

Prediction of multifaceted asymmetric radiation from the edge movement in density-limit disruptive plasmas on Experimental Advanced Superconducting Tokamak using random forest

Wenhui Hu(胡文慧), Jilei Hou(侯吉磊), Zhengping Luo(罗正平), Yao Huang(黄耀), Dalong Chen(陈大龙),Bingjia Xiao(肖炳甲), Qiping Yuan(袁旗平), Yanmin Duan(段艳敏), Jiansheng Hu(胡建生),Guizhong Zuo(左桂忠), and Jiangang Li(李建刚)

Citation: Chin. Phys. B, 2023, 32 (**7**): 075211. DOI: 10.1088/1674-1056/acd2b0
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Study on divertor plasma behavior through sweeping strike point in new lower divertor on EAST

Yu-Qiang Tao(陶余强), Guo-Sheng Xu(徐国盛), Ling-Yi Meng(孟令义), Rui-Rong Liang(梁瑞荣), Lin Yu(余林), Xiang Liu(刘祥), Ning Yan(颜宁), Qing-Quan Yang(杨清泉), Xin Lin(林新), and Liang Wang(王亮)
Chin. Phys. B, 2022, 31 (**6**): 065204. DOI: 10.1088/1674-1056/ac4a65

Reduction of impurity confinement time by combined heating of LHW and ECRH in EAST

Zong Xu(许棕), Zhen-Wei Wu(吴振伟), Ling Zhang(张凌), Yue-Heng Huang(黄跃恒), Wei Gao(高伟), Yun-Xin Cheng(程云鑫), Xiao-Dong Lin(林晓东), Xiang Gao(高翔), Ying-Jie Chen(陈颖杰), Lei Li(黎嫘), Yin-Xian Jie(揭银先), Qing Zang(臧庆), Hai-Qing Liu(刘海庆), and EAST team
Chin. Phys. B, 2021, 30 (**7**): 075205. DOI: 10.1088/1674-1056/abe1a6

Gaussian process tomography based on Bayesian data analysis for soft x-ray and AXUV diagnostics on EAST

Yan Chao(晁燕), Liqing Xu(徐立清), Liqun Hu(胡立群), Yanmin Duan(段艳敏), Tianbo Wang(王天博), Yi Yuan(原毅), Yongkuan Zhang(张永宽)
Chin. Phys. B, 2020, 29 (**9**): 095201. DOI: 10.1088/1674-1056/aba2e4

Estimation of tungsten production from the upper divertor in EAST during edge localized modes

Jing Ou(欧靖), Nong Xiang(项农), Zong-Zheng Men(门宗政), Ling Zhang(张凌), Ji-Chan Xu(许吉禅), Wei Gao(高伟)
Chin. Phys. B, 2019, 28 (**12**): 125201. DOI: 10.1088/1674-1056/ab5279

Synchrotron radiation intensity and energy of runaway electrons in EAST tokamak

Y K Zhang(张永宽), R J Zhou(周瑞杰), L Q Hu(胡立群), M W Chen(陈美文), Y Chao(晁燕), EAST team
Chin. Phys. B, 2018, 27 (**5**): 055206. DOI: 10.1088/1674-1056/27/5/055206

SPECIAL TOPIC — Plasma disruption

Prediction of multifaceted asymmetric radiation from the edge movement in density-limit disruptive plasmas on Experimental Advanced Superconducting Tokamak using random forest

Wenhui Hu(胡文慧)[1], Jilei Hou(侯吉磊)[1,†], Zhengping Luo(罗正平)[1], Yao Huang(黄耀)[1], Dalong Chen(陈大龙)[1], Bingjia Xiao(肖炳甲)[1,2], Qiping Yuan(袁旗平)[1], Yanmin Duan(段艳敏)[1], Jiansheng Hu(胡建生)[1,3], Guizhong Zuo(左桂忠)[1], and Jiangang Li(李建刚)[1]

[1] *Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei* 230031, *China*
[2] *University of Science & Technology of China, Hefei* 230026, *China*
[3] *Key Laboratory of Photovoltaic and Energy Conservation Materials, Chinese Academy of Sciences, Hefei* 230031, *China*

(Received 30 December 2022; revised manuscript received 4 April 2023; accepted manuscript online 5 May 2023)

Multifaceted asymmetric radiation from the edge (MARFE) movement which can cause density limit disruption is often encountered during high density operation on many tokamaks. Therefore, identifying and predicting MARFE movement is meaningful to mitigate or avoid density limit disruption for the steady-state high-density plasma operation. A machine learning method named random forest (RF) has been used to predict the MARFE movement based on the density ramp-up experiment in the 2022's first campaign of Experimental Advanced Superconducting Tokamak (EAST). The RF model shows that besides Greenwald fraction which is the ratio of plasma density and Greenwald density limit, $\mathrm{d}\beta_{\mathrm{p}}/\mathrm{d}t$, H_{98} and $\mathrm{d}W_{\mathrm{mhd}}/\mathrm{d}t$ are relatively important parameters for MARFE-movement prediction. Applying the RF model on test discharges, the test results show that the successful alarm rate for MARFE movement causing density limit disruption reaches $\sim 85\%$ with a minimum alarm time of ~ 40 ms and mean alarm time of ~ 700 ms. At the same time, the false alarm rate for non-disruptive and non-density-limit disruptive discharges can be kept below 5%. These results provide a reference to the prediction of MARFE movement in high density plasmas, which can help the avoidance or mitigation of density limit disruption in future fusion reactors.

Keywords: multifaceted asymmetric radiation from the edge (MARFE) movement prediction, random forest, machine learning, Experimental Advanced Superconducting Tokamak (EAST)

PACS: 52.55.Fa, 28.52.–s, 84.35.+i **DOI:** 10.1088/1674-1056/acd2b0

1. Introduction

High density is essential for the goal of practical fusion power since the fusion reaction rate is proportional to the square of plasma density in tokamak.[1] However, the plasma density has an upper limit and a degradation of plasma confinement or disruption during high density operation.[2] During high density operation, a multifaceted asymmetric radiation from the edge (MARFE) is often encountered and causes plasma disruption.[3–6] It is a toroidally symmetric, poloidally localized, strongly radiating region of high density and low temperature, typically seen on the high-field side or near the X-point of a tokamak.[7] And the formation of the MARFE is accompanied by rapid growth in the global radiation fraction. After the appearance of MARFE, current shrinkage is observed and then disruption is generally triggered.[8–10]

For the sake of improving plasma confinement and avoiding density limit disruption, a scheme of suppressing MARFE by modulating parallel electron thermal flux to increase edge transport barrier is proposed in Ref. [11] based on the model that radiative MARFE instability is caused by edge fluctuations. A few years later, experiments on TEXTOR-94 show that with suppression of unstable MARFE by reduction of recycling, the empirical density limit is overcome and a stationary plasma with density of 1.6× Greenwald limit is obtained.[6,12] While on Tore Supra, high plasma density of 1.4× Greenwald limit is achieved with stabilized MARFE in a high radiation and high recycling regime.[13] Experimental observations on HT-7 also show that stabilized MARFE is good for confinement at high density operation.[14,15] It will be quite beneficial to suppress MARFE movement to maintain high density near or above Greenwald density and avoid disruption for not only EAST but also future reactors, such as ITER and DEMO. In order to avoid or suppress MARFE movement, the first step is to predict and identify MARFE before it moves.[16] On TEXTOR-94, CII and CIII radiation signals near the inner wall are used for MARFE detection,[6] while on Frascati Tokamak Upgrade (FTU) real-time detection of MARFE is realized by processing images of the charge coupled device (CCD) camera.[17,18] Similar to FTU, fast visible camera images are used for automatic recognition of MARFE on JET.[19–21] Apart from directly identifying MARFE with diagnostic signals, the threshold of MARFE onset and the causes

