

# Correlation of the L-mode density limit with edge collisionality

A D Maris<sup>1</sup>, C Rea<sup>1</sup>, A Pau<sup>2</sup>, W Hu<sup>3</sup>, B Xiao<sup>3</sup>, R Granetz<sup>1</sup>, E Marmar<sup>1</sup>, the EUROfusion Tokamak Exploitation team\*, the Alcator C-Mod team, the ASDEX Upgrade team\*\*, the DIII-D team, the EAST team, and the TCV team\*\*\*

<sup>1</sup> Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>École Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center (SPC), CH-1015 Lausanne, Switzerland

<sup>3</sup>Institute of Plasma Physics, Chinese Academy of Sciences, Hefei 230031, CN

\* See the author list of E. Joffrin et al Nucl. Fusion 2024

\*\* See the author list of U. Stroth et al 2022 Nucl. Fusion 62, 042006

\*\*\* See author list of H. Reimerdes et al 2022 Nucl. Fusion 62 042018

E-mail: maris@mit.edu

December 2024

**Abstract.** The “density limit” is one of the fundamental bounds on tokamak operating space, and is commonly estimated via the empirical Greenwald scaling. This limit has garnered renewed interest in recent years as it has become clear that ITER and many tokamak pilot plant concepts must operate near or above the Greenwald limit to achieve their objectives. Evidence has also grown that the Greenwald scaling - in its remarkable simplicity - may not capture the full complexity of the density limit. In this study, we assemble a multi-machine database to quantify the effectiveness of the Greenwald limit as a predictor of the L-mode density limit and compare it with data-driven approaches. We find that a boundary in the plasma edge involving dimensionless collisionality and pressure,  $\nu_{*,\text{edge}}^{\text{limit}} = 3.5\beta_{T,\text{edge}}^{-0.40}$ , achieves significantly higher accuracy (false positive rate of 2.3% at a true positive rate of 95%) of predicting density limit disruptions than the Greenwald limit (false positive rate of 13.4% at a true positive rate of 95%) across a multi-machine dataset including metal- and carbon-wall tokamaks (AUG, C-Mod, DIII-D, and TCV). This two-parameter boundary succeeds at predicting L-mode density limits by robustly identifying the radiative state preceding the terminal MHD instability. This boundary can be applied for density limit avoidance in current devices and in ITER, where it can be measured and responded to in real time.

*Keywords:* tokamak, density limit, machine learning

Submitted to: *Nucl. Fusion*

## 1. Introduction

Plasma electron density ( $n_e$ ) is a critical lever for fusion performance in tokamaks. High density is necessary for many burning plasma tokamak concepts to maximize fusion triple product  $nT\tau_E$  [1], enhance bootstrap current drive (via steeper density gradients) [2], and enable divertor detachment [3]. There has long been an interest in developing scaling laws to describe the highest achievable density in tokamaks [4, 5, 6]. Today, the most widely utilized empirical density limit scaling is the “Greenwald limit” [7], expressed as

$$\frac{\bar{n}}{n_G} = 1, \quad (1)$$

where  $\bar{n}$  is the central line-averaged electron density in units of  $10^{20} \text{ m}^{-3}$ ,  $n_G \equiv I_p/\pi a^2$  is the “Greenwald density,”  $I_p$  is the plasma current in MA, and  $a$  is the minor radius in meters. Operating near or above this limit correlates with confinement regime transitions when the plasma is in the “high” confinement mode (H-mode) and disruptions when the plasma is in the “low” confinement mode (L-mode). Nevertheless, to maximize fusion power, burning plasma experiments such as ITER [8] and fusion power plant (FPP) concepts (such as EU-DEMO [9], the compact advanced tokamak [2], and ARC [10]) are designed to operate near or above the Greenwald limit. Of course, by choosing to operate near an instability limit, future devices run the risk of harmful transients, such as H-to-L back-transitions and disruptions. Even infrequent unmitigated disruptions - such as once a month - could render tokamak power plants uneconomical given the long timescales needed for repairs [11]. Therefore, tokamak power plants require large safety margins and/or extremely effective control solutions for the density limit and other instabilities.

While a complete, first-principles treatment of the density limit remains elusive, theory and experiment have clarified the characteristic dynamics, summarized in Fig. 1. The path to the density limit begins with edge density increasing and/or edge temperature decreasing [12]. Past a certain threshold, a thermal instability occurs at the plasma edge, causing a collapse of the edge temperature. If the plasma is operating in H-mode, it experiences an H-to-L back-transition, referred to as an “H-mode density limit” (HDL). In L-mode, this temperature collapse is associated with the formation of an X-point radiator or a MARFE - a toroidally symmetric ring of cool, dense plasma on the high-field side [13]. The HDL is not a disruptive instability itself, but can be followed by the “L-mode density limit” (LDL). The LDL occurs when the edge cooling causes the plasma current to concentrate in a peaked current profile [14, 15]. When current channel is sufficiently narrow, it loses MHD stability, causing a disruption.

Theories attempting to explain the thermal instability tend to come in two flavors: a radiative instability in the edge [16, 17] or enhanced turbulent transport in the edge [18, 19, 20, 21]. It has been shown that many of these models share qualitative similarities to each other [22].

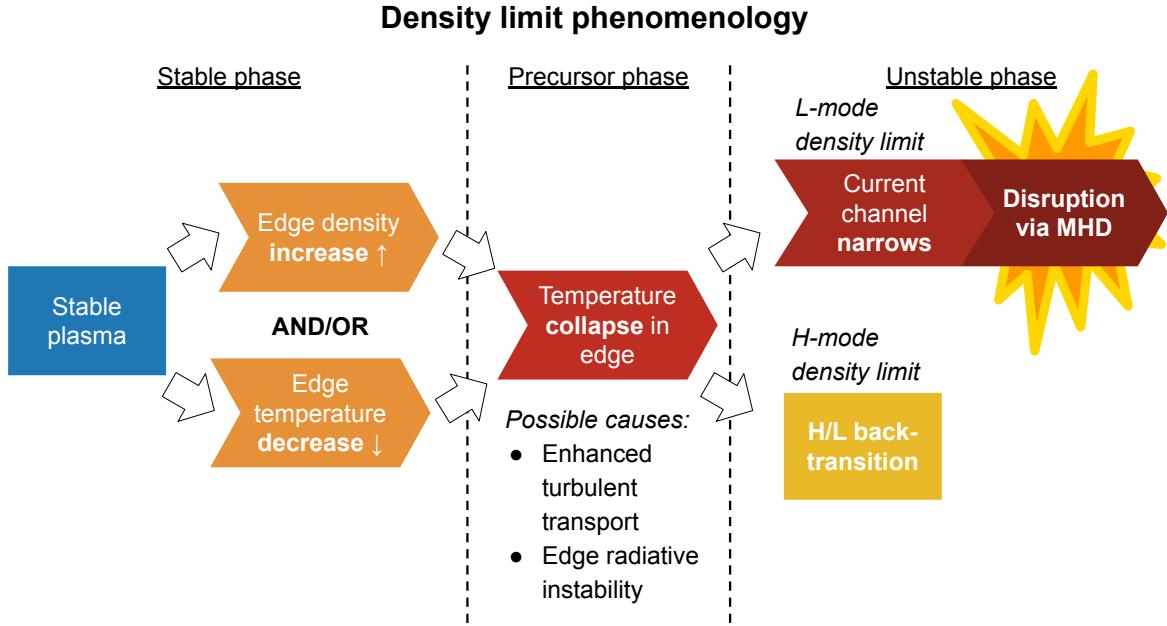


Figure 1: The characteristic chain of events for density limits in tokamaks. If the discharge is in H-mode, it suffers an H-to-L back-transition; this is referred to as an “H-mode density limit.” If it is in L-mode, the current channel shrinks, ultimately resulting in an “L-mode density limit” disruption.

This sharpening picture of the density limit suggests that burning plasmas may be able to exceed the Greenwald limit for two reasons. Because the density limit depends on the edge of the plasma, experiments with density peaking have achieved  $\bar{n}/n_G > 1$  while maintaining  $n_{\text{edge}}/n_G < 1$  [23, 24, 25, 26, 27]. It is expected that burning plasmas will naturally exhibit density peaking due to their low collisionality [28]. Additionally, several studies following Ref. [7] have observed a modest input power scaling of  $P^{0.2-0.6}$  augmenting the Greenwald limit [23, 19, 22, 24, 25, 29, 30, 31]. This power dependence is understood to be due to higher input power raising the temperature at the edge and holding off the onset of the temperature instability.

At the same time, we will show in this paper that these two observations alone are not sufficient to achieve the extremely high disruption prediction accuracy required for ITER [32]. The Greenwald limit was not derived with disruption prediction in mind, and so it is perhaps not surprising that there is room for improvement for a density limit disruption forecaster. It is notable however, that we must go beyond just applying a power scaling and using the edge density, we must instead combine edge temperature and density to predict LDLs with high accuracy.

In this paper, we assemble and study a multi-machine database of LDL events from ASDEX UpGrade (AUG), Alcator C-Mod, DIII-D, EAST, and TCV. We apply a variety of techniques to predict the onset of the instability, thus finding:

Table 1: Number of unique discharges and time steps in the database assembled for this study, divided into discharges that feature an L-mode density limit (LDL) and those that do not (stable).

Device	LDL discharges	LDL time steps	Stable discharges	Stable time steps
AUG	33	1,106	8	16,231
C-Mod	92	3,819	2,429	275,492
DIII-D	42	2,225	1,073	367,171
EAST	13	1,498	669	683,750
TCV	32	1,511	74	11,004
Total	212	10,159	4,253	1,353,648

- 1) The Greenwald limit does not universally predict the onset of L-mode density limit events (false positive rate of  $> 13\%$  for a true positive rate of 95%).
- 2) Data-driven models trained to predict the radiative precursor phase achieve significantly improved L-mode density limit prediction accuracy (false positive rate of  $< 3\%$  for a true positive rate of 95%).
- 3) In particular, a simple stability boundary in terms of the effective collisionality and  $\beta_T$  in the plasma edge,  $\nu_{*,\text{edge}}^{\text{limit}} = 3.5\beta_{T,\text{edge}}^{-0.40}$ , is a highly reliable proximity-to-density-limit metric (false positive rate of 2.3% for a true positive rate of 95%).

The paper is organized as follows: Section 2 describes the methods used for the dataset assembly and the data-driven analysis, Section 3 describes the prediction performance of various models on an unseen test set, Section 4 discusses the relation to existing density limit observations and considers example discharges from the database, and Section 5 summarizes the findings of this study and outlines future work.

## 2. Methods

### 2.1. Dataset and labeling

The dataset utilized for this study is composed of discharges from five tokamaks: AUG, C-Mod, DIII-D, EAST, and TCV. The C-Mod and DIII-D database of LDLs are newly collated for this study based on data fetching workflows utilized in Ref. [33, 34]. The LDL shots from AUG and TCV in this study appeared in Ref. [19], and those from EAST appeared in Ref. [35]. The number of discharges and samples from each device is shown in Table 1, and the global parameters of these devices can be found in Appendix A, Table A1. The reader should note the significant variation in the number discharges available for each device due to different data availability, frequency of density limit experiments, and lifetimes of the machines.

