Multimodal Prediction of Tearing Instabilities in a Tokamak

Jaemin Seo[1,2,*], Rory Conlin[1], Andrew Rothstein[1], SangKyeun Kim[3], Joseph Abbate[1,3], Azarakhsh Jalalvand[1], and Egemen Kolemen[1,2,†]

[1]*Princeton University, Princeton, NJ, US*

[2]*Chung-Ang University, Seoul, South Korea*

[3]*Princeton Plasma Physics Laboratory, Princeton, NJ, US*

[*]jseo@cau.ac.kr, [†]ekolemen@pppl.gov

2023 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-8867-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IJCNN54540.2023.10191359

***Abstract*—Tokamak is a torus-shaped nuclear fusion device that uses magnetic fields to confine fusion fuel in the form of plasma. Tearing instability in plasma is a major issue in which the magnetic field breaks and recombines in tokamak. This instability can lead to plasma disruption that terminates the fusion power generation and damages the plasma-facing wall materials. For a successful steady operation of a large-scale tokamak without disruption, it is required to predict and alarm the tearing instabilities well in advance to avoid them. In this work, we develop and validate a deep neural network-based multimodal prediction system that estimates the future tearing instability likelihood from multi-diagnostics signals in the DIII-D tokamak.**

***Index Terms*—deep neural network, multimodal prediction, nuclear fusion, tokamak, tearing instability**

I. Introduction

A tokamak is one of the most promising concepts for a commercial nuclear fusion reactor, which confines the hydrogen plasmas with magnetic fields in a torus-shaped device. Recently, the Joint European Torus (JET) broke the world record by producing 59 MJ of fusion energy for five seconds [1], and the Korea Superconducting Tokamak Advanced Research (KSTAR) sustained 100 million Kelvin plasma for 30 seconds [2]. ITER, the international tokamak project with a collaboration of 35 nations, is also being constructed and will operate as of 2025 [3].

Although tokamaks have drawn successful achievements, there are still several obstacles we must resolve. While recent advances in tokamak control using deep reinforcement learning (RL) have been very promising [4]–[7], the main hurdle during the control to achieve a high-performance plasma has been plasma instability leading to plasma disruption [4], [8], [9]. Especially in ITER-scale fusion reactors, even a few events of plasma disruption can exert extreme damage on the components. Therefore, it is essential to develop techniques to reliably predict and avoid major disruptive instabilities such as magnetic field tearing. The physics research and experiments for mitigation of the existing tearing instability are advancing [10]–[13], as well as detection and classification of Alfvénic instabilities [14]–[16]. However, prediction of such instabilities to avoid them is still a major challenge.

Fu et al. [17] used decision tree-based machine learning (ML) algorithms to evaluate the likelihood of tearing events, so-called "tearability", in the DIII-D tokamak [18]. They successfully implemented the algorithm in the plasma control system (PCS) of DIII-D and maintained the tearability at a low level by adjusting the injected beam power based on the estimation of the tearability. However, their ML model only provides the tearability metric at a given time, not the dynamics of the future tearability when the actuators change. Such an approach is suitable for simple proportional control of a single actuator, as demonstrated by Fu et al., but not convenient for nonlinear control of multi-variate actuators, which is required to achieve even higher performance without exceeding the stability limit. Especially the tearing instabilities in the ITER-relevant plasma have complex nonlinear dynamics [19]. For a more active and flexible control with multi-variate actuators such as beam power, torque, and plasma shape to avoid the instabilities, we require a dynamic model that predicts the response of future plasma performance and tearability based on the proposed changes of the actuators.

In this paper we propose a deep neural network (DNN)-based system for predicting the dynamics of the future plasma performance and tearability from future actuators' conditions and the current plasma state measured by multiple diagnostics. Our prediction model differs from [17] in two aspects: (1) It uses multimodal inputs including spatial profile information that highly affects tearing stability, and (2) it predicts "future" tearing probabilities based on proposed actuator control commands rather than current tearing likelihood based on measurement signals. The overall diagram of the prediction system is described in Fig. 1. The rest of this paper is structured in four sections. Section II describes the characteristics of the experimental data and the architecture of the DNN model. In Section III, we show the comparison of the training results with different settings and the prediction results on several actual experiments in DIII-D. Finally, Section IV summarizes the paper with a conclusion and possible future research.

II. DNN-based Prediction of Tearing Onset

A. Data collection and preprocessing

Tearing instability is a phenomenon in which the magnetic field tears by finite plasma resistivity at rational surfaces of safety factor $q = m/n$, where m and n are the integer poloidal and toroidal mode numbers, respectively. A possible tearing

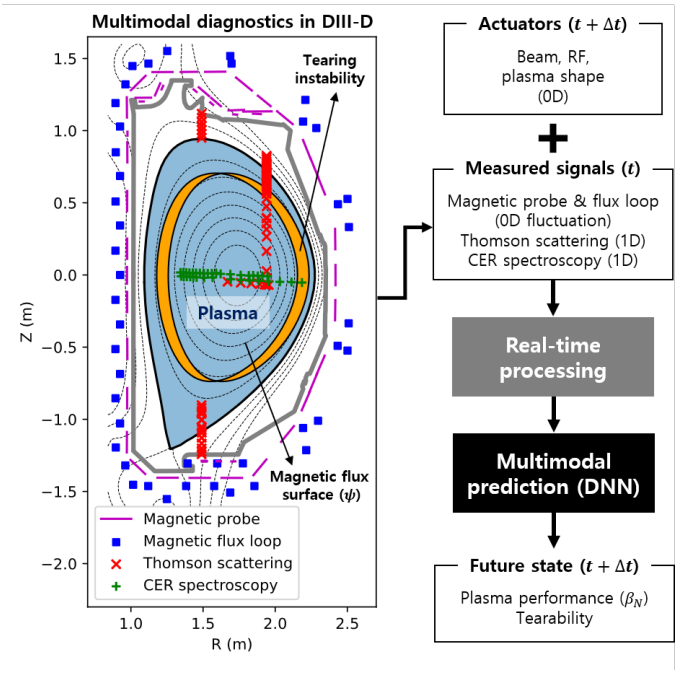

Fig. 1. The diagram of the proposed tearing instability prediction model and the selected diagnostics of DIII-D tokamak that are used as inputs to the model. A possible tearing instability of $m = 2$ and $n = 1$ is illustrated with an orange shade.