[†]Corresponding author. E-mail: jlhou@ipp.ac.cn

© 2023 Chinese Physical Society and IOP Publishing Ltd

of MARFE movement have been investigated extensively with theory, modeling, and experiments. In Ref. [22], the gradient of electron temperature is regarded as the driver of MARFE instability. From a study of Ref. [23], MARFE becomes unstable and moves when radiation rate reaches a critical value. A linear instability analysis shows that higher wall desorption rate leads to higher MARFE density threshold.[24] From a two-dimensional nonlinear fluid model study, the MARFE is presented when there is sufficient radiative cooling.[25] Experiments on JT-60 show higher neutral beam power helps increase MARFE threshold.[26] In EAST's pellet injection discharges, the MARFE becomes unstable and moves when the symmetry of heat flow on the two sides of it is broken.[27] The experiment on J-TEXT shows that localized current shrinkage happens before MARFE causes disruption.[9]

The above results show that there is no unified physical model to use single or few plasma parameters to give warnings of MARFE movement. In recent years, machine learning (ML) techniques are popular to predict different instability events for whom the physics mechanism is complex and to whom multiple physics processes and physics parameters are related, such as thermal quench and current quench of major disruptions,[28,29] vertical displacement events (VDE),[30] and tearing modes prediction.[31,32] Among these frequently-used ML methods, random forest (RF) is a kind of model that has been applied for different tasks such as tearing modes prediction and avoidance,[33] resistive wall mode stability prediction[34] and disruption prediction.[35–38] Besides, from comparison of various ML models the RF model works best for the task of fault-detection for magnetic probes in terms of speed and accuracy.[39] Similar to RF, another ensemble learning method named LightGBM is used for cross-machine disruption prediction on J-TEXT and HL-2A.[40] Above all, what makes RF model more attractive is its interpretable characteristics, i.e., it is able to decompose the final output into the contributions of each input feature named feature contribution. By comparing the feature contributions, one can extract the main causes of the final output[41] which helps guide the control system to take corresponding actions with regard to the prediction results. This paper reports an RF model to predict MARFE movement on EAST. The rest of the paper is organized as follows. Section 2 gives a brief introduction of the RF model. Section 3 is about training data collection and labeling, and the prediction results on test discharges. Section 4 is the discussion of the work and future outlook about application of the MARFE movement prediction model.

2. Brief introduction to random forest

2.1. Decision tree and ensemble learning

The RF is a typical example of ensemble learning with the bootstrap aggregating (bagging) method. The basic idea of ensemble learning is to combine a collection of base learners and aggregate their results as final output to reduce the model errors.[42] For RF models, the base learner is the decision tree.[43] Each regression or classification decision tree is parallelly trained with a subset of samples randomly bagged from the whole training data set. These independent decision trees compose an RF model and the output of the RF model is the average of these decision trees' outputs. For example, an RF model consists of T decision tree learners, then the RF output Y_{RF} is calculated as the mean value of single decision tree's output

$$Y_{\mathrm{RF}} = \frac{1}{T}\sum_{t=1}^{T} Y_t. \tag{1}$$

The growth of a classification tree is the recursive process of splitting a father-node into two child-nodes until the maximum depth is reached or the child-node does not need to be split anymore, i.e., it only contains samples with the same label. If a node does not have a child-node it becomes a leaf of the tree. Consider a dataset of N samples and a sample consisting of F features (parameters) that are represented as $X = [x_1, x_2, ..., x_i, ..., x_F]$. Figure 1 shows a parent node containing S samples is split into two children nodes. Here $D_{\mathrm{P}} = \{X_1, X_2, \ldots, X_S\}$ and $D_{\mathrm{P}} = D_{\mathrm{L}} + D_{\mathrm{R}}$.

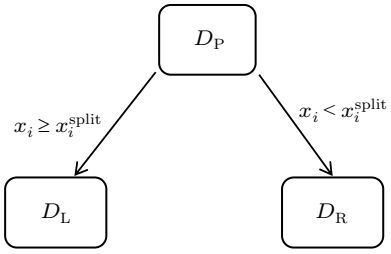


Fig. 1. A parent node is split into two child nodes, the left node and the right node according to the splitting condition of whether the value of the i-th feature x_i is greater or less-equal than the split value x_i^{split}. The set of instances in parent is also split into two child nodes $D_{\mathrm{P}} = D_{\mathrm{L}} + D_{\mathrm{R}}$.

The splitting includes two steps: (i) find the feature for splitting and (ii) find the splitting value of this feature. Both steps are conducted with two cycles: find the splitting value of a feature that maximizes Gini gain, then find the feature that has maximum Gini gain. Here the Gini index of a set that has K kinds of labels is defined as

$$\mathrm{Gini} = 1 - \sum_{k=1}^{K} p_k^2, \tag{2}$$

where p_k is the proportion of samples with the k-th label in the dataset. A dataset having lower Gini index corresponds to a purer dataset. When a set has only one kind of sample, its Gini index is 0. After splitting from a parent node to two child nodes, the Gini gain is calculated using

$$\mathrm{Gini}_{\mathrm{gain}} = \mathrm{Gini}_{\mathrm{P}} - (\mathrm{Gini}_{\mathrm{L}} + \mathrm{Gini}_{\mathrm{R}}). \tag{3}$$

In essence, the decision tree with M leaves divides the feature space into M regions/subspaces. The structure of a tree can

be represented by all the nodes and all routines to the leaf of it. Each node contains a set of samples D_{node} with size S_{node} inside it. The value of a node of the k-th class is represented by the proportion of k-th samples in this node

$$Y_{\text{node}} = \left[p^1_{\text{node}}, p^2_{\text{node}}, \ldots, p^K_{\text{node}}\right],$$
$$p^k_{\text{node}} = \frac{N^k}{S_{\text{node}}}, \ \sum_{k=1}^{K} p^k_{\text{node}} = 1, \tag{4}$$

where N^k is the number of samples of k-th class in the node. In this paper for the task of MARFE-movement prediction, there are only two kinds of labels positive and negative, so that $K = 2$ and the value of a node can be represented as $Y_{\text{node}} = \left[p^{\text{negative}}_{\text{node}}, p^{\text{positive}}_{\text{node}}\right]$.

2.2. Interpretation of random forest

From the interpretation method of RF model in Ref. [44], the interpretability of RF prediction is extracted from every splitting of decision trees. Inside a decision tree, the value of a node is the vector that contains proportions of the samples of different classes as shown in Eq. (4). Inputting a test sample X_{test} to one of the decision trees of the RF model, before it travels to the leaf and gets its final output it will pass a series of nodes. Each time it passes from a parent node to a child node by the splitting performed over some feature f, the value change between parent node and children node is attributed to that feature and is calculated as the feature's contribution at this split

$$C^f_{t,\text{test}} = Y_{t,\text{P}} - Y_{t,\text{C}}. \tag{5}$$

By this way the contributions of each feature to the final output are extracted for one decision tree. For the whole RF model, feature contributions are the mean value of all of the decision trees

$$C^f_{\text{test}} = \sum_{t=1}^{T} C^f_{t,\text{test}}. \tag{6}$$

Now the output of the RF model can be represented as the following equation:

$$Y_{\text{RF}} = \frac{1}{T}\sum_{t=1}^{T}\left[\overline{Y}_t + \sum_{f=1}^{F} C^f_t\right], \tag{7}$$

where $\overline{Y}_t$ is the value of the root node of t-th tree. This method of RF model interpretation is applied for understanding the MARFE movement prediction results in the following section.