Density limit discharges were manually labeled by the authors using the pattern observed across all devices: an increase in density and/or decrease in edge temperature, followed by the formation of a radiator or MARFE near the X-point, which eventually

destabilizes and moves towards the core, resulting in a disruption. For this study, the LDL precursor start time was labeled manually as the time of the X-point radiator formation when such measurements are available (AUG, C-Mod, DIII-D, and TCV) and a fixed 100 ms window before the MARFE destabilizes otherwise (EAST). In AUG and TCV, the DEFUSE tool [36] automatically tags candidate events which are subsequently manually validated by an expert. For DIII-D, the formation time is determined by manual inspection of individual bolometer chords and 2D tomographic reconstructions of the poloidal radiation cross-section. For C-Mod, this is done via inspection of an H-alpha chord and visible camera images. The LDL precursor end time was manually labeled to occur as the radiation front moves inward toward the core of the plasma before the disruption. An example of this labeling is schematically represented in Figure 2.

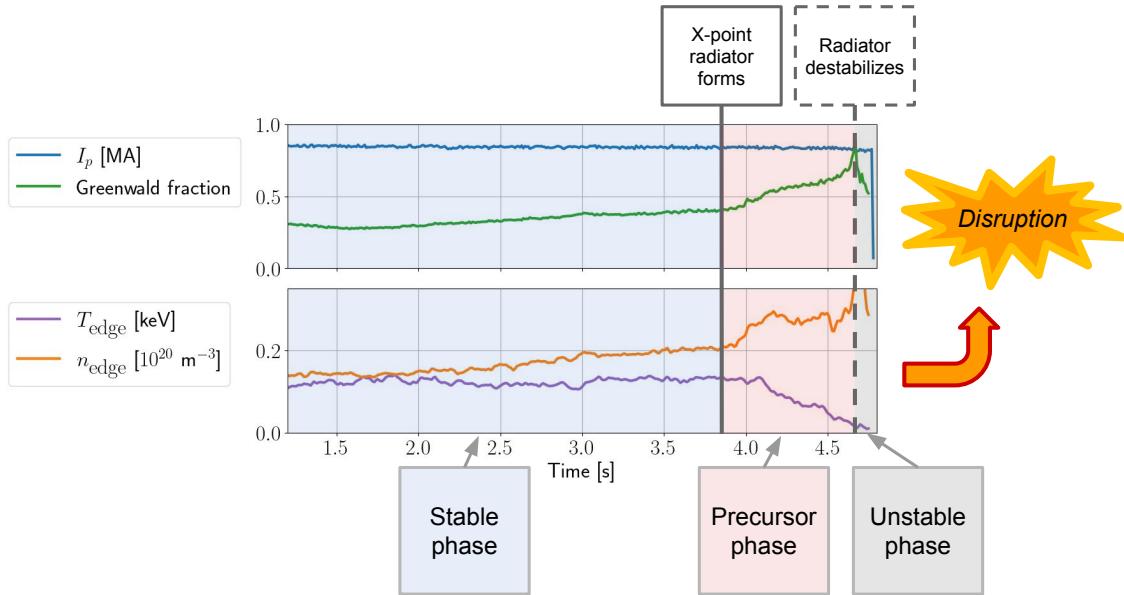


Figure 2: Labeling the L-mode density limit for an example discharge, DIII-D #191794. The LDL precursor phase is the time period between when the X-point radiator forms and when the radiator becomes unstable and moves toward the plasma core. We ignore the window in time after the radiator is destabilized. Here, “edge” density and temperature are defined as the average measurement of the quantity between  $\rho = 0.85$  and  $0.95$ .

To isolate density limit dynamics from other instabilities, LDL candidates were excluded from the database if 1) the operators noted major impurity injections, 2) there was significant MHD activity prior to the formation of the X-point radiator, 3) or the disruption was immediately proceeded by a sudden shutoff or failure of input power.

Non-disruptive discharges, also referred to here as “stable” discharges, are uniformly sampled from the set of discharges in each device that did not result in a disruption during flattop and did not experience control errors. Stable discharges that experienced minor disruptions were also excluded. A correlation matrix for the dataset is reported

Table 2: Features in the dataset, as well as “features sets” used in analysis: “global” features, “edge” features, and dimensionless features.

Symbol	Definition	“Global” features	“Edge” features	Dimensionless features
$\bar{n}$	e <sup>-</sup> density, line avg.	X		
$n_{\text{edge}}$	e <sup>-</sup> density, edge		X	
$P_{\text{in}}$	Input power	X		
$T_{\text{edge}}$	e <sup>-</sup> temperature, edge		X	
$I_p$	Plasma current	X	X	
$a$	Minor radius	X	X	
$q_{95}$	Safety factor		X	X
$\nu_{*,\text{edge}}$	Collisionality, edge			X
$\beta_{T,\text{edge}}$	Toroidal $\beta$ , edge			X
$\rho_{*,\text{edge}}$	Norm. gyroradius, edge			X

in Appendix A.

## 2.2. Feature set

Table 2 lists the signals, which we refer to as “features”, used in this analysis. Edge density and temperature in this study are defined as the Thomson Scattering measurements averaged between a normalized radius ( $\rho$ ) of 0.85 and 0.95, as was used for experimental validation of an edge density limit in Ref. [19]. A simple fitting procedure was used to determine the profiles of AUG and TCV, while linear interpolation was used for C-Mod and DIII-D. Averaging over this relatively large edge region reduces the impact of measurement noise. Normalized radius is defined in terms of the square root of the normalized toroidal flux for DIII-D data and the square root of the normalized poloidal flux for AUG, TCV, and C-Mod data. All signals are resampled onto a 10ms timebase for consistency.

As shown in Table 2, we conduct our analyses using three distinct sets of features: “global” features, “edge” features, and dimensionless features. The global features are macroscopic plasma parameters that are relatively easy to measure (ex.  $\bar{n}$ ,  $I_p$ ) and typically utilized in density limit scalings. The edge features are similar, but with the line-averaged density and input power replaced with the edge density and temperature, respectively. These latter two parameters are expected to be more predictive of the density limit because they are local to the edge region where the density limit is thought to be triggered. We also add the edge safety factor,  $q_{95}$ , which may capture additional information related to the connection length at the edge of the plasma. Because of noise in the measurement of edge density and temperature, a Butterworth filter is applied to these signals with a critical frequency of 8 Hz and 6 Hz, respectively. For the sake of cross-device consistency, the same filter is applied to all devices.

Due to the strong correlation in our dataset between minor radius  $a$ , major radius

$R_0$ , and the toroidal magnetic field  $B_T$ , our primary results will only include the minor radius as in the Greenwald scaling. We seek to avoid multicollinearity in training models because it can mask true variable interactions. For the same reason, we do not train models with both line-average density  $\bar{n}$  and edge density  $n_{\text{edge}}$ . Additionally, we have measurements for inverse aspect ratio  $\epsilon$ , elongation  $\kappa$ , and triangularity  $\delta$  across our database, but we exclude them from the analysis as there is too little variation among these parameters to be of use in this study.

We also analyze a set of dimensionless variables generally following the definitions used in [37], but with ion density and temperature replaced with the electron value. The dimensionless variables we use include  $q_{95}$  and the following three variables:

$$\nu_{*,\text{edge}} \equiv \frac{\nu_{ii} q_{\text{cyl}} R_0}{v_{ti} \epsilon^{3/2}} \approx \frac{e^4 \ln(\Lambda)}{2\pi \epsilon_0^2} \frac{n_{\text{edge}}}{T_{\text{edge}}^2} \frac{q_{\text{cyl}} R_0}{\epsilon^{3/2}}, \quad (2)$$

$$\rho_{*,\text{edge}} \equiv \frac{\rho_i}{a} \approx \frac{\sqrt{m_i T_{\text{edge}}}}{e B_T a}, \quad (3)$$

$$\beta_{T,\text{edge}} \equiv \frac{2n_{\text{edge}} T_{\text{edge}}}{B_T^2 / (2\mu_0)}, \quad (4)$$

where  $\nu_{ii}$  is the ion-ion collision frequency,  $q_{\text{cyl}} \equiv \frac{2\pi}{\mu_0} \frac{B_T a^2}{I_p R_0} \left( \frac{1+\kappa^2}{2} \right)$  is the cylindrical safety factor,  $v_{ti}$  is the ion thermal speed,  $e$  is the elementary charge,  $\ln(\Lambda)$  is the Coulomb logarithm,  $\epsilon_0$  is the permittivity of free space,  $\rho_i$  is the ion gyroradius, and  $m_i$  is the ion mass (assumed to be deuterium), and  $\mu_0$  is the permeability of free space. We use  $q_{95}$  instead of  $q_{\text{cyl}}$  as the fourth feature in the dimensionless feature case to capture effects of shaping (ex. triangularity) not included in the cylindrical approximation.

### 2.3. Problem formulation and performance metrics

We choose to formulate DL prediction as a supervised classification problem: we will attempt to find a model that will accurately classify plasma states as stable or in the LDL precursor phase with sufficient warning time before the instability occurs. Following standard practices, we will hold out 20% of the discharges as the test set: these discharges will be only used to evaluate the performance of the model. The remainder of the data will be used in the training set for the models to learn on.

A discharge is classified as being in the LDL precursor phase - the “positive” class - for a given alarm threshold if two conditions are met: 1) the instability score rises above the alarm threshold for two or more consecutive time steps (i.e. 10ms assertion time) **and** 2) the alarm occurs  $> 30$  ms before the radiator destabilizes. The first condition is intended to prevent spurious alarm triggers due to an anomalous measurement at a single time step, and the second condition discounts predictions that are “too late” for a disruption mitigation system (DMS) to intervene. One could instead define a tardy alarm in terms of the time needed for disruption avoidance, but this would vary depending on tokamak, actuator type, and plasma scenario. For the sake of simplicity, we choose a well-defined DMS timescale for setting the late alarm threshold,

and leave a more thorough treatment of disruption avoidance timescales for a later study. Specifically, we choose a minimum warning time of 30 ms to match the time needed for actuating the ITER DMS [38]. We also note that an alarm that occurs significantly before the LDL time is still considered a true positive, as we do not want to penalize a model for providing a long warning time that could be used in practice for LDL avoidance.

In classification tasks such as this, the goal is to achieve a high true positive rate (TPR) and low false positive rate (FPR). Concretely, the TPR is the fraction of LDL shots (the “positive” case) that are correctly predicted to be an LDL, and the FPR is the fraction of stable shots (the “negative” case) that are incorrectly predicted to have an LDL event.