instability of $m = 2$ and $n = 1$ is also illustrated in Fig. 1, which is the most prone to induce plasma disruption. In many present tokamaks, the tearing instability is linearly stable but becomes unstable nonlinearly by a large enough seed magnetic perturbation. In high-pressure plasmas, which are favorable for a nuclear fusion reactor, the perturbation of the pressure-driven (so-called bootstrap) current becomes a seed that destabilizes the metastable state [20]. Therefore, the plasma pressure in present tokamaks is often limited by the tearing onset. This instability induces irreversible degradation of the plasma performance and often leads to plasma disruption [19], [21], hence, it is required to operate the tokamak below the tearing onset limit while pursuing high plasma pressure. As a first step for avoiding this instability, the dynamics of the tearing likelihood should be modeled.

In order to predict the future tearing likelihood (tearability) with DNN, the 1D plasma profiles should be considered as inputs of the model because the tearing stability strongly depends on the spatial information and the gradient of the kinetic and magnetic profiles near the rational surface [19]. The DIII-D PCS system is already equipped with a data-driven plasma profile predictor [22] and we used the same profiles of that model as the 1D input signals to our proposed DNN model as described in Table I. Electron density and temperature profiles can be estimated by the Thomson scattering measurements [23], and ion rotation profiles can be obtained by the charge exchange recombination (CER) spectroscopy [24]. The safety factor profile and the magneto-hydrodynamics equilibrium quantities can be calculated by RT-EFIT reconstruction based on magnetic measurements [25]. This information can be preprocessed in real-time during the tokamak discharge and provided as inputs for our model [26]. The controllable actuators, such as magnetic field strength, plasma current, beam power/torque, and RF injection, are also set as inputs. The other inputs, the plasma shape parameters, can be controlled by DIII-D PCS [25], [27]. These actuators and plasma shape parameters affect the plasma stability, so we can model the dynamics of the tearability depending on the "suggested" actuators' control.

Our ultimate goal is to push up the plasma pressure as much as possible without crossing a tearing onset limit.

TABLE I
INPUT AND OUTPUT SIGNALS FOR THE DNN MODEL

Name (inputs)	**Units**	**Type**	**Preprocessing**	**Time**
Electron density	$10^{19}\mathrm{m}^{-3}$	Thomson scattering	mtanh fitting	t
Electron temperature	keV	Thomson scattering	mtanh fitting	t
Ion rotation	kHz	CER spectroscopy	spline fitting	t
Safety factor	-	Magnetic measurement	RT-EFIT	t
Plasma pressure	Pa	Magnetic measurement	RT-EFIT	t
Magnetic field	T	Actuator	-	$t + \Delta t$
Plasma current	A	Actuator	-	$t + \Delta t$
Major radius	m	Plasma shape	RT-EFIT	$t + \Delta t$
Plasma elongation	-	Plasma shape	RT-EFIT	$t + \Delta t$
Top triangularity	-	Plasma shape	RT-EFIT	$t + \Delta t$
Bottom triangularity	-	Plasma shape	RT-EFIT	$t + \Delta t$
Inner gap	m	Plasma shape	RT-EFIT	$t + \Delta t$
Beam power	kW	Actuator	-	$t + \Delta t$
Beam torque	Nm	Actuator	-	$t + \Delta t$
RF power	W	Actuator	-	$t + \Delta t$
RF position	-	Actuator	-	$t + \Delta t$
Name (outputs)	**Units**	**Type**	**Preprocessing**	**Time**
Normalized pressure, β_N	-	Magnetic measurement	RT-EFIT	$t + \Delta t$
Tearability	-	Mirnov measurement	Fourier decomposition	$t + \Delta t$

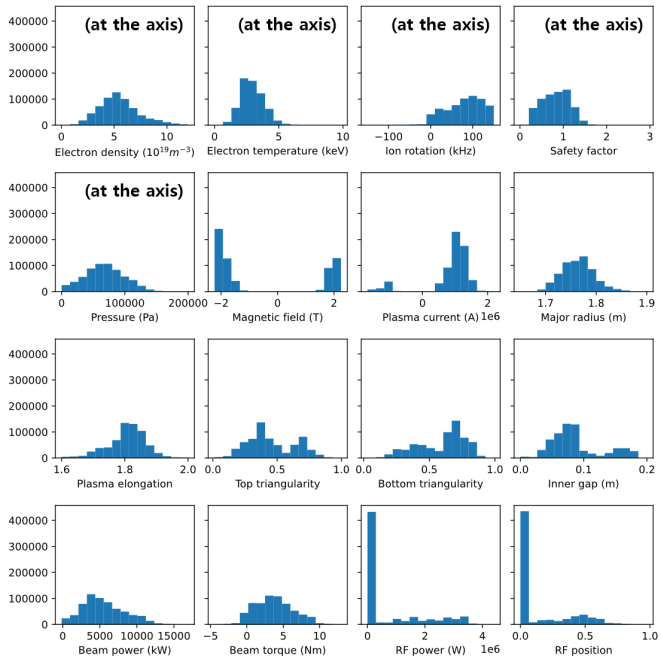

Fig. 2. The histogram of the input signals of the collected dataset. For the 1D signals, such as electron density, electron temperature, ion rotation, safety factor, and plasma pressure, only the axial values are counted.

Therefore, the prediction outputs of the DNN model are set as the normalized plasma pressure (β_N) and the tearability, as shown in Table I. The tearing instability label (tearability) is determined by the root-mean-square (RMS) amplitude of the $n = 1$ fluctuation signals, which are obtained by Fourier decomposition of the Mirnov coil signals [17]. At each time-step t, the model is supposed to predict β_N and tearability at $t + \Delta t$ by looking at the plasma profile signals at time t as well as "suggested" actuators values for $t + \Delta t$. Considering that the energy confinement time of the DIII-D plasmas is $O(25$ ms$)$ [22], setting $\Delta t = 25$ ms is a reasonable choice for capturing the stability dynamics based on the profile variation.

