)$ is larger than unity. Therefore, the real location of the $q = 2$ surface would be located inside this provisional estimate. However, it is pointed out that the profiles of V_t and T_i are peaked and s increases monotonically with the minor radius and local values used in the present analysis change gradually in space. Since the ambiguity of the location of $q = 2$ surface and accompanied errors of local values are systematically in the same direction, even the present analysis would not lead to qualitative misunderstanding. Precise quantitative assessment awaits future experiment. Precise identification of the location of the $q = 2$ surface leads to the detailed discussion about the physical linkage between the local parameters such as $|V_t|$, T_i, and s.

3. Disruption Likelihood

Using four extracted parameters, the characteristics of high-beta disruption in JT-60U has been discussed in terms of disruption likelihood. Note that the likelihood discussed in the present study is different from the disruptivity, which was discussed in previous studies about the operational limit against disruptions [9, 10]. In Fig. 4, the distribution of values of the decision function (Eq. (5)) in the dataset is shown in the bottom panel. The blue and red bars represent non-disruptive and disruptive cases, respectively, and both histograms are normalized to make the sum of bars equal one. The disruptive data are distributed mainly in $f(x) > 1$ region, while the non-disruptive data are distributed mainly in $f(x) < 1$ region. Here, the value of $f(x)$ corresponds to the distance from the decision boundary between the disruptive and non-disruptive regions. In the upper panel of Fig. 4, the percentage of disruptive data in each region of bars is shown with its fitted curve by a sigmoid function. The approximate function is expressed as follows,

$$y_{\text{fit}} = \frac{1}{1 + \exp\left\{-4.02\left(\log_{10} f_{\exp}(\boldsymbol{x}) + 0.1319\right)\right\}}. \quad (6)$$

The value of y_{fit} corresponds to the likelihood of disruption occurrence when the condition of plasma is expressed by $\boldsymbol{x}$. This likelihood is the expansion of the binary classifier to the predictor model with continuous value expression. The identified expression of likelihood quantifies the proximity to disruption, in other words, the risk of disruption. The likelihood expressed by measurable parameters is prerequisite for development of the control system to avoid disruption by means of multiple actuators [24].

