Multimodal Prediction of Tearing Instabilities

in a Tokamak

Jaemin Seo  
Princeton University*,* Princeton, NJ, US  
[js7146@princeton.edu](mailto:js7146@princeton.edu)

Azarakhsh Jalalvand,  
Princeton University*,* Princeton, NJ, US  
[aj17@princeton.edu](mailto:aj17@princeton.edu)Rory Conlin  
Princeton University*,* Princeton, NJ, US  
[wconlin@princeton.edu](mailto:wconlin@princeton.edu)

Joseph Abbate,  
Princeton University*,* Princeton, NJ, US  
[jabbate@princeton.edu](mailto:jabbate@princeton.edu)Andrew Rothstein   
Princeton University*,* Princeton, NJ, US  
[arothstein@princeton.edu](mailto:arothstein@princeton.edu)

Egemen Kolemen   
Princeton University*,* Princeton, NJ, US  
[ekolemen@princeton.edu](mailto:ekolemen@princeton.edu) SangKyeun Kim   
Princeton Plasma Physics Laboratory,  
Princeton, NJ, US  
[skim2@pppl.gov](mailto:skim2@pppl.gov)

*Abstract*—Tearing instability is a major issue in which the magnetic field breaks and recombines in a tokamak, a torus-shaped nuclear fusion device. 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 before the events so that we can 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.

Keywords—deep neural network, multimodal prediction, nuclear fusion, tokamak, plasma instability, tearing instability

# 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 since 2027 [3].

Although tokamaks have drawn successful achievements, there are still several obstacles we must resolve. While advanced tokamak controls using deep reinforcement learning (RL) have been conducted recently [4]–[7], the main hurdle during RL control to achieve a high-performance plasma has been plasma instability and the resulting 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 predict and avoid major disruptive instabilities, such as magnetic field tearing. The researches of numerical modeling and experiment regarding the evolution of the tearing instability and its suppression have been actively conducted [10], [11], and recently, detection and classification of Alfvénic instabilities have also been demonstrated [12]–[14]. However, the techniques of predicting tearing instabilities in advance to avoid them have not been studied much.

Fu et al. [15] used decision tree-based machine learning (ML) algorithms to evaluate a metric, “tearability,” the likelihood of the tearing events, in the DIII-D tokamak. They successfully maintained the tearability at a low level by adjusting the injected beam power according to 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. This way is enough for simple proportional control of a single actuator, as demonstrated by Fu et al., but not suitable 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 more complex nonlinear dynamics than other plasma operation scenarios [16]. 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 from the future change of the actuators.

This paper aims to describe a deep neural network (DNN)-based system for predicting the dynamics of the future plasma performance and tearability from future actuators’ conditions and given plasma state measured by multiple diagnostics. 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.

# DNN-based Prediction of Tearing Onset

## Data collection and preprocessing

Tearing instability is a phenomenon that the magnetic field tears by finite plasma resistivity at rational surfaces of safety factor , where and are the integer poloidal and toroidal mode numbers, respectively. A possible tearing instability of and 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. 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 [16], [17], so 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 since the tearing stability strongly depends on the spatial information and the gradient of the kinetic and magnetic profiles near the rational surface [16]. For possible future combination with the profile prediction algorithm equipped in DIII-D [18], the 1D signals that algorithm uses were set as inputs of our DNN model, which are described in Table I in detail. Electron density and temperature profiles can be estimated by the Thomson scattering measurements [19], and ion rotation profiles can be obtained by the charge exchange recombination (CER) spectroscopy [20]. The safety factor profile and the magneto-hydrodynamics equilibrium quantities can be calculated by RT-EFIT reconstruction based on magnetic measurements [21]. This information can be preprocessed in real-time during the tokamak discharge and provided as inputs for our model [22].

Our ultimate goal is to push up the plasma pressure as long as it does not touch the tearing onset limit. Therefore, the prediction outputs of the DNN model are set as the normalized plasma pressure () and the tearability, as shown in Table II. The tearing instability label (tearability) is determined by the root-mean-square amplitude of the n=1 fluctuation signals, which are obtained by Fourier decomposition of the Mirnov coil signals. The output variables and the actuators in the input variables are selected as values later than the time of the plasma profile signals (). This way allows the model to predict the future dynamics according to actuator changes from the given current plasma state. Considering that the energy confinement time of the DIII-D plasmas is [18], this time interval is suitable for capturing the stability dynamics according to the profile variation.