3. Training data collection and labeling

3.1. Training database collection

Disruption is one of the macro-instabilities of tokamak. During major disruption, plasma current decreases to zero instantly with the speed larger than ~ 2 MA/s. Following current quench is huge thermal and mechanical load to the first wall. Due to enormous damage to the device, disruption remains one of the biggest challenges for tokamak. Direct causes of disruption include high density limit, low q limit, error filed caused locked-modes and plasma elongation caused vertical instability. Density limit disruption usually happens in high density conditions. When increasing plasma density close to or above Greenwald density limit, enhancement of edge radiation might cause plasma cooling and current shrinkage, and consequently plasma disruption. The target of this predictor is to identify the MARFE movement causing density-limit disruption while avoiding false alarms for non-disruptive plasmas and non-density-limit disruptive plasmas, i.e., plasma disruptions due to other causes instead of density limit. In order to train a MARFE movement predictor, density-limit disruptive discharges caused by MARFE movement are used to collect MARFE-unstable samples and non-disruptive discharges are used to collect MARFE-stable samples. Here MARFE-unstable samples refer to the samples close to MARFE movement and MARFE-stable samples refer to the samples far away from MARFE movement. During the 2022's first campaign of EAST, as listed in Table 1, there are 43 densitylimit disruptive discharges due to MARFE movement, 2322 non-density-limit disruptive discharges and totally 4297 non-disruptive discharges.

Table 1. Number of density-limit disruptive and non-density-limit disruptive discharges and non-disruptive discharges of the EAST first campaign of 2022.

Discharge type	density-limit disruption	non-density-limit disruption	non-disruption
Description	disruption due to MARFE movement	disruption due to other causes instead of MARFE movement	no disruption discharges
Number of discharges	43	2322	4279

Table 2. Experiment parameters of density ramp-up experiment of the first EAST campaign of 2022. Including power range of LHW, NBI and ECH, plasma current and plasma configuration.

Experiment parameters	P_{LHW} (MW)	P_{NBI} (MW)	P_{ECH} (MW)	I_{p} (MA)	Configuration
Range	0–4	0–7	0–2	0.25–0.6	upper single null, lower single null, double null, quasi-snowflake

The density ramp-up experiments are carried out with various conditions and 43 density-limit disruptive discharges due to MARFE movement are picked out from them. The main experiment parameters and their range is listed in the Table 2. It has to be noted that for all these density limit disruptive discharges, besides normal gas puffing, their density ramps up under pellet injection at the high field side and supersonic molecular beam injection is added for some of them. Their

auxiliary heating power mixture is shown in Fig. 2, including neutral beam injection (NBI), lower hybrid wave heating (LHW) and electron cyclotron heating (ECH). These dots are samples collected with a time interval of 100 ms from density limit disruption in these 43 density-limit disruptive discharges. A discharge is also often called as shot.

Table 3. Plasma parameters used as input features of RF model.

Input plasma features	Description and calculation
Ip_{errNorm}	normalized plasma current control error $\frac{Ip-Ip_{\mathrm{target}}}{Ip_{\mathrm{target}}}$
z_{errNorm}	normalized vertical displacement control error $\frac{z-z_{\mathrm{target}}}{a_{\mathrm{minor}}}$
H_{98}	confinement factor
$\mathrm{GW}_{\mathrm{frac}}$	Greenwald fraction $\frac{n_{\mathrm{e}}}{n_{\mathrm{GW}}}$
$\mathrm{d}W_{\mathrm{mhd}}/\mathrm{d}t$	time derivative of plasma stored energy
$\mathrm{d}\beta_{\mathrm{p}}/\mathrm{d}t$	time derivative of poloidal beta
κ	plasma shape elongation
β_{N}	normalized beta
q_{95}	edge safety factor

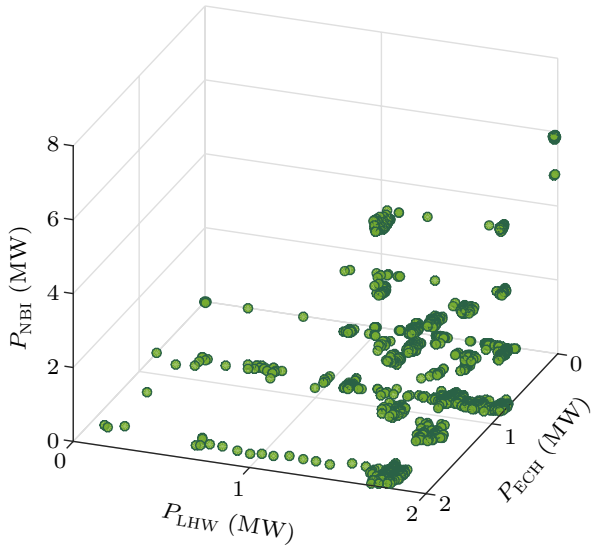


Fig. 2. Distribution of NBI, LHW and ECH powers for all samples from density ramp-up discharges. A dot represents a sample of plasma condition that is heated with mixture of $[P_{\mathrm{LHW}}, P_{\mathrm{ECH}}, P_{\mathrm{NBI}}]$.

As normally applied in machine learning, two-thirds of these 43 discharges are used for training and the other one third of them are used for testing. For density-limit disruptive shots, 28 of them are randomly selected for training and rest 15 shots are for testing. For non-density-limit disruptive and non-disruptive, after deleting the shots with signal missing, totally 5280 shots are left. $\sim$ 3520 shots are randomly selected for training and $\sim$ 1760 are randomly selected for testing. The plasma parameters (features) used to build the training samples are listed in Table 3. Here, W_{mhd}, β_{p}, β_{N}, κ and q_{95} are from equilibrium reconstruction by EFIT. $\mathrm{d}W_{\mathrm{mhd}}/\mathrm{d}t$ is time derivative of W_{mhd} and $\mathrm{d}\beta_{\mathrm{p}}/\mathrm{d}t$ is time derivative of β_{p}. Ip_{errNorm} and z_{errNorm} represent normalized plasma current control error and normalized vertical displacement control error, respectively. The difference between actual values and target values indicates their control errors. Ip error is normalized with Ip_{target} and z error is normalized with minor radius calculated by EFIT.

3.2. Database labeling and RF model training

The RF is a supervised learning method. The RF models are trained by samples with labels. In tokamak plasma abnormal events prediction research, the most frequently used ways to assign labels to samples are using a step function to divide them into two classes according to a classification time constant τ_{class} or using a sigmoid function to assign continuous labels to them.[45–47] For the former labeling method, if denoting the abnormal event time as t_{event}, then samples between $[t_{\mathrm{event}} - \tau_{\mathrm{class}}, t_{\mathrm{event}}]$ are labeled as positive and samples before $t_{\mathrm{event}} - \tau_{\mathrm{class}}$ are labeled as negative. In studies of disruption prediction on HL-2A[48,49] the classification time is 100 ms and on J-TEXT it is 50 ms.[40] In resistive wall mode prediction on NSTX the classification time is also set to 100 ms.[34] Similarly, for MARFE movement prediction, a classification time of $\tau_{\mathrm{class}} \sim 100$ ms regarded to MARFE movement time $t_{\mathrm{MARFEmove}}$ is applied for sample labeling. For example, in Fig. 3, #115858 is a typical density-limit disruption discharge whose MARFE appears at $\sim$ 8.4 s and starts to move after $\sim$ 8.5 s which is marked as $t_{\mathrm{MARFEmove}}$ and leads to disruption at 9.3 s.

Training samples are taken from both density-limit disruptive, non-disruptive discharges and other disruptive shots. Their labeling is shown in Table 4. Positive samples are taken from 100 ms before $t_{\mathrm{MARFEmove}}$ of density limit disruptive shots with a sampling interval of 1 ms. One part of negative samples is from time of more than 100 ms before $t_{\mathrm{MARFEmove}}$ with a sampling interval of 10 ms. Another part of negative samples is taken from the rest of shots including non-disruptive and non-density-limit disruptive shots. With a sampling frequency of 1 kHz, $\sim$ 2800 positive samples are collected from the MARFE-unstable phase of density-limit disruptive shots. At the same time, the negative samples are collected with a sampling frequency of 100 kHz. Because the total number of collected negative samples is much greater than the number of positive samples, in order to avoid imbalance of positive and negative samples, only $\sim$ 5600 negative samples are randomly selected from the whole collected dataset. They are composed of $\sim$ 2800 randomly selected negative samples collected from MARFE-stable phase of density-limit shots and $\sim$ 2800 negative samples taken from the flattop phase of non-disruptive and non-density-limit disruptive shots. In other words, positive samples correspond to MARFE-stable phase and negative samples correspond to MARFE-unstable phase.