For any proximity-to-instability model that provides a continuous instability score, we must choose an alarm threshold above which to predict the shot will end in an LDL. For example,  $n/n_G = 1$  is often considered the standard threshold for the LDL, but the threshold could be adjusted to change the sensitivity level. A lower alarm threshold will be more sensitive, and have a higher TPR and FPR. On the flip side, a higher alarm threshold will have a lower TPR and FPR. In sum, all proximity-to-instability models have a tradeoff between TPR and FPR.

We report two performance metrics for each model: “Area Under the Curve” (AUC) and the FPR at a fixed TPR of 95% (shorthand: FPR @ TPR = 95%). The AUC is the average TPR across the range of FPR  $\in [0, 1]$ , which quantifies the performance of the model across the full range of alarm threshold levels. The FPR @ TPR = 95% metric, by contrast, represents the proportion of stable discharges that are incorrectly classified when we require exceptional detection performance of LDLs. This is important for ITER and future tokamak power plants, where the potential damage from disruptions necessitates near-perfect (TPR  $\geq 95\%$ ) prediction of disruptions.

#### 2.4. LDL prediction models

We hypothesize that we can achieve higher LDL prediction accuracy than the Greenwald fraction by training data-driven models to discriminate the stable and LDL precursor phase. The motivation for this choice is that the precursor phase is a distinct and relatively long-lived regime, which provides a coherent target for the models to fit. The pitfall of this approach is that it results in a subtle difference between the training objective (classifying time steps as stable or LDL precursor) and the performance metrics for judging the models (classifying discharges as stable or LDL disruptions). The precursor regime is, of course, a necessary *but not sufficient* condition for an LDL disruption to occur; these models could therefore be vulnerable to false positives from discharges that enter the LDL precursor regime but do not become MHD unstable. Nevertheless, we will show that this approach results in high LDL prediction accuracy, and leave the treatment of the MHD instability phase of the LDL for a future study.

In this study, we evaluate two types of sequence-to-sequence classification

architectures: non-symbolic and symbolic. The two non-symbolic models are standard machine learning workhorses - the neural network (NN) and random forest (RF). Details about hyperparameters scans are reported in Appendix B. Hyperparameters are selected via the maximum AUC on the validation set, a randomly assigned set of 20% of discharges withheld from the training set.

We also attempt to find a symbolic density limit boundary using two methods: linear regression and linear support vector machines (LSVM). Symbolic models are simply models that take on an analytic form (for example, the Greenwald fraction is a symbolic model). To identify the linear regression boundary, we average over the last 50 ms before the LDL and use multivariate linear regression to find a power law that minimizes the mean squared error over the training set. We encourage a parsimonious model by computing the p-value of each feature, removing the feature with the largest p-value above 0.05, and re-training until all features in the regression model have p-values less than 0.05. We find a power law boundary using LSVMs by training a classifier on all data points in the training set (just as we do for the NN and RF). Feature combinations are explored via sequential feature selection with backward elimination using the Bayesian Information Criterion as the evaluation metric,

$$\text{BIC} = -2 \ln(L) + k \ln(s), \quad (5)$$

where  $L$  is the likelihood of the model evaluated on the training set,  $k$  is the number of regression variables in the model, and  $s$  is the number of samples. The BIC balances a reward for low error (low negative log-likelihood) with a penalty for more parameters in the model, weighted by the log of the number of samples used to fit the parameters. As plasmas states are dynamically evolving in time during discharges and not independently sampled, we approximate  $s$  as the number of discharges. We similarly adjust the likelihood  $L \equiv \sum y \ln(p) + (1 - y) \log(1 - p)$  by rescaling it by the ratio of number of discharges to number of time steps.

Finally, we will compare the model predictions with that of the Greenwald fraction using the line-average density and the edge density. These scalings will be used as baselines for comparison to the data-driven approaches.

### 3. Results

#### 3.1. Predicting the LDL with “global” features

Table 3 shows the test set performance of L-mode density limit (LDL) prediction trained on the “global” features (Table 2) in comparison to the Greenwald scaling. The symbolic boundaries are all written as proportionalities as the magnitude of the coefficient adjusts the alarm threshold. We cannot compute the Greenwald fraction for the edge density as the discharges from EAST lack this signal.

We find that the NN, RF, and LSVM are the most accurate models, far outperforming the AUC and FPR of the Greenwald scaling. The linear regression model, by contrast, has performance levels between the other data-driven models and

Table 3: The test set performance of LDL prediction for models trained on the global features for all devices: AUG, C-Mod, DIII-D, EAST, and TCV. The best performance for each metric (highest AUC, lowest FPR) are bolded for emphasis. The Greenwald fraction is the least accurate of all models tested.

Model	Analytic boundary	AUC	FPR @ TPR = 95%
NN	N/A	<b>0.943</b>	28.4%
RF	N/A	<b>0.943</b>	<b>22.6%</b>
LSVM	$\bar{n}^{\text{limit}} \sim \frac{I_p^{0.67}}{a^{1.80}} P_{\text{in}}^{0.28}$	0.941	26.8%
Lin. Reg.	$\bar{n}^{\text{limit}} \sim \frac{I_p^{0.75}}{a^{2.04}} P_{\text{in}}^{0.17}$	0.925	39.5%
Greenwald	$\bar{n}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.894	46.0%

the Greenwald scaling. This illustrates the value of using a classification algorithm for this problem. Interestingly, the LSVM takes a similar form to the linear regression model – a Greenwald-like scaling with lower current dependence and an additional  $P_{\text{in}}$  scaling – but achieves nearly the same performance as the NN and RF. In Fig. 3, we plot the density against the remaining variables in the LSVM power law.

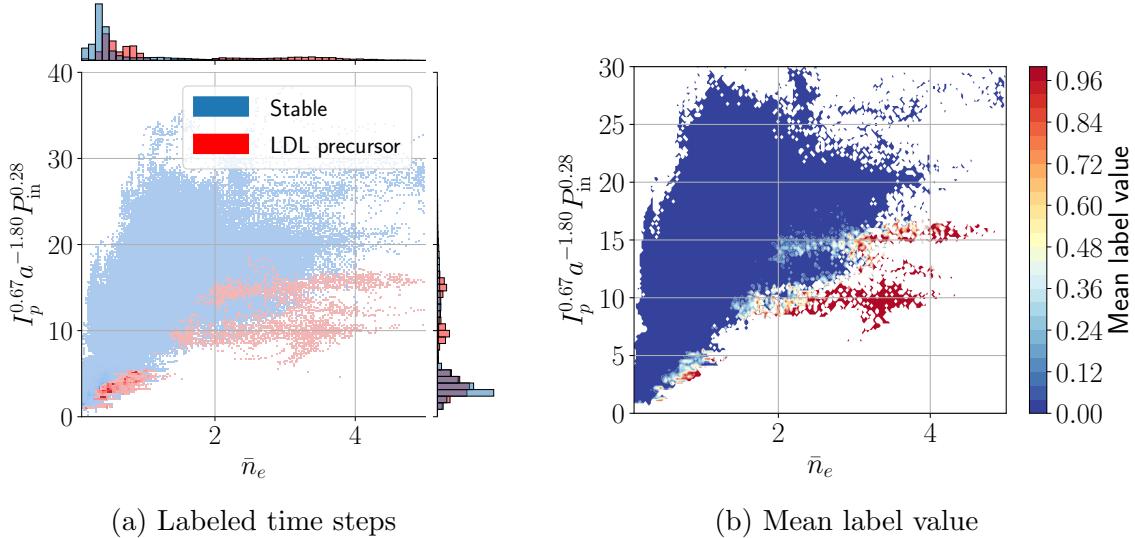


Figure 3: The distribution of line averaged density vs. the remaining terms of the LSVM power law (Table 3) across the database. Subplot 3a shows the LDL points (red) superimposed on the non-DL points (blue), while 3b shows a “stability” heat map, where labels (“stable” = 0, LDL precursor = 1) have been averaged in each bin.

Although the non-symbolic models (NN and RF) achieve higher performance than the Greenwald fraction, we note that a FPR of  $> 20\%$  would still be very costly for the mission of ITER and FPPs.

Table 4: Number of unique discharges and time steps in the database that have edge density and temperature measurements. These data are used for the analysis in section 3.2 and 3.3.

Device	LDL discharges	LDL time steps	Stable discharges	Stable time steps
AUG	30	1,063	8	15,523
C-Mod	52	3,322	2,162	243,221
DIII-D	41	2,205	996	341,717
EAST	0	0	0	0
TCV	30	1,384	27	5,776
Total	153	7,974	3,193	606,237

Table 5: The test set performance of LDL prediction for models trained on the edge features, as well as the Greenwald fraction and edge Greenwald fraction.

Model	Analytic boundary	AUC	FPR @
			TPR = 95%
NN	N/A	0.997	2.8%
RF	N/A	<b>0.998</b>	<b>0.5%</b>
LSVM	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p^{0.79}}{a^{1.30}} T_{\text{edge}}^{1.00}$	0.996	2.3%
Lin. Reg.	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p^{0.86}}{a^{1.48}} T_{\text{edge}}^{0.11} q_{95}^{0.66}$	0.880	54.6%
Greenwald	$\bar{n}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.971	13.9%
Edge Greenwald	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.888	43.7%

### 3.2. Predicting the LDL with “edge” features

As stated previously, edge density and temperature are understood to be key parameters for the onset of the DL. Unfortunately, not all discharges in our dataset have Thomson scattering measurements of the edge. Therefore, the dataset for the “edge” feature analysis has a different composition of discharges, shown in Table 4. Particularly of note is the absence of EAST data. The change in composition of the training and test set leads to different performance for the line-averaged Greenwald scaling compared to the previous section.

As shown in Table 5, the RF is the best performing model, followed closely by the NN and LSVM power law. The linear regression and Greenwald fraction scalings achieve significantly worse performance.