We used the MDS+ database [28] to collect the experimental data of DIII-D shots from the 2011 through 2022 campaigns (shots 147000 to 190985). Since real-time preprocessing of 1D profiles sometimes generates non-physical outlier data due to their loose constraint than the offline processing, we needed to exclude those outliers from the collected data. Also the safety factor (q) profile can diverge to infinity at the plasma boundary. Therefore, the inverse of the safety factor ($1/q$) has been used for the training data to reduce numerical difficulties [22]. The preprocessed dataset includes 8,505 shots containing 639,555 time-slices. The distribution of the collected data for the input and output signals are plotted in Figs. 2 and 3, respectively. Here, the distributions for the 1D signals (electron density, electron temperature, ion rotation, safety factor, and plasma pressure) are counted only for the values at the axis of plasma.

The normalized plasma pressure shows a continuous and well-distributed histogram, as shown in Fig. 3. However, the tearability is a binary variable, where 0 indicates stable against

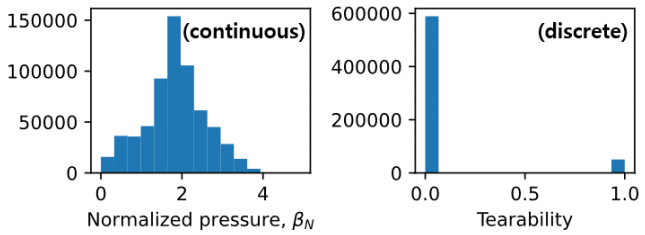

Fig. 3. The histogram of the output signals of the collected dataset. The normalized plasma pressure has a continuous distribution, but the tearability has a discrete one.

the tearing and 1 indicates unstable. In addition, the tearing instability often induces plasma disruption that terminates the tokamak discharge, and this property causes a strong imbalance of the experimental data of the tearability. The difficulties of (1) dealing with the continuous and the discrete outputs at the same time and (2) the highly imbalanced distribution of the tearability will be further discussed in Section III.

B. A DNN architecture for multimodal prediction

The raw signals from multiple diagnostic measurements have different dimensions from 0D to 2D with different spatial resolutions, as described in Fig. 1. Through real-time preprocessing, the dimensions of these signals are reduced into 0D or 1D, and the resolutions of the 1D signals are unified onto 33 equally spaced grids of the magnetic flux coordinate, ψ_N. The preprocessed input signals are then fed into the DNN model. First, the information of the 1D signals is extracted via a sequence of convolutional layers. Then, the extracted features from the 1D signals are concatenated with the 0D signals composed of the actuators and the plasma shape information. The concatenated features are fed to a fully connected multi-layer perceptron (MLP), which ends with the output layer to predict β_N and tearability. Here, all the activation functions of the hidden layers are set as the sigmoid function. The number of total parameters in the model is 12,086. The overall architecture of the DNN model is described in Fig. 4.

The 1D input signals are the measured values at time t, and the 0D inputs and the output signals are the values at time $t + \Delta t$, as shown in Table I. The 0D inputs, such as actuators and plasma shape, can be controlled by PCS in DIII-D. Therefore, by using the DNN model, we can predict the response of future plasma stability from possible actuator controls.

III. Experiments and Discussions

We use the Keras deep learning API [29] to build and train the DNN model. The final DNN model can be converted to C code by Keras2C [30] to be compatible with the control system in DIII-D. In order to reduce the possibility of overfitting during the training, the dropout technique is used before the last layer [31]. We also adopt the early stopping method during the training, which finishes the training process when the validation loss stops decreasing for ten epochs. The collected data samples are split into training, validation, and test sets at

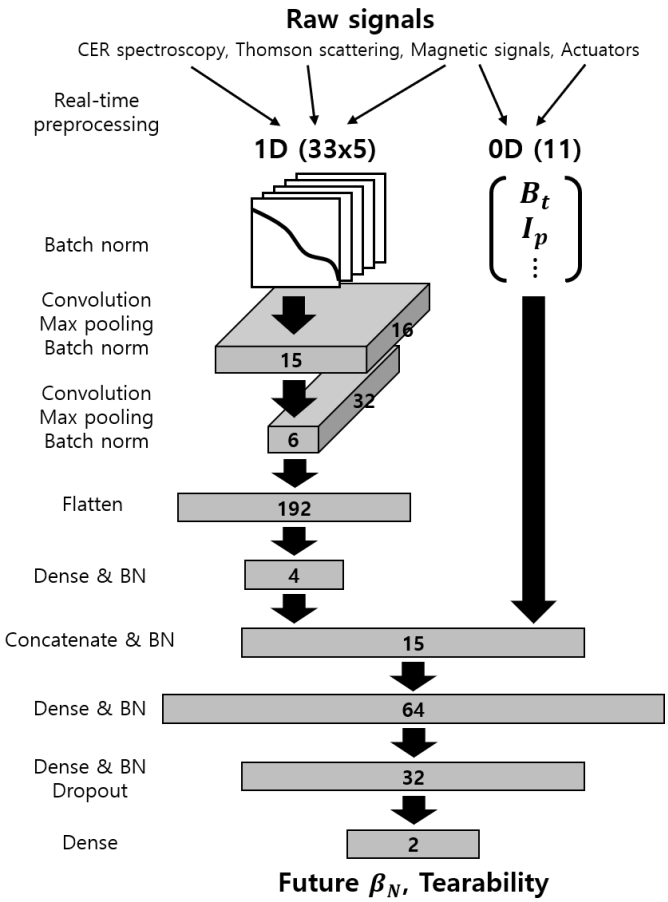

Fig. 4. The DNN model architecture to predict future plasma pressure and tearability from the multiple diagnostic signals. The output signals consist of the normalized plasma pressure (β_N) and the likelihood of tearing instability (tearability) after 25 ms.