After building the training dataset, the RF model is built with an open-source library scikit-learn (https://scikit-

learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html). As bootstrap sampling method is applied for RF training, only part of samples are randomly selected from the whole dataset to build each decision tree and the rest left-out samples which are also named out-of-bag samples are used to estimate the generalization score of the tree. The first parameter that needs to be tuned is the number of decision trees that consist of the RF model. By scanning the number of trees, the mean out-of-bag score is recorded in Fig. 4. By increasing the number of decision trees, the out-of-bag score increases from 0.8 to 0.99. The out-of-bag score changes little after the number of trees reaches above 20 and it becomes stable when number of trees reaches 80. Therefore, with considering the computation resource, the number of decision tree is set to 100 for the RF models. Besides, in order to avoid overfitting, the minimum number of samples in a leaf that can be split is set to 10. In other words, a node with less than 10 samples will not be split anymore.

Table 4. Labeling of density-limit disruptive, non-disruptive and non-density-limit disruptive shots. Here positive labels are encoded as 1 and negative labels are encoded as 0.

Shot type	Time period	Sampling interval	Label
Density-limit (DL) disruption	$(t_{\text{MARFEmove}} - 100\ \text{ms}, t_{\text{MARFEmove}})$	1 ms	1
	$(t_{\text{FlattopStart}}, t_{\text{MARFEmove}} - 100\ \text{ms})$	10 ms	0
No-disruption	$(t_{\text{FlattopStart}}, t_{\text{FlattopEnd}})$	10 ms	0
Non-density-limit disruption	$(t_{\text{FlattopStart}}, t_{\text{disrupt}})$	10 ms	0

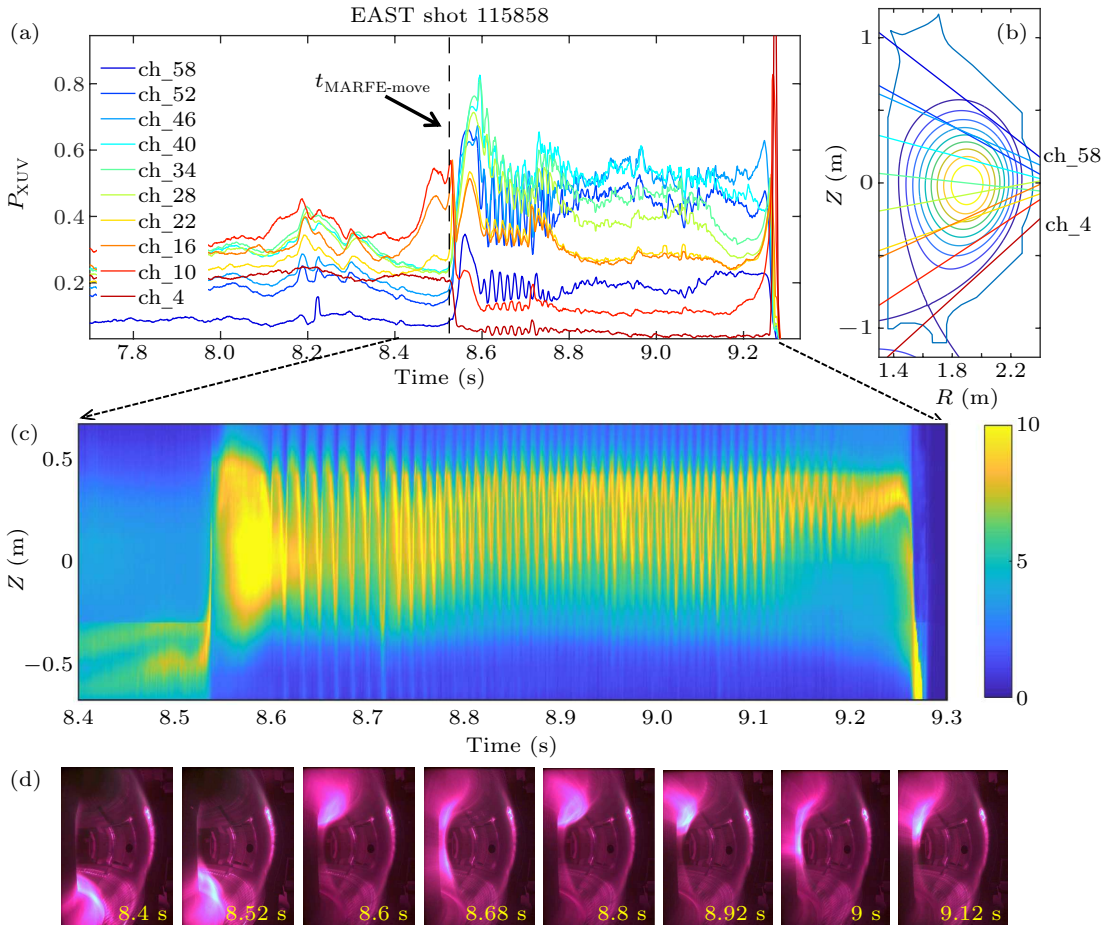


Fig. 3. (a) Time traces of XUV radiation signals for different measurement channels of discharge #115858. (b) XUV channels measurement coverage from top to bottom corresponding to the radiation signal of the same color. (c) The 1-D contour of radiation between time [8.4 s, 9.3 s] reconstructed from XUV diagnostic. (d) CCD images of plasma at different time slices showing MARFE movements.

Figure 5 is one of the decision trees of RF model with only the first three layers shown. In the root node 4680 samples are randomly selected from whole dataset to train this decision tree. Starting from the root node, the first splitting is according to the value of β_{N}. On the second layer, samples with $\beta_{\text{N}} \leq 1.028$ are allocated to the left child node and samples with $\beta_{\text{N}} > 1.028$ are allocated to the right child. The second splitting of the left branch is according to q_{95}. Samples with $q_{95} \leq 8.118$ are allocated to the left child and samples with $q_{95} > 8.118$ are allocated to the right child node of the third layer and so on. As introduced in Section 2, the value of each node is the proportion of negative and positive samples of this

node $Y_{\text{node}} = \left[p_{\text{node}}^{\text{negative}}, p_{\text{node}}^{\text{positive}}\right] = \left[1 - p_{\text{node}}^{\text{positive}}, p_{\text{node}}^{\text{positive}}\right]$. If proportion of positive samples is larger than negative samples, $p_{\text{node}}^{\text{negative}} < p_{\text{node}}^{\text{positive}}$, this node is classified as MARFE-unstable. If a node contains more negative samples than positive samples, $p_{\text{node}}^{\text{negative}} > p_{\text{node}}^{\text{positive}}$, then it is classified as MARFE-stable. If all samples in this node are positive labels, then $p_{\text{node}}^{\text{positive}} = 1$. Similarly, if all samples in this node are labeled as negative, then $p_{\text{node}}^{\text{positive}} = 0$. Higher positive value means higher MARFE-risk for the subspace corresponding to this node. As marked by the pink rectangular, the left child node of the third layer is a node with only negative samples and therefore has a node value of $[1,0]$. It can no longer split and become a leaf because its Gini index is zero and can no longer decrease. As a result, this leaf is classified as MARFE-stable. According to the path that splits to this leaf, it is corresponding to a subspace of $\{\beta_{\text{N}} > 1.028$ and $q_{95} \leq 8.118\}$.