Interestingly, the LSVM boundary takes a similar form to the Greenwald fraction with an approximately linear temperature dependence. Calibrated to TPR = 95%, the limit is

$$n_{\text{edge}}^{\text{limit}} = 1.4 \frac{I_p^{0.79}}{a^{1.30}} T_{\text{edge}}^{1.00}, \quad (6)$$

We show the state space of the edge density vs. the remaining terms in the LSVM power law in Fig. 4. We see stronger separation of the stable and LDL precursor phases

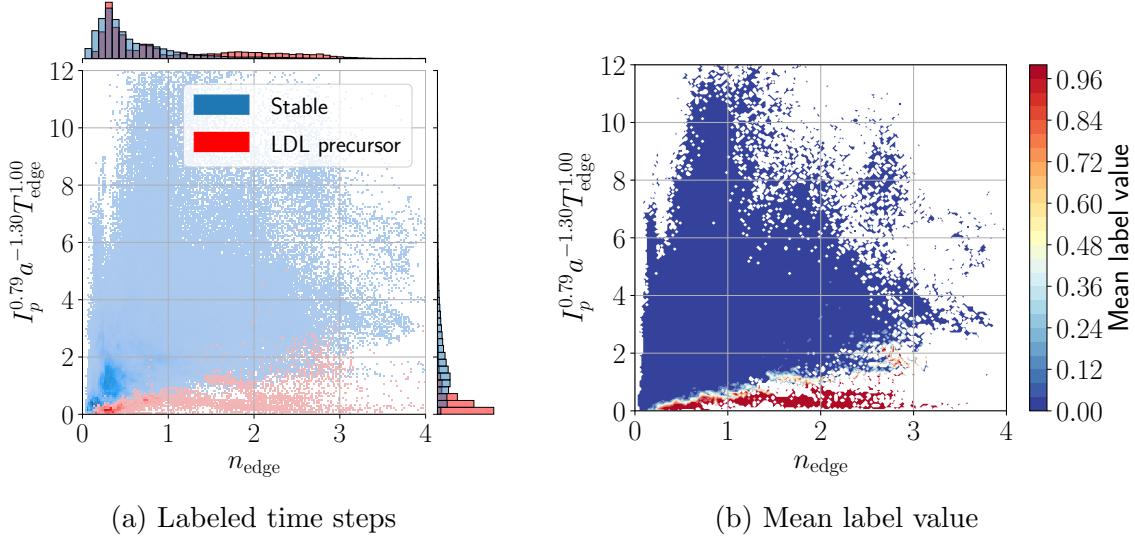


Figure 4: The distribution of edge density vs. the remaining terms of the LSVM power law for the database (Table 5). Subplot 4a shows the LDL points (red) superimposed on the non-DL points (blue), while 4b shows a “stability” heat map, where labels (“stable” = 0, LDL precursor = 1) have been averaged in each bin.

compared to the global features case in Fig. 3.

Once again, the linear regression power law performs significantly worse compared to the LSVM power law. The form of the boundaries are similar except for a much lower edge temperature exponent for the linear regression and the addition of a moderate  $q_{95}$  dependence. The significantly diminished performance is primarily due to this the weaker temperature scaling.

We note that the Greenwald fraction model has higher performance compared to the results in section 3.1 due to the different make-up of the dataset, as stated earlier. Nevertheless, this improved performance is still far below that of the LSVM, NN, and RF.

### 3.3. Predicting the LDL with dimensionless features

When trained on the dimensionless set of features  $\nu_{*,\text{edge}}$ ,  $\rho_{*,\text{edge}}$ ,  $\beta_{T,\text{edge}}$ , and  $q_{95}$ , the data-driven models achieve similarly strong performance as is found in the edge features case (section 3.2). The test set performance metrics are reported in Table 6.

The NN, RF, and LSVM all achieve similar AUC as in the previous section (subsection 3.2) and slightly higher FPRs. The power law boundary identified by the LSVM calibrated to TPR = 95% is

$$\nu_{*,\text{edge}}^{\text{limit}} = 3.5 \beta_{T,\text{edge}}^{-0.40}. \quad (7)$$

The space defined by these two variables is shown in Fig. 5, illustrating relatively strong discrimination between the stable and LDL precursor points. The top marginal

Table 6: The test set performance of LDL prediction for models trained on the dimensionless features, as well as the Greenwald fraction and Edge Greenwald fraction.

Model	Analytic boundary	FPR @ TPR = 95%	
		AUC	
NN	N/A	0.991	3.0%
RF	N/A	0.996	<b>1.6%</b>
LSVM	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.40}$	<b>0.997</b>	2.3%
Lin. Reg.	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.67} \rho_{*,\text{edge}}^{-0.77}$	0.984	6.6%
Greenwald	$\bar{n}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.971	13.9%
Edge Greenwald	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.888	43.7%

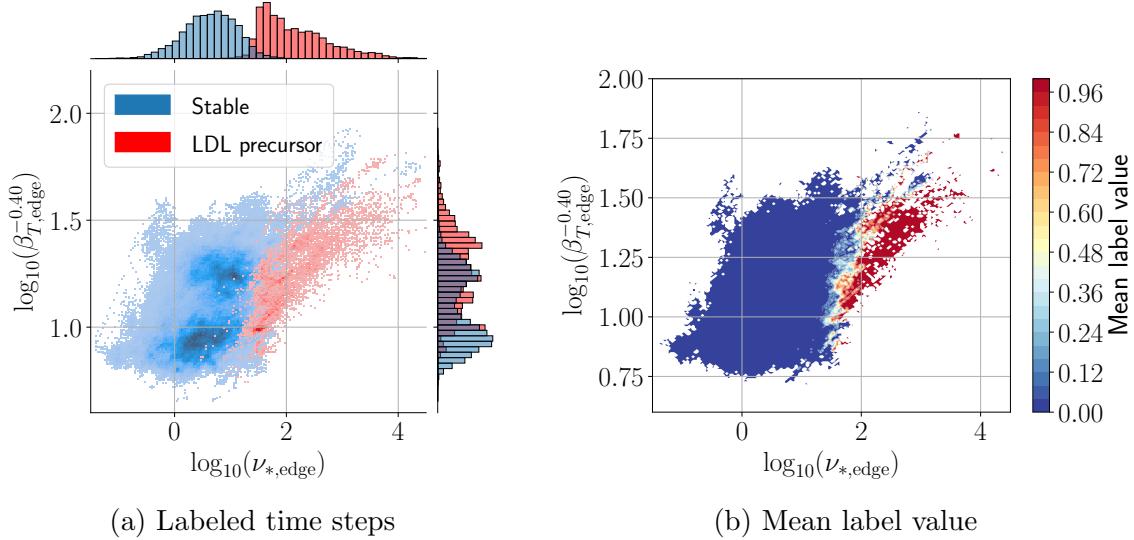


Figure 5: The distribution of edge collisionality versus  $\beta_{T,\text{edge}}^{-0.40}$  (Table 6). Subplot 5a shows the LDL points (red) superimposed on the non-DL points (blue), while 5b shows a “stability” heat map, where labels (“stable” = 0, LDL precursor = 1) have been averaged in each bin.

plot of Fig. 5a shows a histogram of data with respect to  $\nu_{*,\text{edge}}$ , highlighting that the collisionality term alone provides a good degree of discrimination between the cases.

## 4. Discussion

### 4.1. Relation of results to the Greenwald limit

Each of the LSVM-derived instability metrics bears some explicit or implicit resemblance to the Greenwald limit. For ease of reference, we assemble the LSVM boundaries in Table 7 and introduce shorthands for each case.

In the case of the global feature set, the LSVM-G metric takes the form of a Greenwald-like scaling with an additional power dependence. The plasma current and

Table 7: LSVM-derived instability metrics for the LDL-precursor phase, as well as the shorthand used to refer to them.

Feature set	Analytic boundary	Instability metric shorthand
Global features	$\bar{n}^{\text{limit}} \sim \frac{I_p^{0.67}}{a^{1.80}} P_{\text{in}}^{0.28}$	LSVM-G
Edge features	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p^{0.79}}{a^{1.30}} T_{\text{edge}}^{1.00}$	LSVM-E
Dimensionless features	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.40}$	LSVM-D

input power dependencies,  $I_p^{0.67}$  and  $P^{0.28}$ , are within the  $I_p^{0.5-1.0}$  and  $P^{0.2-0.6}$  ranges reported in the literature [17, 19, 22, 25, 29, 30, 39]. Additionally, the minor radius dependence ( $a^{-1.80}$ ) is close to that of the Greenwald fraction ( $a^{-2}$ ). These differences are somewhat subtle, but result in the LSVM-G metric achieving two times lower FPR @ TPR = 95% and significantly higher AUC compared to the Greenwald fraction.

The LSVM-E metric is a Greenwald-like scaling in terms of edge density, with a sub-linear plasma current scaling ( $I_p^{0.79}$ ), a smaller minor radius dependence ( $a^{-1.30}$ ), and a linear edge temperature dependence. As with the LSVM-G metric, these differences result in far better performance than the Greenwald fraction; the LSVM-E metric achieves nearly six times lower FPR @ TPR = 95%.

Interestingly, it can be shown the LSVM-E and LSVM-D scalings are nearly equivalent. In the dimensionless case,

$$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.4}, \quad (8)$$

can be rewritten as

$$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p^{0.7} T_{\text{edge}}^{1.1}}{a^{1.4}} \left( \ln(\Lambda)^{0.7} B_T^{0.1} \epsilon^{-0.6} \kappa^{0.7} \right), \quad (9)$$

which almost exactly matches LSVM-E. The term within the parentheses,  $B_T^{0.1} \epsilon^{0.4} \kappa^{0.7}$ , varies weakly across the entire database (standard deviation < 10% of the mean). Despite the fact the dimensionless case has a more restrictive set of features, the LSVM arrives at a nearly identical solution.

Figure 6 shows a plot of the timeslices associated with stable plasmas and the LDL precursor phase for the Greenwald fraction and LSVM-D. We see the LSVM-D metric is able to better discriminate between the LDL precursor phase and stable plasma states. Even when the LSVM is applied to a subset of machines (Appendix C and Appendix D), it finds the same pattern and achieves similar performance levels.

Notably, the  $n^{\text{limit}} \sim P_{\text{in}}^{0.28}$  and  $n^{\text{limit}} \sim T_{\text{edge}}^{1.00}$  dependencies identified in this study echo the approximation

$$\bar{T}_{e,\text{sep}} \sim P_{\text{SOL}}^{2/7}, \quad (10)$$

where  $\bar{T}_{e,\text{sep}}$  is the average electron temperature at the last closed flux surface and  $P_{\text{SOL}} \equiv P_{\text{in}} - P_{\text{rad}}$  is the power through the SOL (input power minus power radiated within the LCFS), which is valid when parallel heat conduction dominates parallel heat

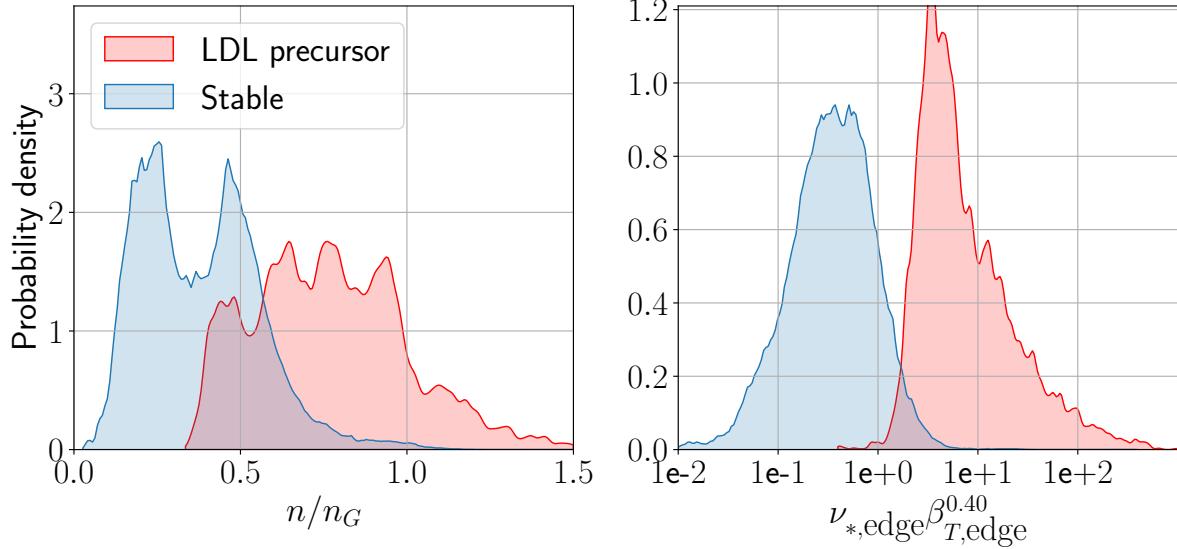


Figure 6: A comparison of the time slices in the stable and LDL precursor phase in terms of both the Greenwald fraction and LSVM-D metric. The LSVM-D metric more clearly separates the precursor phase from stable plasma states.

convection [40]. Of course  $T_{\text{edge}}$  is not  $\bar{T}_e$ , and  $P_{\text{in}}$  is not  $P_{\text{SOL}}$ , but one might expect strong correlations between these parameters.