a ratio of 7:2:1 to evaluate the overfitting. The batch size was determined as 512. For the training of the DNN model, we used the Adam optimizer [32] with a learning rate of 10^{-4}. A challenge in the training of our model is that the distribution of one output signal is discrete and highly imbalanced while another output has continuous and balanced distribution, as shown in Fig. 3. The mean-squared error (MSE) is suitable for the loss function for β_N which is a continuous output signal. On the other hand, either MSE or binary cross entropy (BCE) loss can be used for the tearability, the discrete output signal. In fact, the continuous regression when using MSE can also provide useful information about the possibility occurring of tearing instability. The highly imbalanced distribution of the tearability can be tackled by the weighted loss or the oversampling of the minority classes. In this work, we trained the models in four different ways depending on the type of loss function and the presence of oversampling, as shown in Table II.

The loss function L used for each case is defined in (1) and (2). Equation (1) is for cases 0 and 1 in Table II, where the losses of both outputs are calculated by MSE, and (2) is for

TABLE II
THE INFLUENCE OF THE TYPE OF LOSS FUNCTION AND OVERSAMPLING

Case number	Loss for β_N	Loss for tearability	Over-sampling	R^2 for β_N	AUC for tearability
0	MSE	MSE	No	0.975	0.875
1	MSE	MSE	Yes	0.957	0.903
2	MSE	BCE	No	0.971	0.887
3	MSE	BCE	Yes	0.957	0.907

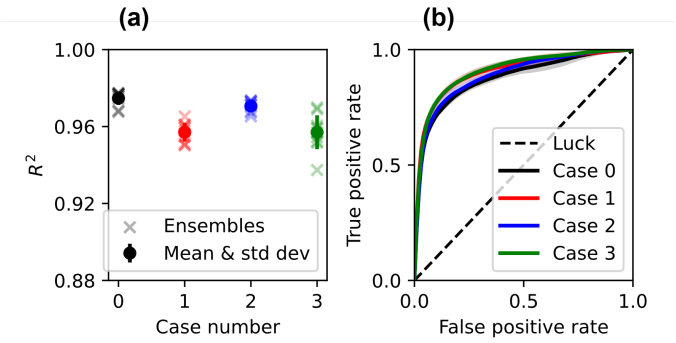

Fig. 5. The distribution of the accuracies of the output prediction for each case. (a) The accuracy for β_N is estimated with the coefficient of determination, R^2, and (b) the accuracy for the tearability is estimated by the area under the ROC curve.

cases 2 and 3, where the loss for the tearability prediction is determined by BCE. Here, N is the batch size, i is the sample index, y_1 is the true value of β_N, y_2 is the true value of the tearability, and $\hat{y}$ is the predicted values of the output signals. Since the BCE loss tends to be larger than the MSE loss for the same errors, we empirically multiplied the BCE loss by a weight ($w_{BCE} = 1/3$) in (2).

$$L_{0or1} = \frac{1}{N}\sum_{i=1}^{N}[(y_{1,i} - \hat{y}_{1,i})^2 + (y_{2,i} - \hat{y}_{2,i})^2] \tag{1}$$

$$\begin{aligned} L_{2or3} = \frac{1}{N}\sum_{i=1}^{N}[&(y_{1,i} - \hat{y}_{1,i})^2 \\ &- w_{BCE}(y_{2,i}\log\hat{y}_{2,i} + (1 - y_{2,i})\log\hat{y}_{2,i})] \end{aligned} \tag{2}$$

For statistically reliable comparison, we trained ten identical-structure ensemble models using cross-validation in each case. After training, the accuracies of β_N and tearability are evaluated with the coefficient of determination (R^2) and the area under the ROC curve (AUC), respectively. Here, the ROC (receiver operating characteristic) curve is a graphical representation of the trade-off between the true positive rate and the false positive rate of a binary classifier. The ensemble-averaged accuracies of the outputs are shown in Table II. The distribution of R^2 for the β_N prediction and the ROC curve for the tearability prediction can be seen in Fig. 5.

In cases 1 and 3, we oversampled the minority classes, the tearing-unstable cases, so that they have the same number of samples as the majority class during the training. As the tearing-unstable cases are included more in the batch, the DNN model becomes better for classifying the tearing cases. However, the plasma with tearing instabilities tends to fluctuate and often be disrupted, which causes uncertainty and noise in the training data. Although this noise can potentially induce the deterioration of the β_N prediction accuracy (as shown in Fig. 5 (a)), the accuracy of the β_N prediction is still high enough ($R^2 \approx 0.957$) despite the oversampling and is sufficient to predict the dynamics of plasma performance. Since the main goal of this work is to predict and alarm the tearing instability, the DNN model trained with oversampling will be used in the later discussion.

While the BCE loss is commonly used in binary class problems such as predicting tearing instability, the results using the MSE loss and the BCE loss do not show a significant difference in prediction accuracy in Table II and Fig. 5. Further analysis of the difference between using MSE and BCE for tearing instability will be conducted in the future. In the rest of this section, we use the DNN model trained with case 3 which yields the highest AUC value for the tearability prediction.

After training the DNN model, we conducted DIII-D experiments to test the feasibility of alarming the instability in real experiments using the trained model. The future ITER baseline scenario (IBS) characterizes the low edge safety factor ($q_{95} \approx 3$) and low toroidal rotation, which make the plasma prone to disruption by tearing instability [19]. Therefore, it is important to be capable of predicting the tearing likelihood of the IBS plasmas. For this reason, the IBS demonstration discharges in DIII-D with stable and unstable plasmas are targeted for this test. Figs. 6, 7, and 8 show the prediction results for different IBS discharges in DIII-D, which are all unseen shots by the DNN model during the training. In these discharges, we maintain the edge safety factor as $q_{95} = 3.2$ and the beam torque below 1 Nm to constrain the IBS condition. Shot 193207 (Fig. 6) and 193208 (Fig. 7) are operated under the almost identical preprogrammed setting, but the tearing instability occurred in the former and not in the latter.