In Fig. 5, the likelihood is expressed as a color contour on the plane of β_N and the term of other extracted parame-
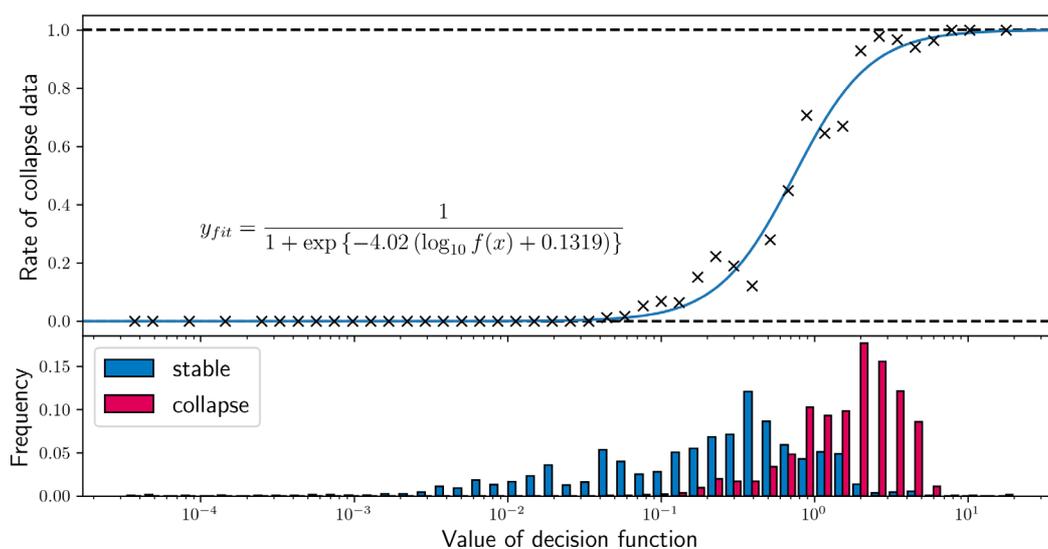
Fig. 4 (lower) The distribution of values of the decision function shown in Eq. (5) when each data in the dataset is input and (upper) the percentage of disruptive data in each region of bars. The curve in the upper figure is a fitted curve of the percentage of disruptive data. The function being expressed is also shown in the figure.
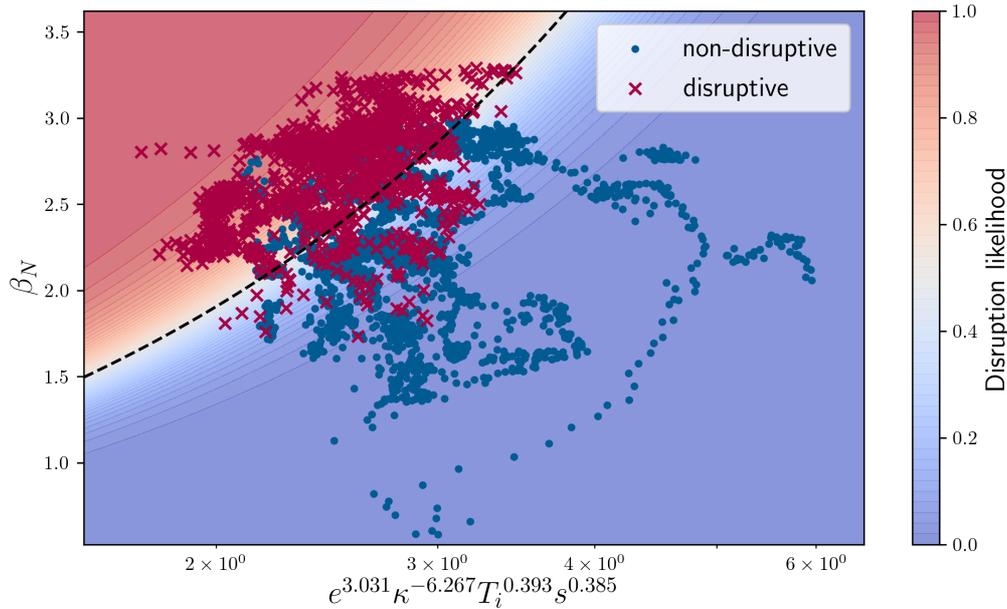
Fig. 5 The contour plot of the likelihood of disruption against β_N and the terms of other extracted parameters. The red crosses and blue dots show data points from disruptive and non-disruptive discharges, respectively. Note the e in the label of vertical axis is Napier's constant.
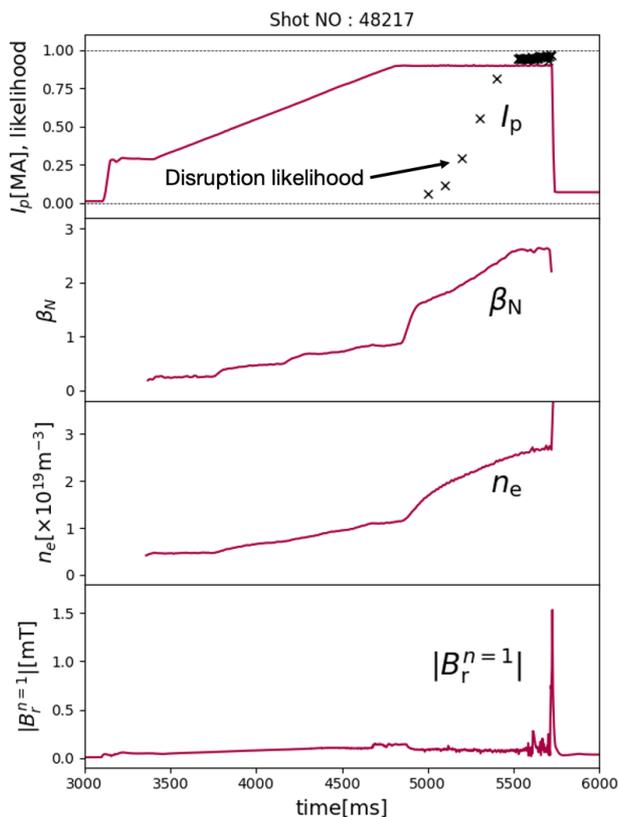
Fig. 6 Typical discharge with disruption in the JT-60U plasma experiment targeted in this research. The black crosses in the top panel shows the disruption likelihood, which is estimated within 200 ms before the current quench.

ters, that is, κ, T_i, and s along with Eq. (5). The likelihood shows that the higher ion temperature and magnetic shear can extend the β_N region with low disruption likelihood. Although this seems to go against the knowledge that the high elongation raises β_N, it should be pointed out that this trend means that lower elongation could bring a safer high β_N discharge but not that lower elongation leads to higher β_N.

