

Interpretable logistic regression for predicting 1/1 kink-driven regular complete sawtooth phenomenon in EAST tokamak plasmas

Haipeng Quan(全海鹏), Liqing Xu(徐立清), Chaowei Mai(麦晁玮), Yan Chao(晁燕), Yuhang Wang(王宇航), and EAST Team

Citation: Chin. Phys. B, 2026, 35 (**4**): 045203. DOI: 10.1088/1674-1056/ae37f5
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

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

Effects of coil structure and electromagnetic shielding on plasma distribution and uniformity in large-area radio-frequency inductively coupled plasmas

Cheng Xin(辛程), Xiang-Yun Lyu(吕翔云), Si-Yu Xing(邢思雨), Yu-Ru Zhang(张钰如), Tao Liu(刘涛), Wei-Ping Le(乐卫平), Fei Gao(高飞), and You-Nian Wang(王友年)
Chin. Phys. B, 2026, 35 (**2**): 025201. DOI: 10.1088/1674-1056/ae2113

Experimental studies of H_2/Ar plasma in a cylindrical inductive discharge with an expansion region

Shi-Bo Li(李世博), Si-Yu Xing(邢思雨), Fei Gao(高飞), and You-Nian Wang(王友年)
Chin. Phys. B, 2024, 33 (**10**): 105201. DOI: 10.1088/1674-1056/ad6250

Global dust density in two-dimensional complex plasma

Yi-Zhen Zhao(赵逸真), Song-Fen Liu(刘松芬), Wei Kong(孔伟), and Fang Yang(杨芳)
Chin. Phys. B, 2024, 33 (**6**): 065201. DOI: 10.1088/1674-1056/ad2a76

Development of a monochromatic crystal backlight imager for the recent double-cone ignition experiments

Chenglong Zhang(张成龙), Yihang Zhang(张翌航), Xiaohui Yuan(远晓辉), Zhe Zhang(张喆), Miaohua Xu(徐妙华), Yu Dai(戴羽), Yufeng Dong(董玉峰), Haochen Gu(谷昊琛), Zhengdong Liu(刘正东), Xu Zhao(赵旭), Yutong Li(李玉同), Yingjun Li(李英骏), Jianqiang Zhu(朱健强), and Jie Zhang(张杰)
Chin. Phys. B, 2024, 33 (**2**): 025201. DOI: 10.1088/1674-1056/ad1091

Ignition dynamics of radio frequency discharge in atmospheric pressure cascade glow discharge

Ya-Rong Zhang(张亚容), Qian-Han Han(韩乾翰), Jun-Lin Fang(方骏林), Ying Guo(郭颖), and Jian-Jun Shi(石建军)
Chin. Phys. B, 2023, 32 (**2**): 025201. DOI: 10.1088/1674-1056/ac673d

---

Interpretable logistic regression for predicting 1/1 kink-driven regular complete sawtooth phenomenon in EAST tokamak plasmas

Haipeng Quan(全海鹏)[1,2,†], Liqing Xu(徐立清)[2], Chaowei Mai(麦晁玮)[3], Yan Chao(晁燕)[2,‡], Yuhang Wang(王宇航)[4], and EAST Team[2]

[1] *School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan* 232001, *China*
[2] *Plasma Institute, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei* 230031, *China*
[3] *Guangdong Ocean University, School of Electronics and Information Engineering, Zhanjiang* 524088, *China*
[4] *School of Physical Science and Information Technology, Anhui University, Hefei* 230601, *China*

(Received 20 October 2025; revised manuscript received 12 January 2026; accepted manuscript online 14 January 2026)

Sawtooth phenomena are a central topic in tokamak fusion research; nevertheless, different sawtooth classes differ markedly in underlying physics, statistical abundance and diagnostic definition. This paper focuses exclusively on regular complete sawtooth — ideal, 1/1 internal-kink-driven events with full magnetic reconnection — which are the most frequent, unambiguously identifiable and theoretically best characterized (hereafter, all references to "sawtooth" denote this specific class). Compound crashes, partial-reconnection events and fast-ion-induced giant sawtooth are deliberately excluded because of their limited data volume and still-contested classification criteria, the inclusion of which would introduce intolerable label noise and theoretical ambiguity. To this aim, we propose a machine-learning-based temporal binary-classification framework that converts multi-diagnostic, high-resolution signals into precise labels distinguishing "oscillation" from "quiet" windows for this specific category, thereby supplying critical timing information for active control. A large-scale database covering 124 discharges and hundreds of millions of samples was constructed, and three representative algorithms — logistic regression, decision tree and random forest — were trained and compared. Among them, logistic regression achieved the best and most robust performance, reaching 95% accuracy on an independent test set and significantly outperforming the other models. Furthermore, shapley additive explanations (SHAP) was innovatively employed to quantify the contribution magnitude and direction of key physical features to the onset of regular 1/1 sawtooth, substantially enhancing model interpretability and physical fidelity. The study provides an efficient and robust predictor for the active intervals of ordinary 1/1 sawtooth; the uncovered correlations between physical drivers and sawtooth behavior lay a solid foundation for deepening the understanding of regular sawtooth evolution and for optimizing control strategies, thereby holding significant promise for improving the operational stability of fusion plasmas.