In Fig. 6, the top graph shows the time traces of the plasma current and injected beam power, which are key input features of the actuators. Other actuators are not significantly varying in this discharge. The next four graphs present the 1D input profiles, namely, electron density, electron temperature, safety factor, and ion rotation. In each graph, only four values (out of 33) at $\psi_N = 0.0, 0.3, 0.6$, and 0.9 are shown for the sake of visibility. The last two graphs show the normalized plasma pressure and the tearability, which are the outputs of the DNN model. The ground truth values are shown in black dashed lines, and the predicted ones are in blue solid lines. The uncertainty ranges from ten ensemble models are indicated with blue-filled areas.

In Fig. 6, the tearing instability occurs at $t = 3250$ ms, the rotating instability gets locked at $t = 3400$ ms, and finally, the plasma is disrupted at $t = 3800$ ms. This sequence is a

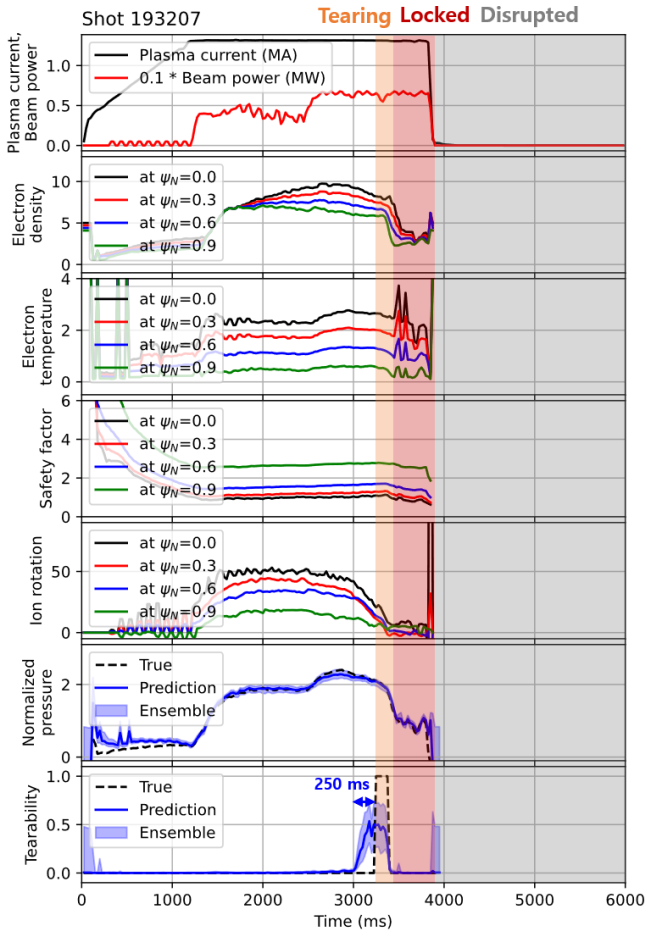

Fig. 6. Tearing prediction result for shot 193207 in DIII-D. The tearing instability occurs at $t = 3250$ ms, it locks to the wall at $t = 3400$ ms, and the plasma is disrupted at $t = 3800$ ms.

typical process when the disruptive tearing instability occurs in a tokamak. The tearability plot shows that the DNN model successfully predicts the tearing instability long enough before the event happens, so that we can respond or avoid it in advance. In shot 193207, the model could predict a positive value of the tearability 250 ms before the instability occurred. This predicted tearability could be interpreted as a likelihood of the occurrence of tearing events. 250 ms is several times the typical value of energy confinement time ($O(25$ ms$)$) of the DIII-D plasmas and is sufficient time to avoid instability by changing the kinetic profiles using external actuators such as beam, RF, and plasma shape.

Even though shot 193208 has been operated under almost the same condition as shot 193207, the plasma of shot 193208 is stable until the end of the discharge, as shown in Fig. 7. The DNN model also estimates almost zero tearability throughout the discharge. The only difference of shot 193207 from shot 193208 at $t = 3000$ ms is that the toroidal rotation starts to slow down. In IBS plasmas, the rotation drop opens the stabilizing ion-polarization current gate, which can induce tearing instability [21]. The lower rotation also deepens the

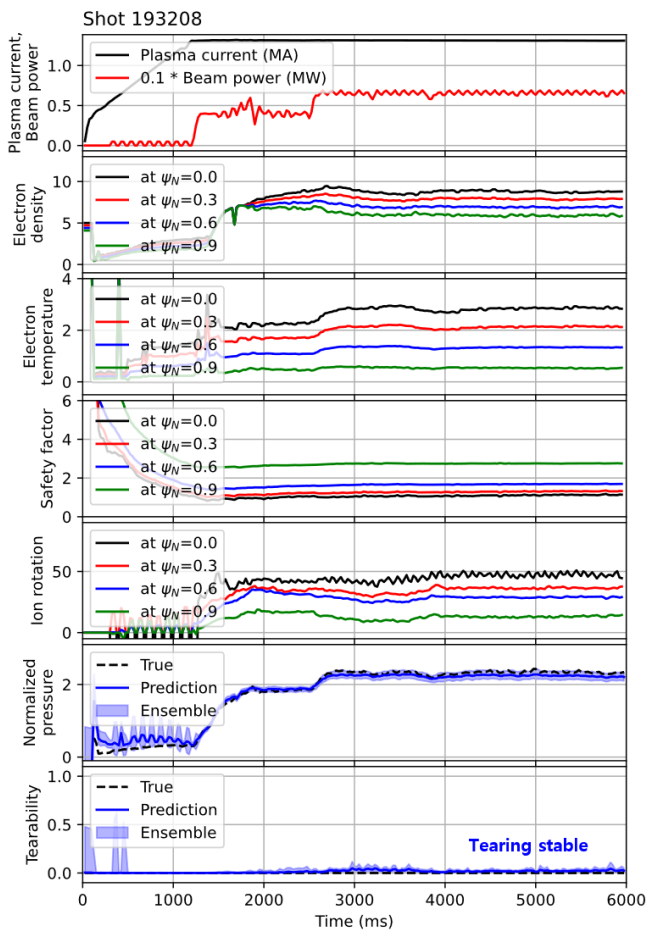

Fig. 7. Tearing prediction result for shot 193208 in DIII-D. The plasma is stable against tearing in this shot.