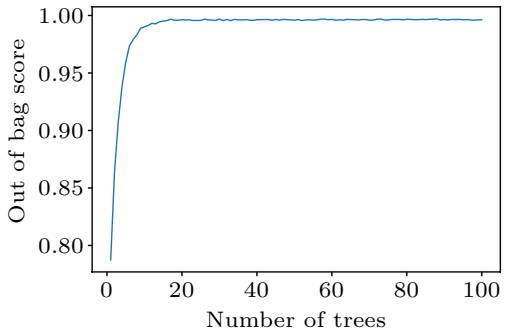


Fig. 4. Out-of-bag score of RF models with increasing number of decision trees.

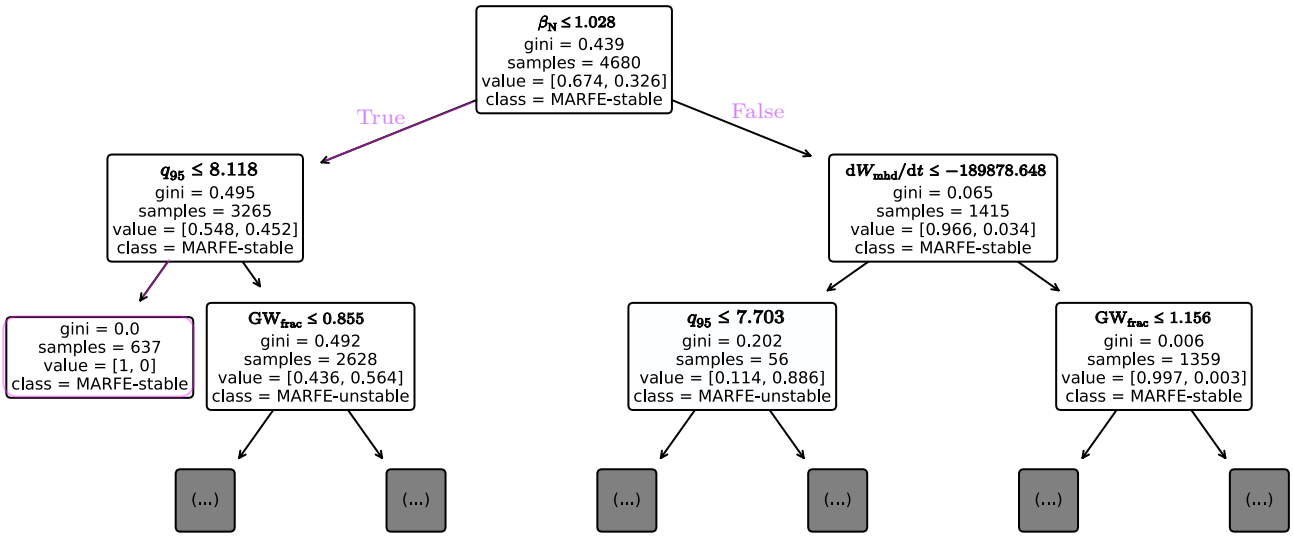


Fig. 5. The structure of first three layers of one of the decision trees of the RF model. The pink line highlights a routine to a leaf that is classified as MARFE-stable with all the samples inside it are negative and the Gini index of it is zero.

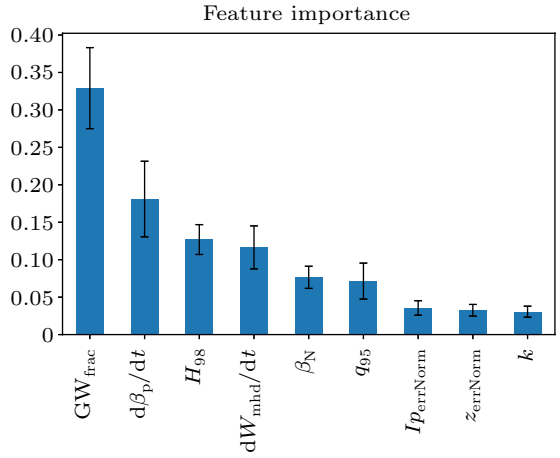


Fig. 6. Feature importance of the input parameters of RF model calculated with the mean impurity decrease method. The parameter that is used to split a dataset and results in larger impurity decrease after splitting is assigned with relatively higher feature importance. Standard deviation is used as their error bar.

Figure 6 is the feature importance of the RF model calculated through mean decrease impurity. As the number of test density limit disruption shots is only 15, to increase the generalizability of the results, the training and testing process is carried out repeatedly five times. At each round of training and testing, the training and testing shots are randomly selected from the shots listed in Table 1. The feature importances plotted in Fig. 6 are mean values of five rounds' training and testing. It indicates that the most important feature for MARFE-move detection is GW_{frac}, which is reasonable because MARFE is closely related to density limit disruption. The relative important features are $\text{d}\beta_{\text{p}}/\text{d}t$, $\text{d}W_{\text{mhd}}/\text{d}t$ and H_{98} while Ip_{errNorm}, z_{errNorm} and κ are relatively unimportant features.

3.3. RF model testing and overall performance

After training, the RF model is tested with the test shots. For all testing shots, their parameters are taken the same way as the training shots including the time derivative parameters, except that all the input parameters for test shots are then interpolated to a sampling frequency of 1 kHz, so that a value of MARFE-risk can be outputted every 1 ms from the RF model,

which is what should be like in real-time experiments. For a DL disruptive test shot if alarming conditions are satisfied and an alarm can be given out, then the alarm is regarded as a successful alarm (SA). While for non-disruptive and non-DL disruptive test shots the alarms for them are regarded as false alarm (FA). The SA rate is defined as ratio of number of SA shots and number of test DL disruptive shots. The FA rate is defined as ratio of number of FA shots and number of test non-disruptive and non-DL disruptive shots

$$\text{SA rate} = \frac{N_{\text{SA}}}{N_{\text{DL-disrupt}}}, \quad \text{FA rate} = \frac{N_{\text{FA}}}{N_{\text{non-disrupt}} + N_{\text{non-DL-disrupt}}},$$

where N_{SA} is the number of SA shots, $N_{\text{DL-disrupt}}$ is the number of density-limit disruptive test shots, $N_{\text{non-disrupt}}$ is the number of non-disruptive test shots and $N_{\text{non-DL-disrupt}}$ is the number of non-DL disruptive test shots.

Figure 7 shows the SA rate and FA rate for different thresholds assigned to MARFE risk alarming. Similar to Fig. 6, SA rate, mean SA time and FA rate plotted in Fig. 7 are mean values of five rounds' training and testing. In Fig. 7, with increasing alarm threshold, FA rate decreases while SA rate and mean SA time decrease as well. The SA time is the time when an alarm is given before MARFE movement. This is reasonable, because with a higher alarm threshold, alarm condition is harder to meet and SA and FA rates get lower together. For alarm threshold of 0.4, the SA rate is $\sim 95\%$ and FA rate is $\sim 4\%$. While for alarm threshold of 0.6, the SA rate is $\sim 80\%$ and FA rate is $\sim 1\%$. To get a balance between SA rate and FA rate, the threshold can be set to 0.5. Under such conditions, the SA rate is $\sim 87\%$ and FA rate is $\sim 2\%$ so that most MARFE instability can be detected while false alarm rate is as low as can be acceptable for EAST experiment.

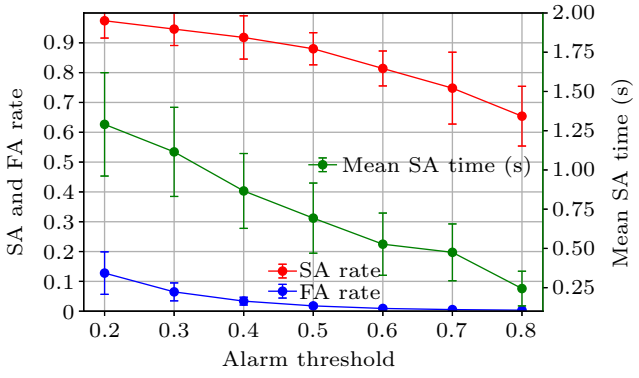


Fig. 7. SA rate and FA rate with different alarm thresholds. Mean alarm time for SA shots. The success alarm time is calculated as SA time = $t_{\text{MARFEmove}} - t_{\text{alarm}}$. Mean SA-time is the mean of SA time for all SA shots. Standard deviation is used as their error bar.