In summary, the LSVM-derived scalings are generally consistent with past observations of a Greenwald-like scaling for the density limit and a moderate power dependence. Despite these similarities, the LSVM scalings achieve significantly improved LDL prediction accuracy.

#### 4.2. Relation of collisionality boundary to electron adiabaticity

Theoretical treatments of the density limit [18, 41] and empirical studies on individual devices [42, 43] have suggested electron adiabaticity

$$\alpha \equiv \frac{k_{\parallel}^2 v_{te}^2}{\nu_{ee} \omega}, \quad (11)$$

in the plasma edge is a critical parameter for the density limit, where  $k_{\parallel}$  is the wavenumber along the magnetic field line (usually taken to be the inverse of the connection length  $L_c \sim qR$ ),  $v_{te}$  is the electron thermal speed,  $\nu_{ee}$  is the electron-electron collision frequency, and  $\omega$  is the peak turbulence frequency. The regime  $\alpha < 1$  is thought to result in increased turbulent transport through the emergence of Resistive Ballooning Modes (RBMs) [18] or by shear layer collapse [41]. If the enhanced transport sufficiently cools the edge, the current channel narrows and an X-point radiator (XPR) or MARFE can form. Although the degradation of the shear layer is not itself a radiative mechanism, it can cause a collapse of the edge temperature to force the plasma into the radiative precursor state of the LDL.

The adiabaticity parameter  $\alpha$  is challenging to measure in practice because it involves the measuring fluctuation in the plasma edge. We can show, however, that the LSVM-D metric can be re-written in a form similar to the electron adiabaticity. Taking  $k_{\parallel} \sim 1/q_{95}R_0$  in the plasma edge (as in Ref. [42]), one can show

$$(\nu_{*,\text{edge}}\beta_{T,\text{edge}}^{0.40})^{-1} \sim \frac{k_{\parallel}^2 v_{te}^2}{\nu_{ee}\omega_{\text{imp}}}, \quad (12)$$

where the implied frequency,  $\omega_{\text{imp}}$ , is

$$\omega_{\text{imp}} \equiv \frac{T_{\text{edge}}^{0.9}}{B_T^{0.8}} \frac{n_{\text{edge}}^{0.4} k_{\parallel}}{\epsilon^{3/2}}. \quad (13)$$

The implied frequency has temperature and magnetic field dependencies similar to those of the electron diamagnetic drift frequency

$$\omega_{*e} = \frac{T}{B} \frac{k_{\perp}}{en} \frac{dn}{dr}. \quad (14)$$

Beyond the leading  $T$  and  $B$  terms, and the  $\epsilon^{3/2}$  term that is relatively fixed across the database, the remaining terms do not obviously agree. We might expect a discrepancy because, as adiabaticity breaks down, the turbulence should no longer be purely drift waves. Direct measurements of the fluctuation frequency or the density gradient at the edge would help elucidate this matter.

#### 4.3. Relation of collisionality boundary to Stroth et al. X-point radiator model [16]

Reference [16] presents a scaling for the formation of an XPR by identifying the threshold at which power conducted through the edge of the plasma no longer balances the ionization and charge exchange losses. They estimate the threshold density for XPR formation to be

$$n_u^{\text{XPR}} \sim \frac{T_u^{5/2}}{n_0} \frac{a}{f_{\text{exp}} R_0^2 q_s^2} \quad (15)$$

where  $n_u^{\text{XPR}}$  is the upstream density for XPR formation,  $T_u$  is the upstream temperature,  $n_0$  is the neutral density,  $f_{\text{exp}}$  is the flux expansion factor, and  $q_s$  is the safety factor of a cylindrical plasma. We cannot evaluate the full LDL model from Ref. [16] across our database, as they present a second condition involving impurity concentration for the XPR to become an unstable MARFE. However, we can arrive at an approximation of the XPR scaling

$$n_{\text{edge}}^{\text{limit}} \sim \frac{T_{\text{edge}}^{5/4} \sqrt{a}}{q_s R_0}. \quad (16)$$

by taking the neutral density to be proportional to the upstream density (as suggested in Ref. [22]), taking the upstream density and temperature to correspond to the edge density and temperature, and assuming roughly fixed flux expansion in the XPR region across all scenarios. This expression has a similar density and temperature relationship as in LSVM-E, however we note significant discrepancies between the macroscopic

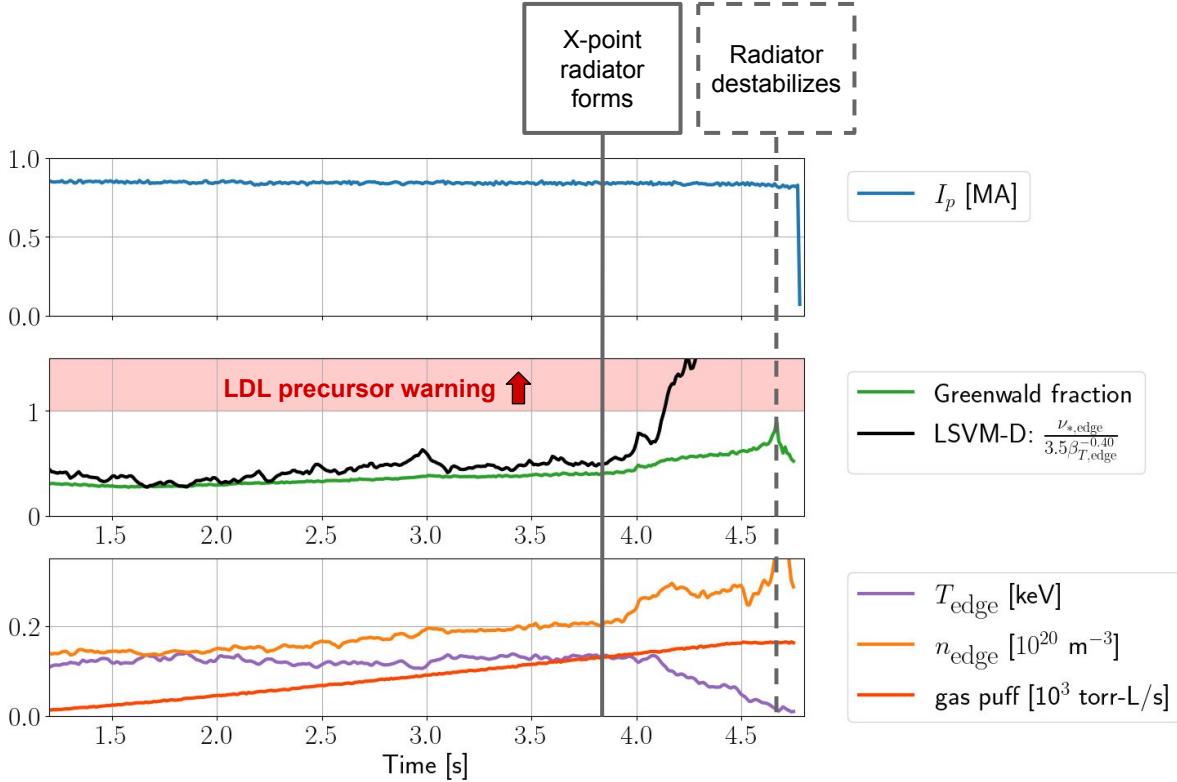


Figure 7: Time traces of DIII-D #191794, previously shown in Fig. 2, which ends in a major disruption. The middle panel shows the Greenwald fraction and LSVM-D metric calibrated to TPR = 95%.

parameters such as  $q$  and  $R_0$ . Naively using eq. 16 as a density limit indicator results in a prediction performance (AUC = 0.973, FPR = 19.7% @ TPR = 95%) significantly below the LSVM-derived boundaries, and similar to that of the Greenwald fraction.

#### 4.4. Example discharges

Here, we consider two example discharges to show how the Greenwald fraction and LSVM-D metric compare as LDL warning indicators. The LSVM-D metric has been calibrated for TPR = 95%.

Figure 7 shows a standard density limit discharge at DIII-D (previously illustrated in Fig. 2 to describe the labeling). As is typical for this device, the LDL occurs at a Greenwald fraction less than 1. This would therefore be a false negative if an LDL warning threshold of  $n/n_G = 1$  was used. By contrast, the LSVM-G correctly warns of the LDL, rising above unity about half a second before the radiator destabilizes. This long warning time would be useful for disruption avoidance, as it would provide the control system time to recover the discharge by reducing fueling or increasing heating power.