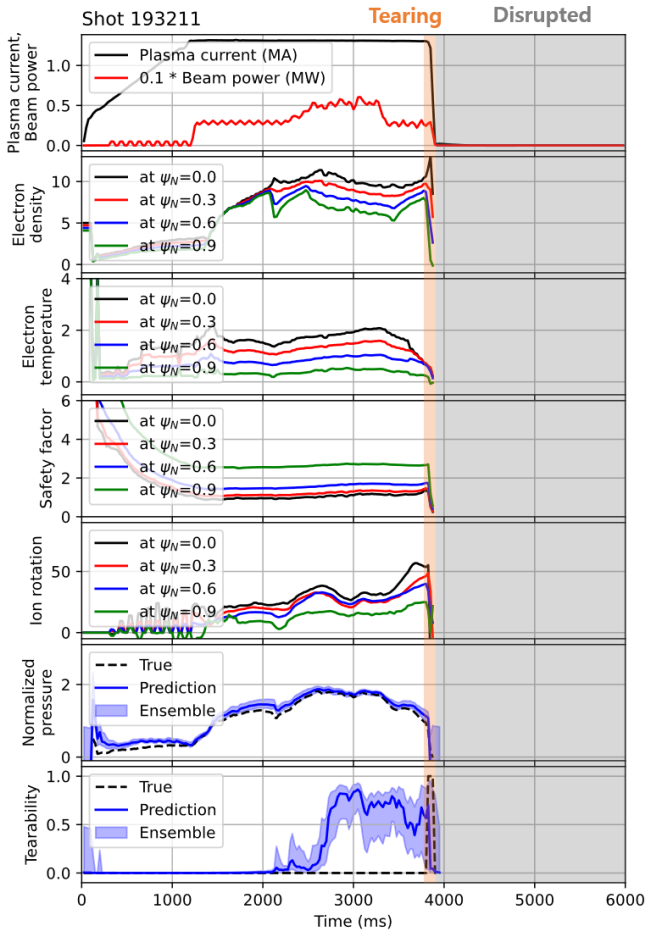

Fig. 8. Tearing prediction result for shot 193211 in DIII-D. The model predicts high tearability from $t \approx 2500$ ms, but the plasma is sustained longer than 1000 ms until the tearing event eventually occurs at $t \approx 3800$ ms

well of the plasma current profile at the rational surface, which is also correlated with instability [19]. The data-driven DNN model could cover the complicated physics of the interaction among the rotation, current profile, and tearing instability. The prediction of tearability only requires a single forward propagation of the DNN model, which takes less than 1ms per inference. This suggests that the DNN model is suitable for evaluating the tearability of the DIII-D plasmas in real-time.

An interesting feature of the tearability prediction can be observed in Fig. 8. Even though the predicted tearability is close to 1 from $t \approx 2500$ ms, the plasma is sustained longer than 1000 ms without the instability before the tearing event eventually happens at $t \approx 3800$ ms. This indicates that high tearability does not always lead to a tearing event. This is because, as illustrated in Fig. 9, the tearing instability requires not only high tearability but also a seed perturbation to grow. The tearing event is a metastable phenomenon that is linearly stable but nonlinearly unstable. High tearability only means that the gate is opened for tearing events, but the actual occurrence of the events requires another factor, a seed perturbation.

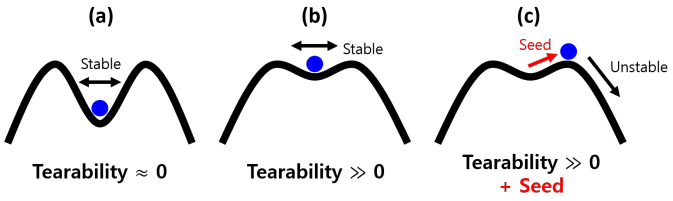

Fig. 9. The illustration of the metastable state of tearing instability that requires a seed perturbation to become unstable. High tearability makes the plasma marginally stable, and a seed perturbation excites the instability.

Therefore, even the plasma with high tearability can persist stably if there is no large enough seed magnetic perturbation. This physical property of tearing instability makes it difficult to increase the AUC value beyond a certain level, as shown in Table II and Fig. 5.

The seed can be a magnetic perturbation by sawtooth crash, edge localized mode, higher mode number instabilities, or external factors such as 3D field coils [21], [33]. Because these phenomena are irregular and evolve in a time scale much shorter than the prediction time interval, the actual onset of

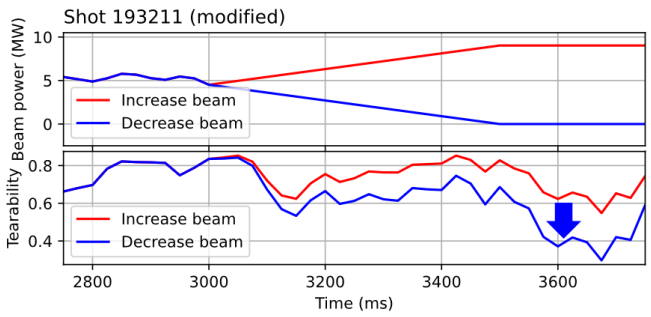

Fig. 10. The prediction of the dynamics of tearability when the injected beam power changes from the original shot 193211. Adjusting beam power can lower the tearability.

tearing instability is nearly stochastic even after the tearability increases. However, this property provides a time opportunity to avoid instabilities before the seed perturbation occurs after the model alarms. Fig. 10 shows the prediction of the dynamics of tearability when the beam power actuator changes from the original shot 193211. As the injected beam power decreases, the tearability is also reduced compared to the other case, which means the plasma moves away from the tearing onset limit. This implies a possibility of avoiding tearing instabilities by adjusting the tokamak actuators.