When setting alarm threshold to be 0.5, the mean SA time in Fig. 7 is shown to be ~ 550 ms and the accumulation fraction of detected MARFE-movement for DL-disruptive shots is shown in Fig. 8. It shows that around 85% of unstable MARFE (MARFE movement) can be detected at 40 ms before it, which is at least enough for disruption prediction mitigation with massive gas injection or shattered pellet injection. But a better way to handle MARFE instability is to suppress it and avoid disruption.

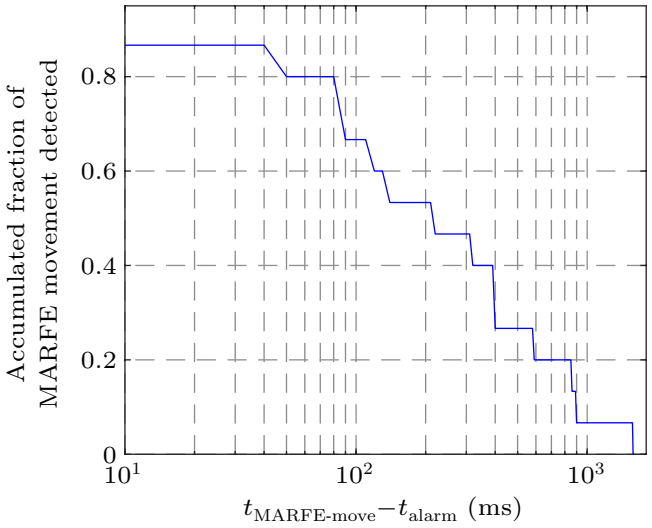


Fig. 8. Accumulated fraction of MARFE-movement detected for density-limit disruptive shots for different $t_{\text{MARFEmove}} - t_{\text{alarm}}$ when alarm threshold is set to be 0.5.

3.4. MARFE movement prediction results

The #113323 is a density-limit disruptive shot that is not used for training. The plasma current, GW_{frac}, $\mathrm{d}\beta_{\text{p}}/\mathrm{d}t$ and H_{98} of this shot are plotted in Figs. 9(a)–9(d). GW_{frac}, $\mathrm{d}\beta_{\text{p}}/\mathrm{d}t$ and H_{98} are the three relatively important features of the RF model as shown in Fig. 6. In Fig. 9(b), GW_{frac} increases as density ramps up above Greenwald density. Until GW_{frac} reaches ~ 1.1, the sudden drop of GW_{frac} at 6.28 s implies that MARFE movement happens which causes density diagnostic signal to drop suddenly and leads to the drop of GW_{frac} signal. As a consequence, in Fig. 9(a), plasma current signal shows that disruption happens ~ 300 ms after MARFE movement. Applying the RF model to it, the output value MARFE-risk is plotted in Fig. 9(e). If setting the alarming conditions to be reaching alarming threshold of 0.5 and lasting for a duration longer than 10 ms, an alarming would be given at t_{alarm} as noted in Fig. 9(e). Feature contributions in Fig. 9(f) show that it is the $\mathrm{d}\beta_{\text{p}}/\mathrm{d}t$ (orange), $\mathrm{d}W_{\text{mhd}}/\mathrm{d}t$ (red) and GW_{frac} (purple) that mainly contribute to increase of MARFE risk at around ~ 6 s. A more clear view of MARFE risk enhancement and the three main contributions of it are shown in Figs. 9(g) and 9(h) that are the zoom in Figs. 9(e) and 9(f) between time [5.8 s, 6.3 s]. Figures 9(c) and 9(d) show that before MARFE movement, $\mathrm{d}\beta_{\text{p}}/\mathrm{d}t$, $\mathrm{d}W_{\text{mhd}}/\mathrm{d}t$ and H_{98} have an obvious decrease, which corresponds to MARFE-risk increase of Fig. 9(e).

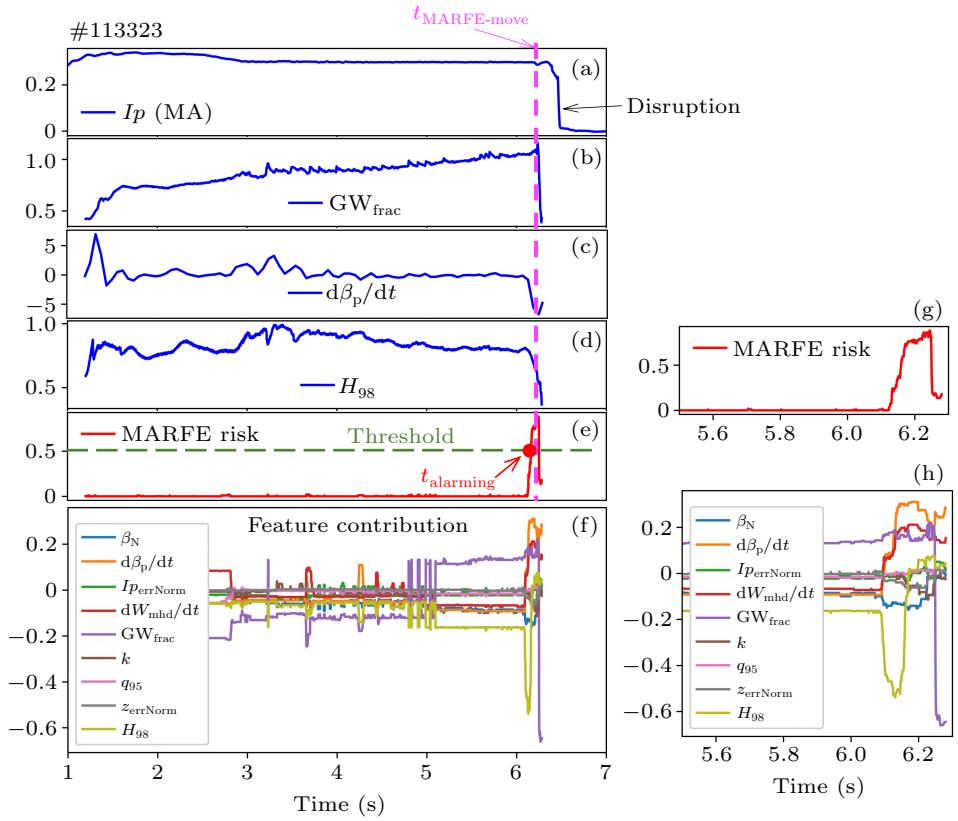


Fig. 9. Time traces of parameters of an example of a successful alarm shot. (a) The plasma current I_{p}, (b) $\mathrm{GW_{frac}}$ signal, (c) $\mathrm{d}\beta_{\mathrm{p}}/\mathrm{d}t$ signal, (d) H_{98} signal of shot #113323. (e) The output of RF model which is represented as MARFE risk identifying the probability of instability MARFE. (f) The feature contributions that consist of the MARFE risk. (g) Zoom in of MARFE risk signal between time [5.8 s, 6.3 s]. (h) Zoom in of feature contribution signals between time [5.8 s, 6.3 s].