A failure case for the LSVM-D metric is shown in Fig. 8. This example is the most common failure case for Alcator C-Mod: transient H-modes with low heating

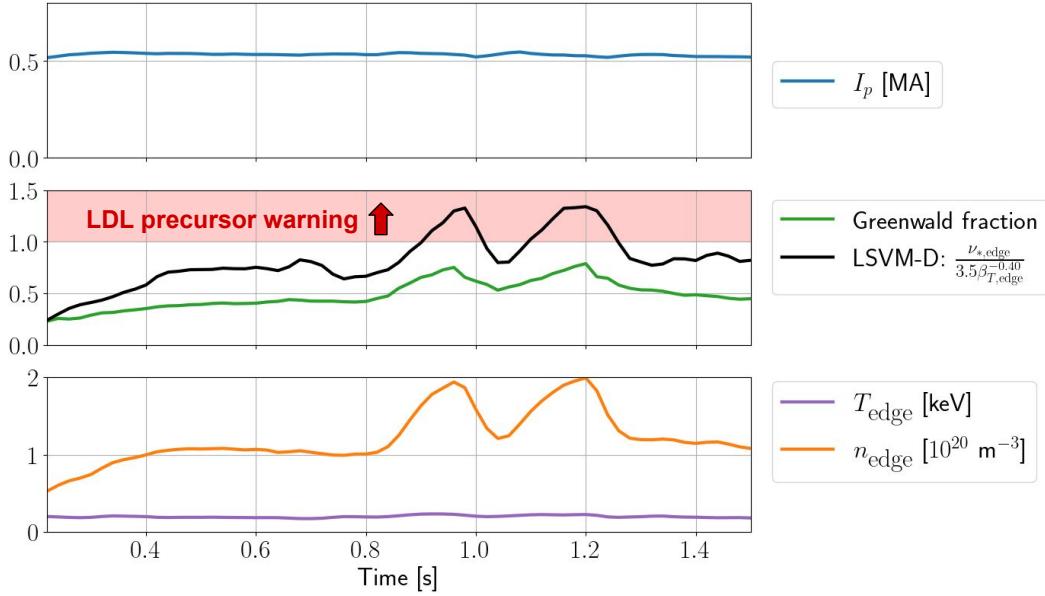


Figure 8: Time traces of non-disruptive C-Mod #1150806028, a low-auxiliary power discharge with two brief H-modes visible in the peaks in edge density around  $t = 0.9$ s and  $t = 1.2$ s.

power. This incorrect classification may be due to the presence of the H-mode pedestal changing the correlation between the “edge” density and temperature (as defined in this study) with the separatrix density and temperature. If indeed the separatrix conditions set the density limit, once might expect failures when this correlation is broken.

From another perspective, this might not necessarily be a considered a failure at instability prediction, as the brief H-modes are not stable; while an LDL instability does not occur, H-to-L back-transition instabilities occur during both the excursions above the stability boundary, restoring the plasma state below the threshold after each return to L-mode. In general, other false positives for the LSVM-D can also occur for discharges with low heating power, and discharges with non-disruptive MARFEs or X-point radiators.

#### 4.5. Comparison of data-driven models

In each of the three LDL prediction cases (global features, edge features, and dimensionless features), the LSVM achieves comparable LDL prediction performance with the NN and RF. The RF has the lowest FPR in each case, but the AUC is very similar to the LSVM. This demonstrates that a highly-parameterized ML architecture such as a NN or RF is not necessary for achieving high accuracy for predicting the LDL. NNs and RFs are well suited for problems where simple, analytic functions cannot describe the observed behavior; in this case, however, a power law appears to describe the LDL precursor boundary well.

By comparing the LSVM and linear regression results, it is also evident that the way one determines the power law is critical. The linear regression approach records notably worse performance in all cases compared to the LSVM, despite the fact that both search over the same set of features and utilize the same functional form. The large gulf between these models boils down to the fact that the LSVM is utilized as a classification algorithm that leverages information from both LDL and stable plasma states, while linear regression can only use information from LDL cases. The task at hand - distinguishing LDLs from stable plasma states - is a classification problem; there is valuable information in the plasma states that end in LDLs as well as those that do not. Simply fitting an expression to the density near the LDL is not appropriate for this task.

#### *4.6. Limitations*

We note that the number of discharges in our database from the five devices is not uniform, as shown in Tables 1 and 4. In particular, the large number of stable discharges from the C-Mod and DIII-D discharges give us good statistics for the FPR in the test set, but also means that the FPR is mostly determined by discharges from those two devices.

We also note there are strong correlations among  $a$ ,  $R_0$ , and  $B_T$  in our database (see Appendix A); it is therefore impossible to disentangle the independent causal effect of these three variables. Shaping variables  $\epsilon$ ,  $\kappa$ , and  $\delta$  are also not included in this analysis, as there is relatively little variance (standard deviation  $\leq 25\%$  of the mean) across the dataset. Additionally, there are no negative triangularity discharges in this study.

The effect of isotope mass and impurities on the density limit are also not captured in this study. The database only contains deuterium majority density limit discharges, does not include effective charge  $Z_{\text{eff}}$  as a parameter, and excludes discharges where the operators noted a major impurity injection. We note that the LSVM metrics achieve high accuracy across metal- and carbon-wall devices in this study, including shots in DIII-D with impurity seeding, but we also underline that these instability metrics should be applied to relatively clean, hydrogenic discharges.

#### *4.7. Potential applicability for real-time control of burning plasmas*

The LDL instability metrics identified in this study could be used for real-time disruption prediction and plasma control. The most challenging measurements needed for the LSVM-E and -D metrics are the edge density and temperature signals computed from Thomson scattering (TS). While TS systems have low sampling rates relative to other diagnostics, such as the magnetics, TS measurements are frequent relative to the energy and particle confinement times that set the evolution rate of the temperature and density. For example, ITER edge TS ( $r/a > 0.85$ ) will have a temporal resolution of

10ms and spatial resolution of 5mm, compared to several seconds of energy confinement time and a 2.8m plasma minor radius [44].

Estimating the LSVM-E and -D metrics in fusion power plants (FPPs) will be more challenging as TS will likely be unavailable; the large windows for gathering scattered photons conflict with tritium breeding requirements. In this case, other diagnostics (reflectometers, interferometers, ECE) would be needed to measure or infer the edge density and temperature.

In burning plasmas, however, edge collisionality will be very low due to the tremendous self-heating. Measuring the distance to the stability boundary in real time may only be necessary during the ramp-up and ramp-down phases.

## 5. Conclusion

In this multi-machine study, we leverage a manually labelled database of density limit disruption to identify  $\nu_{*,\text{edge}} \beta_{T,\text{edge}}^{0.40}$  as a reliable predictor of the LDL precursor phase. This instability limit achieves excellent prediction performance (AUC = 0.997, FPR = 2.3% @ TPR = 95%), significantly outperforming line-averaged Greenwald and edge Greenwald scalings. We are able to uncover this scaling by training an LSVM to identify the radiative precursor phase to the LDL. Other non-symbolic data-driven approaches, such as NNs and RFs achieve similar accuracy as the LSVM power law. By contrast, standard linear regression is unable to identify a highly reliable LDL instability metric.

The LSVM power law boundaries (Table 7) are reminiscent of the Greenwald limit in that they favors smaller devices with higher current. However, we achieve higher prediction accuracy by accounting for  $T_{\text{edge}}$  explicitly in LSVM-E or through dimensionless quantities that include edge temperature in LSVM-D. Despite the different set of features, the LSVM-E and -D metrics can be shown to be nearly identical. These scalings appears to be consistent with some theoretical models of the density limit, but additional measurements of impurities and turbulent fluctuations would be needed to confirm the association.

This study also demonstrates the utility of LSVMs for identifying stability boundaries for specific events such as the LDL. For the edge and dimensionless features case, the LSVM identifies a power law with comparable performance to the highly-parameterized NN and RF models. The LSVM power law also consistently outperforms the power law identified via linear regression. These results illustrate that for a specific instability and descriptive set of features, an LSVM can identify an accurate analytic stability boundary.

This analysis is somewhat limited by non-uniform number of discharges available across devices and correlations among some parameters (ex.  $a$ ,  $R_0$ , and  $B_T$ ). Future work will seek to address these limitations by increasing the number of non-disruptive discharges from underrepresented machines in the database (AUG, TCV), expanding the database to new devices (JET), adding data from uncommon scenarios (DIII-D negative triangularity), and potentially including devices with other shapes and aspect

ratios, such as spherical tokamaks.

We also discuss the potential applicability of the collisionality boundary as an instability metric for real-time control. Current experiments, such as DIII-D, and future experiments, such as ITER, have TS systems capable of measuring edge density and temperature at sufficiently high spatial and temporal resolution in real time. FPPs will face a more constrained sensing environment that preclude TS, but other diagnostics may be able to measure or infer the relevant parameters. Additionally, the low collisionality in the edge of burning plasmas may obviate the need for density limit avoidance during the flattop. We will explore applying this indicator for real-time density limit avoidance in future work.

## Acknowledgments

The authors would like to thank A. Hubbard, J. Hughes, N. Logan, and X. Chen for providing insightful feedback on this study; A. Miller for guidance in gathering C-Mod Thomson Scattering data; and M. Tobin for providing useful code for plotting. 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 Awards DE-FC02-04ER54698, DE-SC0014264.

This work has been carried out within the frame-work of the EUROfusion Consortium, via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion) and funded by the Swiss State Secretariat for Education, Research, and Innovation (SERI). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, the European Commission, or SERI. Neither the European Union nor the European Commission nor SERI can be held responsible for them.

This work is partially supported by the National Natural Science Foundation of China under Grant number 12005264 and the International Atomic Energy Agency under Research Contract number 26478.

Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## References

- [1] J Rand McNally. The ignition parameter  $tn\tau$  and the energy multiplication factor  $k$  for fusioning plasmas. *Nuclear Fusion*, 17(6):1273, 1977.
- [2] RJ Buttery, JM Park, JT McClenaghan, D Weisberg, J Canik, J Ferron, A Garofalo, CT Holcomb, J Leuer, PB Snyder, et al. The advanced tokamak path to a compact net electric fusion pilot plant. *Nuclear Fusion*, 61(4):046028, 2021.
- [3] SI Krasheninnikov, AS Kukushkin, and AA Pshenov. Divertor plasma detachment. *Physics of Plasmas*, 23(5), 2016.
- [4] Masanori Murakami, JD Callen, and LA Berry. Some observations on maximum densities in tokamak experiments. *Nuclear Fusion*, 16:347, 1976.
- [5] SJ Fielding, J Hugill, GM McCracken, JWM Paul, R Prentice, and PE Stott. High-density discharges with gettered torus walls in dite. *Nuclear Fusion*, 17:1382–1385, 1977.
- [6] Martin Greenwald. Density limits in toroidal plasmas. *Plasma Physics and Controlled Fusion*, 44(8):R27, 2002.
- [7] Martin Greenwald, JL Terry, SM Wolfe, S Ejima, MG Bell, SM Kaye, and GH Neilson. A new look at density limits in tokamaks. *Nuclear Fusion*, 28(12):2199, 1988.
- [8] M Shimada, DJ Campbell, V Mukhovatov, M Fujiwara, N Kirneva, K Lackner, MPVD Nagami, VD Pustovitov, N Uckan, J Wesley, et al. Overview and summary. *Nuclear Fusion*, 47(6):S1, 2007.
- [9] G Giruzzi, JF Artaud, Matteo Baruzzo, Tommaso Bolzonella, E Fable, L Garzotti, I Ivanova-Stanik, R Kemp, DB King, M Schneider, et al. Modelling of pulsed and steady-state demo scenarios. *Nuclear Fusion*, 55(7):073002, 2015.
- [10] BN Sorbom, J Ball, TR Palmer, FJ Mangiarotti, JM Sierchio, P Bonoli, C Kasten, DA Sutherland, HS Barnard, CB Haakonsen, et al. Arc: A compact, high-field, fusion nuclear science facility and demonstration power plant with demountable magnets. *Fusion Engineering and Design*, 100:378–405, 2015.
- [11] Andrew D Maris, Allen Wang, Cristina Rea, Robert Granetz, and Earl Marmar. The impact of disruptions on the economics of a tokamak power plant. *Fusion Science and Technology*, pages 1–17, 2023.
- [12] B LaBombard, RL Boivin, M Greenwald, J Hughes, B Lipschultz, D Mossessian, CS Pitcher, JL Terry, SJ Zweben, and Alcator Group. Particle transport in the scrape-off layer and its relationship to discharge density limit in alcator c-mod. *Physics of Plasmas*, 8(5):2107–2117, 2001.
- [13] B Lipschultz, B LaBombard, ES Marmar, MM Pickrell, JL Terry, R Watterson, and SM Wolfe. Marfe: an edge plasma phenomenon. *Nuclear Fusion*, 24(8):977, 1984.
- [14] Peng Shi, Ge Zhuang, K Gentle, Qiming Hu, Jie Chen, Qiang Li, Yang Liu, Li Gao, Xiaolong Zhang, Hai Liu, et al. First time observation of local current shrinkage during the marfe behavior on the j-text tokamak. *Nuclear Fusion*, 57(11):116052, 2017.
- [15] Xin Li, Shouxin Wang, Yuqi Chu, Hui Lian, Yinxian Jie, Rongjie Zhu, Yi Yuan, Liqing Xu, Tonghui Shi, Ang Ti, et al. Local current shrinkage induced by the marfe in l mode discharges on east tokamak. *AIP Advances*, 13(3), 2023.
- [16] U Stroth, M Bernert, D Brida, M Cavedon, R Dux, E Huett, T Lunt, O Pan, M Wischmeier, the ASDEX Upgrade Team, et al. Model for access and stability of the x-point radiator and the threshold for marfes in tokamak plasmas. *Nuclear Fusion*, 62(7):076008, 2022.
- [17] Paolo Zanca, F Sattin, DF Escande, and JET Contributors. A power-balance model of the density limit in fusion plasmas: application to the l-mode tokamak. *Nuclear fusion*, 59(12):126011, 2019.
- [18] BN Rogers, JF Drake, and A Zeiler. Phase space of tokamak edge turbulence, the l-h transition, and the formation of the edge pedestal. *Physical Review Letters*, 81(20):4396, 1998.
- [19] M Giacomin, A Pau, P Ricci, O Sauter, T Eich, JET Contributors, ASDEX Upgrade Team, et al. First-principles density limit scaling in tokamaks based on edge turbulent transport and