The experimental data of DIII-D shots from the 2011 through 2022 campaigns (shots 147000 to 190985) are collected from the MDS+ database [23]. Since real-time preprocessing sometimes generates unphysical outlier data due to their loose constraint than the offline processing, we needed to exclude those outliers from the collected data. The safety factor () profile can diverge to infinity at the plasma boundary, so the inverse of the safety factor () has been used for the training data to reduce numerical difficulties [18]. The preprocessed dataset includes 8,505 shots containing 639,555 samples. The distribution of the collected data for the input and output signals are plotted in Fig. 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.

The normalized plasma pressure shows a continuous and well-distributed histogram, as shown in Fig. 3. However, the tearability is a discontinuous variable, where 0 indicates stable against the tearing and 1 indicates unstable. In addition, the tearing instability often induces plasma disruption that terminates the tokamak discharge, and this property induces a strong imbalance of the experimental data of the tearability. The difficulties of (i) dealing with the continuous and the discontinuous outputs at the same time and (ii) the highly imbalanced distribution of the tearability will be discussed again in Section III.

## 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, . 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 through another sequence of the fully connected layers, which are finally connected with the output layer composed of the signals of the future 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 values at time , and the 0D inputs and the output signals are the values at time , as shown in Table I and II. The 0D inputs, such as actuators and plasma shape, can be controlled by the plasma control system equipped in DIII-D. Therefore, by using the DNN model, we can predict the response of future plasma stability from possible actuator controls.

# Experiments and Discussions

We use the Keras deep learning API [24] to build and train the DNN model, for conversion to the real-time capable C code [25] 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 [26]. 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 a ratio of 7:2:1 to evaluate the overfitting. The batch size was determined as 512. For the training of the DNN model, the Adam optimizer [27] is used with a learning rate of .

One thing to be careful of while training the 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 which is a continuous output signal, however, either MSE or binary cross entropy (BCE) loss can be used for the tearability, the discrete output signal. Even though the ground truth values of the tearability are discrete, which is a typical case of BCE being used, the continuous regression using MSE will also provide useful information regarding the possibility of tearing instability to occur. The highly imbalanced distribution of the tearability can be dealt with by the weighted loss or the oversampling of the minority classes. In this work, we trained the models in 4 different ways depending on the type of loss function and the presence of oversampling, as shown in Table III.

The loss function used for each case is defined in (1) and (2). Equation (1) is for cases 0 and 1, where both losses of the outputs are set as MSE, and (2) is for cases 2 and 3, where the loss for the tearability prediction is set as BCE. Here, is the batch size, is the sample index, is the true value of , is the true value of the tearability, and 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 additionally multiplied the BCE loss by a weight () in (2).

(1)

(2)

For statistically reliable comparison, we trained ten identical-structure ensemble models using cross-validation in each case. After training, the accuracies for and tearability are evaluated with the coefficient of determination () and the AUC value, respectively. The ensemble-averaged accuracies of the outputs are shown in Table III. The distribution of for the 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 the uncertainty and noise of the training data. This noise in the data induces the deterioration of the prediction accuracy, as shown in Fig. 5 (a). This can be a trade-off of using the oversampling method, but the accuracy of the prediction is still high () 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.

Although the BCE loss is commonly used in binary class problems like predicting tearing instability, the results using the MSE loss and the BCE loss don’t show a significant difference in prediction accuracy in Table III and Fig. 5. 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.