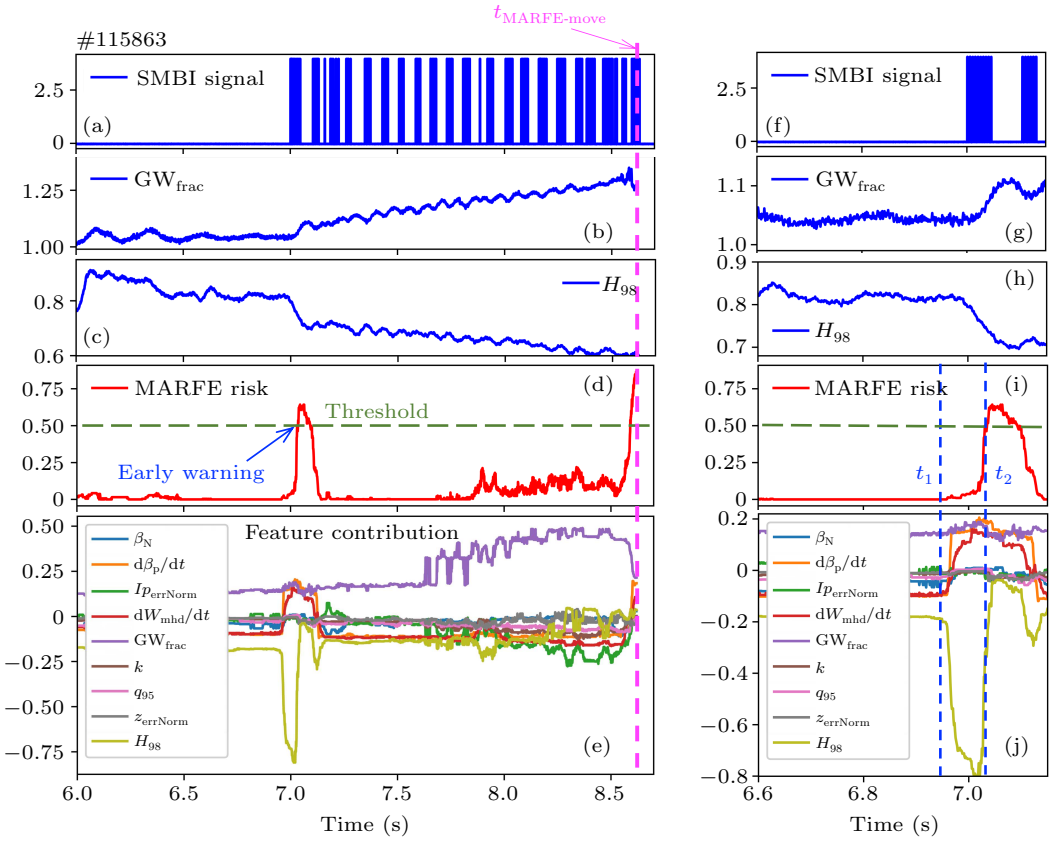


Fig. 10. The #115863 is an early warning shot. Panels (a)–(e) are time traces of its discharge signals, the RF model evaluated MARFE-risk and its feature contributions. Panels (f)–(j) are zoom-in images of left images.

3.5. Early warning analysis

Figure 8 shows that there are a few early warning shots whose alarming time is more than 1 s before MARFE movement. For example, #115863 is an early warning shot that reaches the alarming threshold first at $t_2 = 7.02$ s which is ~ 1.56 s before MARFE movement and is therefore regarded as early warning. In Figs. 10(i) and 10(e), together with rapid enhancement of $\mathrm{d}\beta_\mathrm{p}/\mathrm{d}t$ and $\mathrm{d}W_\mathrm{mhd}/\mathrm{d}t$ feature contributions, its MARFE-risk starts to increase from t_1. While the sharp drop of H_{98} feature contribution helps MARFE-risk maintains a low level below the threshold until t_2 when a sudden increase of H_{98} feature contribution causes MARFE-risk quickly gets to the alarming threshold leading to early warning. Figure 10(c) shows that the increase of H_{98} feature contribution at around 7.01 s is caused by fast reduction of confinement factor H_{98} from 0.8 to 0.7. Figures 10(g) and 10(h) show that the decline of H_{98} is a consequence of increase of plasma density as shown in Fig. 10(b). Figure 10(a) shows that it is the opening of super-sonic molecular beam injection (SMBI) that results in rise of plasma density. Here the opening of SMBI is pre-programmed for the purpose of density ramp up experiment.

It indicates that besides plasma density, confinement and equilibrium parameters, other parameters such as radiation and temperature signals should be applied for MARFE movement prediction, in order to avoid early warning caused by sudden change of plasma density or confinement.

4. Summary and outlook

Prediction of MARFE movement is the first step towards actively handling density limit disruption and maintaining a steady-state high density operation. This paper reports an RF model to predict movement of MARFE for high density discharges on EAST. The model is able to classify MARFE movement with an accuracy of 85%–90% for density limit disruption discharges, while the false alarm for non-density-limit disruption and non-disruption discharges is below 5%. Except for the $\mathrm{GW}_\mathrm{frac}$, relatively important parameters for MARFE movement prediction are $\mathrm{d}\beta_\mathrm{p}/\mathrm{d}t$, $\mathrm{d}W_\mathrm{mhd}/\mathrm{d}t$ and H_{98}. The results show that it is feasible to use a machine learning model to forecast MARFE movement before density limit disruption while avoiding false alarms for discharges without unstable MARFE, which is helpful to design the high-density operation scenario for future fusion reactor. For instance, when the unstable MARFE is predicted, we can increase the auxiliary heating power or reduce the edge gas puffing to mitigate or avoid it.

On the other hand, density limit disruption will happen easily if the unstable MARFE is difficult to mitigate or avoid. Then some external means will have to be used to mitigate the disruption. At present, a major gas injection system and shattered pellet injection system have been built for disruption mitigation on EAST. And some experiments for disruption mitigation have been performed with these systems. Future work will be implemented the MARFE movement predictor into EAST plasma control system for real-time MARFE instability detection in high-density experiments.

However, it should be noted that this model in the paper is only trained with limited density limit data (totally 43 shots) from EAST now. Thus, future work will be explored using more density limit discharges data of EAST to increase model's successful alarm rate. Meanwhile, besides the 9 parameters used in this model, it is worth investigating more effective parameters to avoid false alarm and early warning. What's more, it is of significant importance to use cross-machine density limit data and cross-machine unified parameters to train the MARFE movement predictor so that the model can be generalized to be applied to future devices.

Acknowledgments

This work is supported by the National MCF Energy R&D Program of China (Grant Nos. 2018YFE0302100 and 2019YFE03010003), the National Natural Science Foundation of China (Grant Nos. 12005264, 12105322, and 12075285), the National Magnetic Confinement Fusion Science Program of China (Grant No. 2022YFE03100003), the Natural Science Foundation of Anhui Province of China (Grant No. 2108085QA38), the Chinese Postdoctoral Science Found (Grant No. 2021000278), and the Presidential Foundation of Hefei institutes of Physical Science (Grant No. YZJJ2021QN12).

References

[1] Connor J W and You S 2001 *Plasma Phys. Control. Fusion* **44** 121
[2] ITER Physics Expert Group on Disruptions, Plasma Control, MHD and ITER Physics Basis Editors 1999 *Nucl. Fusion* **39** 2251
[3] Wesson J A, Gill R D, Hugon M, *et al.* 1989 *Nucl. Fusion* **29** 641
[4] Stabler A, McCormick K, Mertens V, Muller E R, Neuhauser J, Niedermeyer H, Steuer K H, Zohm H, Dollinger F, Eberhagen A, Fussmann G, Gehre O, Gernhardt J, Hartinger T, Hofmann J V, Kakoulidis E, Kaufmann M, Kyriakakis G, Lang R S, Murmann H D, Poschenrieder W, Ryter F, Sandmann W, Schneider U, Siller G, Soldner F X, Tsois N, Vollmer O and Wagner F 1992 *Nucl. Fusion* **32** 1557
[5] Hosogane N, Asakura N, Kubo H, Itami K, Sakasai A, Shimizu K, Nakamura H, Shimada M, Neyatani Y and Yoshino R 1992 *J. Nucl. Mater.* **196–198** 750
[6] Samm U, Brix M, Durodié F, Lehnen M, Pospieszczyk A, Rapp J, Sergienko G, Schweer B, Tokar M Z and Unterberg B 1999 *J. Nucl. Mater.* **266–269** 666
[7] Greenwald M 2002 *Plasma Phys. Control. Fusion* **44** R27
[8] Schuller F C 1995 *Plasma Phys. Control. Fusion* **37** A135
[9] Shi P, Zhuang G, Gentle K, Hu Q, Chen J, Li Q, Liu Y, Gao L, Zhang X, Liu H, Chen Z, Zhu L, Li F, Zhou Y, Zeng Z, Liu L and He J 2017 *Nucl. Fusion* **57** 116052
[10] Zheng X, Li J, Hu J, Liu H, Jie Y, Wang S, Li J, Duan Y, Li M, Li Y, Zhang L, Ye Y, Yang Q, Zhang T, Cheng Y, Xu J, Wang L, Xu L, Zhao H, Wang F, Lin S, Wu B, Lyu B, Xu G, Gao X, Shi T, He K, Lan H, Chu N, Cao B, Sun Z, Zuo G, Ren J, Zhuang H, Li C, Yuan X, Yu Y, Wang H, Chen Y and Wu J 2016 *Plasma Phys. Control. Fusion* **58** 55013