Figure 6 shows the typical discharge with disruption in JT-60U and the disruption likelihood approaches unity before the occurrence of disruption. The data points correspond to the points calculated for the dataset.

4. Conclusion

In the present study, a disruption predictor model has been developed based on the high-beta experiment in JT-60U [19]. As key parameters of disruption prediction, four parameters that β_N, κ, T_i, and s were extracted using ES and SVM. The disruption likelihood has been estimated as a function of these four parameters in a power-law form.

In conclusion, the likelihood of disruption in high-β discharges on JT-60U is quantified by the distance from the decision boundary which is defined by the machine learning technique of sparse modeling (ES) and SVM. This boundary is expressed in the power law scaling with the limited number of parameters that are extracted according to their importance.

The disruption likelihood could be a way to apply the extracted features to predict and avoid disruption of tokamak plasma and also to give a hint to explore the underlying physics of disruption. Therefore, further validation of

the likelihood and extracted features using a dataset with a broader range of parameters and comparative study of different machines will be part of future works.

Acknowledgments

[1] M. Lehnen *et al.*, J. Nucl. Mater. **463**, 39 (2015).
[2] E. Strait *et al.*, Nucl. Fusion **59**(11), 112012 (2019).
[3] A.H. Boozer, Phys. Plasmas **19**(5), 058101 (2012).
[4] T. Hender *et al.*, Nucl. Fusion **47**(6), S128 (2007).
[5] A. Murari *et al.*, Nucl. Fusion **59**(8), 086037 (2019).
[6] A. Pau *et al.*, IEEE Trans. Plasma Sci. **46**(7), 2691 (2018).
[7] J. Kates-Harbeck, A. Svyatkovskiy and W. Tang, Nature **568**(7753), 526 (2019).
[8] K.J. Montes *et al.*, Nucl. Fusion **59**(9), 096015 (2019).
[9] P. de Vries *et al.*, Nucl. Fusion **49**(5), 055011 (2009).
[10] S. Gerhardt *et al.*, Nucl. Fusion **53**(4), 043020 (2013).
[11] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, volume 362 (John Wiley & Sons, 2011).
[12] K. Olofsson, B. Sammuli and D. Humphreys, Fusion Eng. Des. **146**, 1476 (2019).
[13] C. Rea *et al.*, Nucl. Fusion **59**(9), 096016 (2019).
[14] T. Yokoyama *et al.*, Fusion Eng. Des. **140**, 67 (2019).
[15] Y. Igarashi *et al.*, J. Phys.: Conf. Series **1036**, 012001 (2018).
[16] S. Ishida *et al.*, Plasma Physics and Controlled Nuclear Fusion Research: Proceedings of 14th IAEA Fusion Energy Conference **1**, 219 (1993).
[17] R. Sweeney *et al.*, Nucl. Fusion **57**(1), 016019 (2016).
[18] W. Howl *et al.*, Phys. Fluids B: Plasma Physics **4**(7), 1724 (1992).
[19] G. Matsunaga *et al.*, Nucl. Fusion **50**(8), 084003 (2010).
[20] S. Tokuda and T. Watanabe, Phys. Plasmas **6**(8), 3012 (1999).
[21] C. Cortes and V. Vapnik, Machine Learning **20**(3), 273 (1995).
[22] T.S. Taylor *et al.*, Phys. Plasmas **2**(6), 2390 (1995).
[23] M. Takechi *et al.*, Phys. Rev. Lett. **98**(5), 055002 (2007).
[24] Y. Miyoshi and Y. Ogawa, Plasma Fusion Res. **9**, 1405015 (2014).