For a more practical test of the trained model, we conducted the prediction of the time-evolution of the plasma performance and the tearability of actual plasma discharges in DIII-D, as shown in Fig. 6, 7, and 8. The future ITER baseline scenario (IBS) characterizes the low edge safety factor () and low toroidal rotation, which make the plasma prone to disruption by tearing instability [16], so 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. Fig. 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 and the beam torque below 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 are about the 1D input profiles of electron density, electron temperature, safety factor, and ion rotation. In each graph, only four values at 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 , the rotating instability becomes locked at , and finally, the plasma is disrupted at . This sequence is a typical process when the disruptive tearing instability occurs in a tokamak. Fortunately, it is seen that the DNN model trained with previous experimental data successfully predicts the tearing instability 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 before the instability occurred. is several times the energy confinement time 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 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 [17]. The lower rotation also deepens the well of the plasma current profile at the rational surface, which can either destabilize the plasma [16]. The data-driven DNN model could cover the complicated physics of the interaction among the rotation, current profile, and tearing instability. Since the prediction of tearability only requires a single forward propagation of the DNN model, which takes less than per each inference after the C conversion [25], it can be a useful tool to evaluate the tearability of the DIII-D plasmas in real-time.

An interesting feature of the tearability prediction can be found in Fig. 8. Even though the predicted tearability is close to 1 from , the plasma is sustained longer than without the instability before the tearing event eventually happens at . 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. Since the tearing event is a metastable phenomenon that is linearly stable but nonlinearly unstable, even the plasma with high tearability can persist stably if there is no large enough seed magnetic perturbation. 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 [17], [28]. Because these phenomena are irregular and evolve in a time scale much shorter than the prediction time interval, the actual onset of tearing instability is nearly stochastic even after the tearability becomes high. 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 also becomes 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 , 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 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 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.

# 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 real-time preprocessed 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 demonstrated the feasibility of avoiding instability using this prediction system with a primitive beam power control. As the beam power is adjusted after the tearing instability alarmed, the tearability becomes reduced 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 [18], 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.

##### References

[1] J. Mailloux *et al.*, “Overview of JET results for optimising ITER operation,” *Nucl. Fusion*, vol. 62, no. 4, p. 042026, Jun. 2022, doi: 10.1088/1741-4326/ac47b4.

[2] H. Han *et al.*, “A sustained high-temperature fusion plasma regime facilitated by fast ions,” *Nature*, vol. 609, no. 7926, pp. 269–275, Sep. 2022, doi: 10.1038/s41586-022-05008-1.

[3] M. Shimada *et al.*, “Progress in the ITER Physics Basis - Chapter 1: Overview and summary,” *Nucl. Fusion*, vol. 47, pp. S1–S17, 2007.

[4] J. Seo *et al.*, “Feedforward beta control in the KSTAR tokamak by deep reinforcement learning,” *Nucl. Fusion*, vol. 61, no. 10, p. 106010, Oct. 2021, doi: 10.1088/1741-4326/AC121B.

[5] J. Seo *et al.*, “Development of an operation trajectory design algorithm for control of multiple 0D parameters using deep reinforcement learning in KSTAR,” *Nucl. Fusion*, vol. 62, no. 8, p. 086049, Jul. 2022, doi: 10.1088/1741-4326/AC79BE.

[6] J. Degrave *et al.*, “Magnetic control of tokamak plasmas through deep reinforcement learning,” *Nature*, vol. 602, no. 7897, pp. 414–419, Feb. 2022, doi: 10.1038/s41586-021-04301-9.

[7] I. Char, Y. Chung, M. Boyer, E. Kolemen, and J. Schneider, “A Model-Based Reinforcement Learning Approach for Beta Control,” Nov. 2021.

[8] I. Char *et al.*, “Offline Contextual Bayesian Optimization,” *Adv. Neural Inf. Process. Syst.*, vol. 32, p. 4629, 2019, Accessed: Jan. 06, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/7876acb66640bad41f1e1371ef30c180-Abstract.html.

[9] Y. Chung *et al.*, “Offline Contextual Bayesian Optimization for Nuclear Fusion,” Jan. 2020, doi: 10.48550/arxiv.2001.01793.

[10] R. J. La Haye *et al.*, “Control of neoclassical tearing modes in DIII–D,” *Phys. Plasmas*, vol. 9, no. 5, p. 2051, Apr. 2002, doi: 10.1063/1.1456066.

[11] M. Park, Y. S. Na, J. Seo, M. Kim, and K. Kim, “Effect of electron cyclotron beam width to neoclassical tearing mode stabilization by minimum seeking control in ITER,” *Nucl. Fusion*, vol. 58, no. 1, p. 016042, Jan. 2018, doi: 10.1088/1741-4326/aa95d1.