[11] Sen A K 1993 *Phys. Fluids B: Plasma Phys.* **5** 3997

[12] de Vries P C, Rapp J, Schüller F C and Tokar M Z 1998 *Phys. Rev. Lett.* **80** 3519

[13] Dachicourt R, Monier-Garbet P, Gil C, Guirlet R, Tamain P, Meyer O, Devynck P, Pégourié B, Clairet F, Bucalossi J, Corre Y and Ségui J L 2013 *J. Nucl. Mater.* **438** S334

[14] Gao X, Zhao Y P, Luo J R, *et al.* 1999 *Plasma Phys. Control. Fusion* **41** 1349

[15] Gao X, Xie J K, Wan Y X, Ushigusa K, Wan B N, Zhang S Y, Li J and Kuang G L 2001 *Phys. Rev. E* **65** 017401

[16] Stroth U, Bernert M, Brida D, Cavedon M, Dux R, Huett E, Lunt T, Pan O, Wischmeier M and ASDEX Upgrade Team 2022 *Nucl. Fusion* **62** 076008

[17] Arena P, Basile A, De Angelis R, Fortuna L, Mazzitelli G, Migliori S, Vagliasindi G and Zammataro M 2005 *IEEE Trans. Plasma Sci.* **33** 1106

[18] Vagliasindi G, Murari A, Arena P, Fortuna L and Mazzitelli G 2009 *IEEE Trans. Instrum. Meas.* **58** 2417

[19] Murari A, Camplani M, Cannas B, Mazon D, Delaunay F, Usai P and Delmond J F 2010 *IEEE Trans. Plasma Sci.* **38** 3409

[20] Portes de A M, Portes de A M, Chacon G T, de Faria E L and Murari A 2012 *IEEE Trans. Plasma Sci.* **40** 3485

[21] Portes de A M, Murari A, Giovani M, Alves N and Portes de A M 2013 *IEEE Trans. Plasma Sci.* **41** 341

[22] Mazzotta C, Spizzo G, Pucella G, Giovannozzi E, Tudisco O, Apruzzese G, Bin W and Esposito B 2017 *Nucl. Mater. Energy* **12** 808

[23] Stacey W M 2007 *Fusion Sci. Technol.* **52** 29

[24] Marchuk O, Tokar M Z and Kelly F A 2006 *Contrib. Plasma Phys.* **46** 744

[25] Chen X P and Shi B R 1999 *Commun. Theor. Phys.* **31** 625

[26] Nishitani T, Ishida S, Hosogane N, Sugie T, Itami K and Takeuchi H 1990 *J. Nucl. Mater.* **176–177** 763

[27] Hou J, Chen Y, Zuo G, Hu J, Mao S, Yuan X, Huang J, Wu M, Xu L, Zhao H, Yuan J, Wang S, Liu H, Meng L, Shi T, Li P and Li J 2022 *Plasma Phys. Control. Fusion* **64** 055010

[28] Kates-Harbeck J, Svyatkovskiy A and Tang W 2019 *Nature* **568** 526

[29] Vega J, Murari A, Dormido-Canto S, *et al.* 2022 *Nat. Phys.* **18** 741

[30] Sammuli B S, Barr J L and Humphreys D A 2021 *Fusion Eng. Des.* **169** 112492

[31] Olofsson K E J, Humphreys D A and Haye R J L 2018 *Plasma Phys. Control. Fusion* **60** 084002

[32] Pucella G, Buratti P, Giovannozzi E, Alessi E, Auriemma F, Brunetti D, Ferreira D R, Baruzzo M, Frigione D, Garzotti L, Joffrin E, Lerche E, Lomas P J, Nowak S, Piron L, Rimini F, Sozzi C, Van Eester D and JET Contributors 2021 *Nucl. Fusion* **61** 046020

[33] Fu Y, Eldon D, Erickson K, Kleijwegt K, Lupin-Jimenez L, Boyer M D, Eidietis N, Barbour N, Izacard O and Kolemen E 2020 *Phys. Plasmas* **27** 022501

[34] Piccione A, Berkery J W, Sabbagh S A and Andreopoulos Y 2022 *Nucl. Fusion* **62** 036002

[35] Rea C, Montes K J, Erickson K G, Granetz R S and Tinguely R A 2019 *Nucl. Fusion* **59** 096016

[36] Rea C and Granetz R S 2018 *Fusion Sci. Technol.* **74** 89

[37] Montes K J, Rea C, Granetz R S, Tinguely R A, Eidietis N, Meneghini O M, Chen D L, Shen B, Xiao B J, Erickson K and Boyer M D 2019 *Nucl. Fusion* **59** 096015

[38] Hu W, Rea C, Yuan Q P, Erickson K, Chen D, Shen B, Huang Y, Xiao J, Chen J, Duan Y, Zhang Y, Zhuang H, Xu J, Montes K J, Granetz R, Zeng L, Qian J, Xiao B and Li J 2021 *Nucl. Fusion* **61** 066034

[39] Mohapatra D, Subudhi B and Daniel R 2020 *Fusion Eng. Des.* **151** 111401

[40] Zhong Y, Zheng W, Chen Z Y, Xia F, Yu L M, Wu Q Q, Ai X K, Shen C S, Yang Z Y, Yan W, Ding Y H, Liang Y F, Chen Z P, Tong R H, Bai W, Fang J G and Li F 2021 *Plasma Phys. Control. Fusion* **63** 075008

[41] Rea C, Montes K J, Pau A, Granetz R S and Sauter O 2020 *Fusion Sci. Technol.* **76** 912

[42] Altman N and Krzywinski M 2017 *Nat. Methods* **14** 933

[43] Breiman L 2001 *Mach. Learn.* **45** 5

[44] Palczewska A, Palczewski J and Robinson R M 2014 *Integr. Reusable Syst.* 193

[45] Zheng W, Hu F R, Zhang M, Chen Z Y, Zhao X Q, Wang X L, Shi P, Zhang X L, Zhang X Q, Zhou Y N, Wei Y N and Pan Y 2018 *Nucl. Fusion* **58** 056016

[46] Guo B H, Shen B, Chen D L, Rea C, Granetz R S, Huang Y, Zeng L, Zhang H, Qian J P, Sun Y W and Xiao B J 2021 *Plasma Phys. Control. Fusion* **63** 025008

[47] Zhu J X, Rea C, Granetz R S, Marmar E S, Montes K J, Sweeney R, Tinguely R A, Chen D L, Shen B, Xiao B J, Humphreys D, Barr J and Meneghini O 2021 *Nucl. Fusion* **61** 114005

[48] Yang Z, Xia F, Song X, Gao Z, Huang Y and Wang S 2020 *Nucl. Fusion* **60** 016017

[49] Yang Z, Xia F, Song X, Gao Z, Wang S and Dong Y 2021 *Nucl. Fusion* **61** 126042