Fig. 11 shows the time trace of shot 193210, where a primitive beam power control is tried to avoid the tearing instability in the experiment. As the predicted tearability starts to increase from $t = 3000$ ms, the beam power is reduced from 7 MW to 3 MW. Then, the tearability decreases down to zero again, and the plasma could sustain more than 1000 ms after that. The observation that the increased tearability can be restored to zero by adjusting an actuator implies the feasibility of the instability avoidance control while pursuing high fusion performance. Although the tearability increased again at $t \approx 4000$ ms and the tearing event occurred eventually, we expect a suitable control could avoid this instability event as well. The tearability dynamic model using DNN in this work can be an environment to train the tearing avoidance control model by using deep reinforcement learning (RL) in the future. The trained RL agent will be able to actively adjust multiple actuators to avoid tearing instability while keeping high plasma performance.

IV. Conclusion and Future Work

In this work, we propose a multimodal prediction system based on DNN that predicts the future dynamics of the plasma pressure and the likelihood of tearing instability. The multimodal signals obtained by measurements are processed in real-time into 0D and 1D inputs, and the DNN model estimates the normalized plasma pressure and the tearability from the inputs. We tested this DNN-based tearing instability alarming system with ITER-relevant experiments in DIII-D, and it shows a reasonable prediction of the instability a few hundred milliseconds ahead of the event. We also

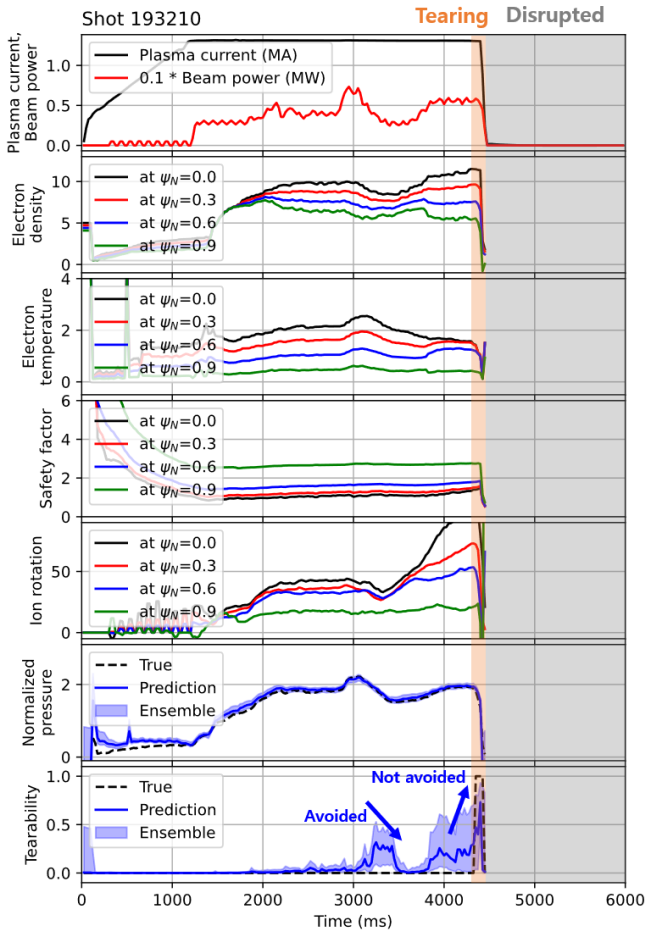

Fig. 11. The result of tearing prediction and primitive avoidance control for shot 193210 in DIII-D.

demonstrated the feasibility of avoiding instability using this prediction system with a primitive beam power control. Adjusting the beam power upon the tearing instability alarm, reduced the tearability again. This indicates that a more active avoidance control using multiple actuators is also possible in the future. By using the actuator control based on the tearing instability prediction, we can achieve long sustainment of high-performance plasmas just below the instability onset limit. This tearing prediction system can also be combined with the profile prediction technique installed in DIII-D [22], which enables us to predict the dynamics of plasma stability of a farther future by autoregressive prediction. Then, we can optimize the whole trajectory of a tokamak discharge which yields the highest performance without instability. Recently, it has been demonstrated that deep reinforcement learning can be used for the tokamak control and optimization [4]–[7]. We can train the RL agent in the environment of the tearing instability dynamic model, to obtain a robust tokamak controller to achieve higher performance with a more stable condition. Especially, the main hurdle of controlling fusion plasma in previous trials is the tearing instability and disruption [4], [7]. By resolving this issue using instability prediction and

avoidance, a basic technology for autonomous fusion reactor control can be established.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Award DE-FC02-04ER54698. This work was also supported by the National Research Foundation of Korea(NRF) funded by the Korea government. (Ministry of Science and ICT) (RS-2023-00255492)

DISCLAIMER

REFERENCES

[1] J. Mailloux, N. Abid, K. Abraham *et al.*, "Overview of jet results for optimising iter operation," *Nuclear Fusion*, vol. 62, no. 4, p. 042026, jun 2022.

[2] H. Han, S. J. Park, C. Sung *et al.*, "A sustained high-temperature fusion plasma regime facilitated by fast ions," *Nature*, vol. 609, no. 7926, pp. 269–275, Sep 2022.

[3] M. Shimada, D. Campbell, V. Mukhovatov *et al.*, "Chapter 1: Overview and summary," *Nuclear Fusion*, vol. 47, no. 6, p. S1, jun 2007.

[4] J. Seo, Y.-S. Na, B. Kim *et al.*, "Feedforward beta control in the kstar tokamak by deep reinforcement learning," *Nuclear Fusion*, vol. 61, no. 10, p. 106010, sep 2021.

[5] J. Degrave, F. Felici, J. Buchli *et al.*, "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, no. 7897, pp. 414–419, Feb 2022.