[12] J. Seo *et al.*, “Parametric study of linear stability of toroidal Alfvén eigenmode in JET and KSTAR,” *Nucl. Fusion*, vol. 60, p. 066008, 2020, doi: 10.1088/1741-4326/ab7fe2.

[13] A. Jalalvand *et al.*, “Alfvén eigenmode classification based on ECE diagnostics at DIII-D using deep recurrent neural networks,” *Nucl. Fusion*, vol. 62, no. 2, p. 026007, Dec. 2021, doi: 10.1088/1741-4326/AC3BE7.

[14] A. A. Kaptanoglu *et al.*, “Exploring data-driven models for spatiotemporally local classification of Alfvén eigenmodes,” *Nucl. Fusion*, vol. 62, no. 10, p. 106014, Aug. 2022, doi: 10.1088/1741-4326/AC8A03.

[15] Y. Fu *et al.*, “Machine learning control for disruption and tearing mode avoidance,” *Phys. Plasmas*, vol. 27, p. 022501, 2020, doi: 10.1063/1.5125581.

[16] F. Turco, T. C. Luce, W. Solomon, G. Jackson, G. A. Navratil, and J. M. Hanson, “The causes of the disruptive tearing instabilities of the ITER Baseline Scenario in DIII-D,” *Nucl. Fusion*, vol. 58, no. 10, p. 106043, Sep. 2018, doi: 10.1088/1741-4326/AADBB5.

[17] R. La Haye *et al.*, “Disruptive neoclassical tearing mode seeding in DIII-D with implications for ITER,” *Nucl. Fusion*, vol. 62, no. 5, p. 056017, Mar. 2022, doi: 10.1088/1741-4326/AC351F.

[18] J. Abbate, R. Conlin, and E. Kolemen, “Data-driven profile prediction for DIII-D,” *Nucl. Fusion*, vol. 61, no. 4, p. 046027, Mar. 2021, doi: 10.1088/1741-4326/ABE08D.

[19] T. N. Carlstrom *et al.*, “Design and operation of the multipulse Thomson scattering diagnostic on DIII‐D (invited),” *Rev. Sci. Instrum.*, vol. 63, no. 10, p. 4901, Jun. 1998, doi: 10.1063/1.1143545.

[20] R. P. Seraydarian and K. H. Burrell, “Multichordal charge‐exchange recombination spectroscopy on the DIII-D tokamak,” *Rev. Sci. Instrum.*, vol. 57, no. 8, p. 2012, Jun. 1998, doi: 10.1063/1.1138774.

[21] J. R. Ferron, M. L. Walker, L. L. Lao, H. E. S. T. John, D. A. Humphreys, and J. A. Leuer, “Real time equilibrium reconstruction for tokamak discharge control,” *Nucl. Fusion*, vol. 38, p. 1055, 1998, doi: 10.1088/0029-5515/38/7/308.

[22] R. Shousha, J. Ferron, Z. Xing, A. Nelson, K. Erickson, and E. Kolemen, “Improved Real-time Equilibrium Reconstruction with Kinetic Constraints on DIII-D and NSTX-U,” Oct. 2022, Accessed: Jan. 06, 2023. [Online]. Available: https://meetings.aps.org/Meeting/DPP22/Session/PP11.11.

[23] J. A. Stillerman, T. W. Fredian, K. A. Klare, and G. Manduchi, “MDSplus data acquisition system,” *Rev. Sci. Instrum.*, vol. 68, no. 1, pp. 939–942, Jun. 1997, doi: 10.1063/1.1147719.

[24] F. Chollet, “Keras: the Python deep learning API,” 2015. https://keras.io/ (accessed Jan. 06, 2023).

[25] R. Conlin, K. Erickson, J. Abbate, and E. Kolemen, “Keras2c: A library for converting Keras neural networks to real-time compatible C,” *Eng. Appl. Artif. Intell.*, vol. 100, p. 104182, Apr. 2021, doi: 10.1016/J.ENGAPPAI.2021.104182.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[27] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–15.

[28] 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, no. 5, p. 055002, Jul. 2021, doi: 10.1103/PHYSREVLETT.127.055002/FIGURES/3/MEDIUM.