- implications for iter. *Physical Review Letters*, 128(18):185003, 2022.
- [20] Thomas Eich, Peter Manz, ASDEX Upgrade Team, et al. The separatrix operational space of asdex upgrade due to interchange-drift-alfvén turbulence. *Nuclear Fusion*, 61(8):086017, 2021.
  - [21] Rameswar Singh and PH Diamond. Zonal shear layer collapse and the power scaling of the density limit: old lh wine in new bottles. *Plasma Physics and Controlled Fusion*, 64(8):084004, 2022.
  - [22] Peter Manz, Thomas Eich, Ondrej Grover, ASDEX Upgrade Team, et al. The power dependence of the maximum achievable h-mode and (disruptive) l-mode separatrix density in asdex upgrade. *Nuclear Fusion*, 63(7):076026, 2023.
  - [23] A Gibson. Fusion relevant performance in jet. *Plasma Physics and Controlled Fusion*, 32(11):1083, 1990.
  - [24] Y Kamada, N Hosogane, R Yoshino, T Hirayama, and T Tsunematsu. Study of the density limit with pellet fuelling in jt-60. *Nuclear fusion*, 31(10):1827, 1991.
  - [25] A Stabler, K McCormick, V Mertens, ER Muller, J Neuhauser, H Niedermeyer, K-H Steuer, H Zohm, F Dollinger, A Eberhagen, et al. Density limit investigations on asdex. *Nuclear fusion*, 32(9):1557, 1992.
  - [26] TH Osborne, AW Leonard, MA Mahdavi, M Chu, ME Fenstermacher, R La Haye, G McKee, TW Petrie, E Doyle, G Staebler, et al. Gas puff fueled h-mode discharges with good energy confinement above the greenwald density limit on diii-d. *Physics of Plasmas*, 8(5):2017–2022, 2001.
  - [27] G Pucella, O DArcangelo, O Tudisco, F Belli, W Bin, A Botrugno, P Buratti, G Calabò, B Esposito, E Giovannozzi, et al. Analytical relation between peripheral and central density limit on ftu. *Plasma Physics and Controlled Fusion*, 59(8):085011, 2017.
  - [28] C Angioni, H Weisen, OJWF Kardaun, M Maslov, A Zabolotsky, C Fuchs, L Garzotti, C Giroud, B Kurzan, P Mantica, et al. Scaling of density peaking in h-mode plasmas based on a combined database of aug and jet observations. *Nuclear Fusion*, 47(9):1326, 2007.
  - [29] J Rapp, PC De Vries, FC Schüller, W Biel, R Jaspers, HR Koslowski, A Krämer-Flecken, A Kreter, M Lehnen, A Pospieszczyk, et al. Density limits in textor-94 auxiliary heated discharges. *Nuclear Fusion*, 39(6):765, 1999.
  - [30] Alexander Huber, S Brezinsek, M Groth, PC De Vries, V Riccardo, G Van Rooij, G Sergienko, G Arnoux, A Boboc, P Bilkova, et al. Impact of the iter-like wall on divertor detachment and on the density limit in the jet tokamak. *Journal of Nuclear Materials*, 438:S139–S147, 2013.
  - [31] M Bernert, T Eich, A Kallenbach, D Carralero, A Huber, PT Lang, S Potzel, F Reimold, J Schweinzer, E Viezzier, et al. The h-mode density limit in the full tungsten asdex upgrade tokamak. *Plasma Physics and Controlled Fusion*, 57(1):014038, 2014.
  - [32] M. Lehnen, P. Aleynikov, and B. Bazylev. Plasma Disruption Management in ITER. Presented at 2018 IAEA Fusion Energy Conference, Gandhinagar.
  - [33] Cristina Rea, RS Granetz, K Montes, Roy Alexander Tinguely, N Eidietis, Jeremy M Hanson, and B Sammuli. Disruption prediction investigations using machine learning tools on diii-d and alcator c-mod. *Plasma Physics and Controlled Fusion*, 60(8):084004, 2018.
  - [34] Kevin Joseph Montes, Cristina Rea, RS Granetz, Roy Alexander Tinguely, N Eidietis, OM Meneghini, DL Chen, Biao Shen, BJ Xiao, Keith Erickson, et al. Machine learning for disruption warnings on alcator c-mod, diii-d, and east. *Nuclear Fusion*, 59(9):096015, 2019.
  - [35] Wenhui Hu, Jilei Hou, Zhengping Luo, Yao Huang, Dalong Chen, Bingjia Xiao, Qiping Yuan, Yanmin Duan, Jiansheng Hu, Guizhong Zuo, et al. Prediction of multifaceted asymmetric radiation from the edge movement in density-limit disruptive plasmas on experimental advanced superconducting tokamak using random forest. *Chinese Physics B*, 32(7):075211, 2023.
  - [36] A Pau et al. A modern framework to support disruption studies: the eurofusion disruption database. Preprint: 2023 IAEA Fusion Energy Conference, London.
  - [37] Geert Verdoollaeghe, Stanley M Kaye, Clemente Angioni, Otto JWF Kardaun, Mikhail Maslov, Michele Romanelli, François Ryter, Knud Thomsen, JET Contributors, ASDEX Upgrade Team, et al. The updated itpa global h-mode confinement database: description and analysis. *Nuclear*

*Fusion*, 61(7):076006, 2021.

- [38] PC De Vries, G Pautasso, D Humphreys, M Lehnert, S Maruyama, JA Snipes, A Vergara, and L Zabeo. Requirements for triggering the iter disruption mitigation system. *Fusion Science and Technology*, 69(2):471–484, 2016.
- [39] G Duesing. First results of neutral beam heating on jet. *Plasma Physics and Controlled Fusion*, 28(9A):1429, 1986.
- [40] Peter C Stangeby et al. *The plasma boundary of magnetic fusion devices*, volume 224. Institute of Physics Pub. Philadelphia, Pennsylvania, 2000.
- [41] RJ Hajjar, PH Diamond, and MA Malkov. Dynamics of zonal shear collapse with hydrodynamic electrons. *Physics of Plasmas*, 25(6), 2018.
- [42] R Hong, GR Tynan, PH Diamond, L Nie, D Guo, T Long, R Ke, Y Wu, B Yuan, M Xu, et al. Edge shear flows and particle transport near the density limit of the hl-2a tokamak. *Nuclear Fusion*, 58(1):016041, 2017.
- [43] T Long, PH Diamond, R Ke, L Nie, M Xu, XY Zhang, BL Li, ZP Chen, X Xu, ZH Wang, et al. Enhanced particle transport events approaching the density limit of the j-text tokamak. *Nuclear Fusion*, 61(12):126066, 2021.
- [44] M Bassan, P Andrew, G Kurskiev, E Mukhin, T Hatae, G Vayakis, E Yatsuka, and M Walsh. Thomson scattering diagnostic systems in iter. *Journal of Instrumentation*, 11(01):C01052, 2016.

## Appendix A. Range of parameters in dataset

Table A1 shows the average value and standard deviation of macroscopic parameters of the five devices in the database of this study. All parameters come from experiment measurements or equilibrium reconstructions except for the major radius of DIII-D and EAST, which are set as constant values (note: the major and minor radius of Alcator C-Mod have a finite but small standard deviation). We also show the range of dimensionless edge parameters in Table A2. The Pearson correlation matrix of several parameters used in this study are shown in Fig. A1, including the binary LDL precursor label.

Table A1: Average value and standard deviation of macroscopic parameters for each device in the database.