[6] J. Seo, Y.-S. Na, B. Kim *et al.*, "Development of an operation trajectory design algorithm for control of multiple 0d parameters using deep reinforcement learning in kstar," *Nuclear Fusion*, vol. 62, no. 8, p. 086049, jul 2022.

[7] I. Char, J. Abbate, L. Bardoczi *et al.*, "Offline model-based reinforcement learning for tokamak control," in *Advances in neural information processing systems*, vol. 36, December 2022. [Online]. Available: https://neurips.cc/Conferences/2022/ScheduleMultitrack?event=56844

[8] I. Char, Y. Chung, W. Neiswanger *et al.*, "Offline contextual bayesian optimization," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer *et al.*, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://papers.nips.cc/paper/2019/file/7876acb66640bad41f1e1371ef30c180-Paper.pdf

[9] Y. Chung, I. Char, W. Neiswanger *et al.*, "Offline contextual bayesian optimization for nuclear fusion," 2020. [Online]. Available: https://arxiv.org/abs/2001.01793

[10] R. J. La Haye, S. Günter, D. A. Humphreys *et al.*, "Control of neoclassical tearing modes in diii–d," *Physics of Plasmas*, vol. 9, no. 5, pp. 2051–2060, 2002.

[11] F. A. G. Volpe, M. E. Austin, R. J. La Haye *et al.*, "Advanced techniques for neoclassical tearing mode control in diii-d," *Physics of Plasmas*, vol. 16, no. 10, p. 102502, 2009.

[12] E. Kolemen, A. Welander, R. L. Haye *et al.*, "State-of-the-art neoclassical tearing mode control in diii-d using real-time steerable electron cyclotron current drive launchers," *Nuclear Fusion*, vol. 54, no. 7, p. 073020, may 2014.

[13] M. Park, Y.-S. Na, J. Seo *et al.*, "Effect of electron cyclotron beam width to neoclassical tearing mode stabilization by minimum seeking control in iter," *Nuclear Fusion*, vol. 58, no. 1, p. 016042, 2017.

[14] J. Seo, J. Kim, J. Mailloux *et al.*, "Parametric study of linear stability of toroidal alfvén eigenmode in jet and kstar," *Nuclear Fusion*, vol. 60, no. 6, p. 066008, 2020.

[15] A. Jalalvand, A. A. Kaptanoglu, A. V. Garcia *et al.*, "Alfvén eigenmode classification based on ece diagnostics at diii-d using deep recurrent neural networks," *Nuclear Fusion*, vol. 62, no. 2, p. 026007, 2021.

[16] A. A. Kaptanoglu, A. Jalalvand, A. V. Garcia *et al.*, "Exploring data-driven models for spatiotemporally local classification of alfvén eigenmodes," *Nuclear Fusion*, vol. 62, no. 10, p. 106014, 2022.

[17] Y. Fu, D. Eldon, K. Erickson *et al.*, "Machine learning control for disruption and tearing mode avoidance," *Physics of Plasmas*, vol. 27, no. 2, p. 022501, 2020.

[18] M. E. Fenstermacher, for the DIII-D Team:, J. Abbate *et al.*, "Diii-d research advancing the physics basis for optimizing the tokamak approach to fusion energy," *Nuclear Fusion*, vol. 62, no. 4, p. 042024, apr 2022.

[19] F. Turco, T. Luce, W. Solomon *et al.*, "The causes of the disruptive tearing instabilities of the iter baseline scenario in diii-d," *Nuclear Fusion*, vol. 58, no. 10, p. 106043, 2018.

[20] R. J. La Haye, "Neoclassical tearing modes and their control," *Physics of Plasmas*, vol. 13, no. 5, p. 055501, 2006.

[21] R. J. La Haye, C. Chrystal, E. J. Strait *et al.*, "Disruptive neoclassical tearing mode seeding in diii-d with implications for iter," *Nuclear Fusion*, vol. 62, no. 5, p. 056017, 2022.

[22] J. Abbate, R. Conlin, and E. Kolemen, "Data-driven profile prediction for diii-d," *Nuclear Fusion*, vol. 61, no. 4, p. 046027, 2021.

[23] T. N. Carlstrom, G. L. Campbell, J. C. DeBoo *et al.*, "Design and operation of the multipulse thomson scattering diagnostic on diii-d (invited)," *Review of Scientific Instruments*, vol. 63, no. 10, pp. 4901–4906, 1992.

[24] R. P. Seraydarian and K. H. Burrell, "Multichordal charge-exchange recombination spectroscopy on the diii-d tokamak," *Review of Scientific Instruments*, vol. 57, no. 8, pp. 2012–2014, 1986.

[25] J. Ferron, M. Walker, L. Lao *et al.*, "Real time equilibrium reconstruction for tokamak discharge control," *Nuclear Fusion*, vol. 38, no. 7, p. 1055, jul 1998.

[26] R. Shousha, J. Ferron, Z. Xing *et al.*, "Improved real-time equilibrium reconstruction with kinetic constraints on diii-d and nstx-u," in *Bulletin of the American Physical Society*, vol. 64. APS, October 2022.

[27] J. Barr, B. Sammuli, D. Humphreys *et al.*, "Development and experimental qualification of novel disruption prevention techniques on diii-d," *Nuclear Fusion*, vol. 61, no. 12, p. 126019, oct 2021.

[28] J. A. Stillerman, T. W. Fredian, K. Klare *et al.*, "Mdsplus data acquisition system," *Review of Scientific Instruments*, vol. 68, no. 1, pp. 939–942, 1997.

[29] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[30] R. Conlin, K. Erickson, J. Abbate *et al.*, "Keras2c: A library for converting keras neural networks to real-time compatible c," *Engineering Applications of Artificial Intelligence*, vol. 100, p. 104182, 2021.

[31] N. Srivastava, G. Hinton, A. Krizhevsky *et al.*, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[33] L. Bardóczi, N. C. Logan, and E. J. Strait, "Neoclassical tearing mode seeding by nonlinear three-wave interactions in tokamaks," *Phys. Rev. Lett.*, vol. 127, p. 055002, Jul 2021.