Device	$n_e$ [10 <sup>20</sup> m <sup>-3</sup> ]	$I_p$ [MA]	$a$ [m]	$R_0$ [m]	$B_T$ [T]	$P_{in}$ [MW]
AUG	$0.68 \pm 0.16$	$0.72 \pm 0.13$	$0.50 \pm 0.01$	$1.60 \pm 0.01$	$2.41 \pm 0.23$	$6.61 \pm 1.66$
C-Mod	$1.45 \pm 0.68$	$0.83 \pm 0.18$	$0.22 \pm 0.00$	$0.68 \pm 0.00$	$5.44 \pm 0.81$	$1.73 \pm 1.05$
DIII-D	$0.44 \pm 0.17$	$1.04 \pm 0.21$	$0.59 \pm 0.02$	$1.67 \pm 0.00$	$1.94 \pm 0.17$	$5.87 \pm 3.17$
EAST	$0.33 \pm 0.09$	$0.35 \pm 0.06$	$0.44 \pm 0.01$	$1.83 \pm 0.00$	$2.43 \pm 0.00$	$3.29 \pm 1.77$
TCV	$0.47 \pm 0.20$	$0.18 \pm 0.07$	$0.23 \pm 0.01$	$0.88 \pm 0.01$	$1.42 \pm 0.03$	$0.57 \pm 0.49$

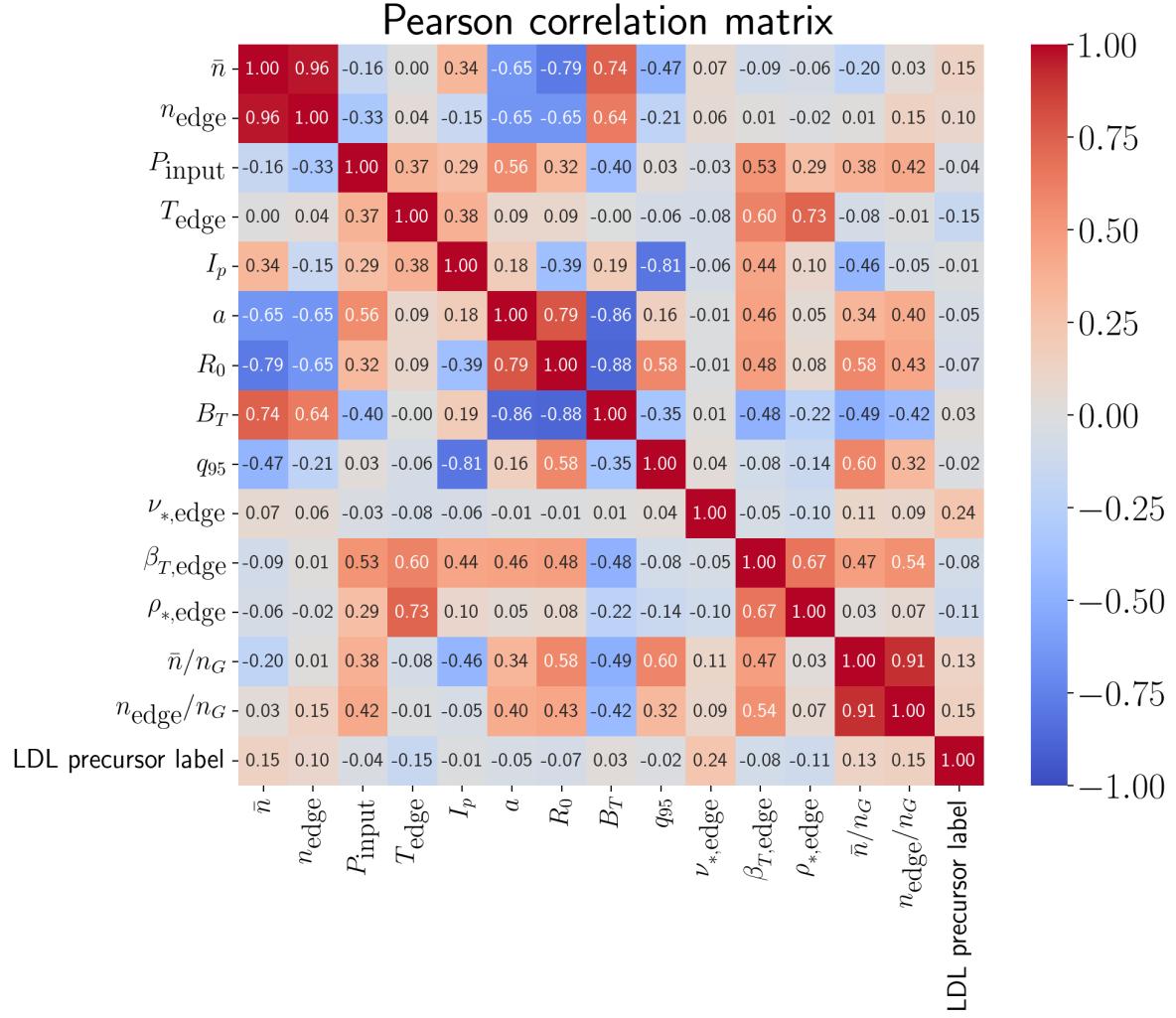


Figure A1: Correlation of several parameters in the dataset utilized in this study, including the binary LDL precursor label.

Table A2: Average value and standard deviation of several dimensionless parameters in the edge of the plasma for each device in the database.

Device	$q_{95}$	$\nu_{*,edge}$	$\beta_{T,edge}$ [%]	$\rho_{edge}^*$ [%]
AUG	$6.00 \pm 0.97$	$16.51 \pm 73.24$	$0.42 \pm 0.17$	$0.37 \pm 0.07$
C-Mod	$4.53 \pm 0.96$	$19.97 \pm 272.04$	$0.16 \pm 0.16$	$0.38 \pm 0.10$
DIII-D	$5.08 \pm 1.48$	$3.69 \pm 18.74$	$0.67 \pm 0.39$	$0.51 \pm 0.14$
EAST	$7.97 \pm 1.29$	N/A	N/A	N/A
TCV	$4.85 \pm 1.19$	$52.53 \pm 72.99$	$0.19 \pm 0.17$	$0.77 \pm 0.23$

Table B1: The hyperparameter ranges utilized for training the neural network.

Hyperparameter	Range or values	Sampling distribution
Learning rate	0.001 - 0.2	log uniform
Batch size	32, 64	uniform
# epochs	10 - 800	log uniform
# layers	1 - 5	uniform
# hidden units	16, 32, 64, 128	uniform
drop out proportion	0 - 0.5	uniform
activation function	relu, sigmoid	uniform

Table B2: The hyperparameter ranges utilized for training the random forest.

Hyperparameter	Range or values	Sampling distribution
# estimators	10, 30, 100	uniform
max # features	3 - 8	uniform
max depth	3, 5, 8	uniform
min # samples per split	2, 5, 10	uniform
min # samples per leaf	1, 2, 5, 10	uniform

Table B3: The hyperparameter ranges utilized for training the LSVM.

Hyperparameter	Range or values	Sampling distribution
C	0.1, 1, 10	uniform

## Appendix B. Hyperparameter ranges

The neural network, random forest, and LSVM were trained over a range of hyperparameters reported in Tables B1, B2, and B3. The NN and RF hyperparameters were sampled randomly, while the LSVM hyperparameter was evaluated in a grid scan. Several sample-weighting methods were also explored, with minimal effect on the final models.

## Appendix C. Generalizing to an unseen device

As long as fusion remains an experimental science, data-driven disruption predictors must be robust to “domain shifts,” i.e. differences between the training set and the cases observed during deployment. The best disruption prediction performances reported in the literature often come from highly expressive machine learning architectures, such as NNs and RFs, which can be especially vulnerable to domain shifts. Given the potentially catastrophic consequences of disruptions during a full-power discharge on ITER [32], robustness to domain shifts is a critical question.

Here, we consider an example of a domain shift: training on all AUG, C-Mod, and TCV discharges and then testing on DIII-D discharges. We utilize the dimensionless variables as in section 3.3, and therefore EAST is excluded due to lack of edge density

Table C1: The test set performance of LDL prediction for models trained on the edge features of AUG, C-Mod, and TCV but evaluated on DIII-D. The performance of the Greenwald fraction and Edge Greenwald fraction is also reported.

Model	Analytic boundary	AUC	FPR @ TPR = 95%
NN	N/A	0.987	3.3%
RF	N/A	<b>0.995</b>	<b>1.5%</b>
LSVM	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.41}$	0.992	1.6%
Lin. Reg.	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-1.06}$	0.974	9.6%
Greenwald	$\bar{n}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.847	53.7%
Edge Greenwald	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.705	82.6%

and temperature measurements. DIII-D was chosen for the test set because it has the largest current, major radius, and minor radius of devices in the database that includes edge measurements.

The results are shown in Table C1. The RF achieved the highest performance of all, with the LSVM close behind. The LSVM instability metric here is nearly identical to the one derived when the training set included DIII-D data (Section 3.3). Despite the domain shift, all data-driven models have better performance than the Greenwald scaling.

#### Appendix D. Assessing Giacomin-Ricci scaling [19] as a disruption predictor on AUG, DIII-D, and TCV

Several theoretical models for the density limit, such as in Ref. [19], offer compelling explanations of the density limit. This first-principles scaling is not explicitly intended for providing a warning to the density limit, but it could be used for this purpose. We therefore evaluate the scaling for the maximum density from Ref. [19] to see how it fairs as a LDL predictor.

To do so, we must rely on only AUG, DIII-D, and TCV, where we have consistent measurements of the power through the SOL. Additionally, we note that the scaling in [19] estimates the maximum density 'in the proximity of the separatrix,' not the 'edge' (as defined both here and in [19] as the average TS measurement between  $\rho = 0.85$  and  $\rho = 0.95$ ). Just as in [19], however, we will proceed with utilizing the edge density in the absence of reliable measurements at the separatrix. In light of this, we emphasize that we are not attempting a complete validation of this model in this exercise. Finally, we note that our analysis overlaps in terms of AUG and TCV, but our study has data from DIII-D instead of JET and includes stable discharges to quantify the accuracy of the metric for LDL prediction.

In Table D1, we show the LDL prediction accuracy for a NN, RF, LSVM, linear regression model, the Greenwald fraction, the edge Greenwald fraction, and

Table D1: The test set performance of LDL prediction for models trained and tested on dimensionless features for AUG, TCV, and DIII-D to compare with the Giacomin-Ricci scaling [19]. The performance of the Greenwald fraction and edge Greenwald fraction is also reported.

Model	Analytic boundary	AUC	FPR @ TPR = 95%
NN	N/A	0.984	6.4%
RF	N/A	0.985	<b>3.0%</b>
LSVM	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-0.41}$	<b>0.992</b>	3.5%
Lin. Reg.	$\nu_{*,\text{edge}}^{\text{limit}} \sim \beta_{T,\text{edge}}^{-1.14}$	0.976	11.4%
Giacomin	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p^{22/21} P_{SO_L}^{10/21} A^{1/6} R_0^{1/42}}{a^{79/42} B_T^{8/21} (1+\kappa^2)^{1/3}}$	0.915	22.8%
Greenwald	$\bar{n}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.896	44.1%
Edge Greenwald	$n_{\text{edge}}^{\text{limit}} \sim \frac{I_p}{\pi a^2}$	0.745	79.2%

the Giacomin-Ricci scaling. We see that the Giacomin-Ricci scaling achieves higher AUC and lower FPR @ TPR = 95% than the Greenwald fractions. Compared with the data-driven models, however, the Giacomin-Ricci scaling has a significantly lower performance. The LSVM boundary achieves the highest AUC and nearly matches the RF for lowest FPR @ TPR = 95%. The LSVM boundary is nearly the same as the case presented earlier in section 3.3.

Again, we emphasize that this is not an attempted validation of the Giacomin-Ricci scaling, as we do not utilize the density at the separatrix to conduct this analysis. Our focus is to analyze the scaling as a method of forecasting the LDL given readily available measurements. On this count, it improves upon the Greenwald limit, but is not as reliable as the LSVM-derived metric.