Keywords: sawtooth prediction, logistic regression, SHAP, interpretative model

PACS: 52.35.Py, 52.55.Fa, 52.30.–q, 07.05.Mh

DOI: 10.1088/1674-1056/ae37f5 **CSTR:** 32038.14.CPB.ae37f5

1. Introduction

Sawtooth oscillations represent a ubiquitous and critical magnetohydrodynamic (MHD) instability in tokamak discharges, routinely observed in devices such as EAST ($q_{\min} < 1$, easily observed). These oscillations manifest as periodic relaxations of the core electron temperature and density. When the central pressure exceeds a certain threshold, an $m = 1/n = 1$ magnetic island forms and triggers magnetic reconnection, leading to abrupt drops in temperature and density along with rapid energy release. Such relaxation processes not only degrade confinement performance but may also excite secondary instabilities, jeopardizing device safety.[1,2] To suppress sawtooth, EAST has implemented various active control techniques, including neutral beam injection (NBI),[3–6] electron cyclotron resonance heating and current drive (ECRH/ECCD),[7–9] ion cyclotron resonance frequency (ICRF) heating,[10,11] and resonant magnetic perturbations (RMP).[12] However, achieving effective real-time intervention hinges on the ability to predict the sawtooth time window Δt_{saw} (i.e., from the onset time t_{onset} to the crash time t_{crash}), rather than merely detecting the crash event itself. Existing research has predominantly focused on single-point crash identification, for instance, Gude *et al*. utilized multi-channel SXR signals to distinguish crashes from noise;[13] Van den Brand *et al*. employed an interacting multiple model (IMM) filter;[14] and Isei *et al*. applied neural networks to locate the crash time.[15] These methods generally fail to reliably capture the onset time t_{onset}, let alone identify the full sawtooth evolution window. Implementing control only after a crash occurs cannot prevent early energy loss and is unlikely to meet the operational requirements of future long-

†Corresponding author. E-mail: haipeng.quan@ipp.ac.cn
‡Corresponding author. E-mail: chaoyan@ipp.ac.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

pulse, high-performance fusion reactors. Therefore, we propose an interpretable machine learning framework that integrates logistic regression[16–20] with Shapley additive explanations (SHAP)[21,22] to achieve end-to-end identification of sawtooth-active intervals (hereafter referred to as LR–SHAP). The innovations of this framework include: (i) constructing a logistic regression-based binary classifier to directly label "sawtooth-active" and "quiet" phases, and (ii) introducing a transferability metric based on SHAP value distributions to quantify the physical mapping between diagnostic features and interval states, thereby enhancing the model's physical interpretability and generalization capability.

The paper is structured as follows. Section 1 reviews the relevant scientific background and elaborates on the motivation for constructing the LR-SHAP framework. Section 2 describes the database construction process, model architecture, and training and testing pipeline. Section 3 presents the model's prediction performance and interprets the results using both global and single-point SHAP analysis. Section 4 summarizes the main research findings and outlines directions for future work.

2. Materials and methods

2.1. Soft-x-ray

At present, sawtooth identification on EAST still relies mainly on visual inspection of soft-x-ray (SXR) or electron cyclotron emission (ECE) traces.[23–25] Among the available diagnostics, the SXR array is nevertheless the preferred and most robust tool for such studies due to the following reasons. (i) Its signal is a line-integrated measure of electron pressure and therefore directly sensitive to the core physics driving the instability. (ii) It remains reliable even at high density, free of the cut-off limitations that frequently affect ECE. (iii) The simultaneous recording of up to 46 radial channels provides an unparalleled spatio-temporal pattern for machine-learning-based recognition.[26] Figure 1 shows a representative discharge exhibiting clear sawtooth; the enlarged window demonstrates that the characteristic slow-rise/fast-decline shape is already evident in the raw SXR trace. Figure 2 illustrates the synchronized evolution of the core and edge SXR channels around a sawtooth crash during an EAST discharge. In Fig. 2(a), the core (red) and edge (blue) SXR signals are temporally aligned, displaying the characteristic anti-correlation during a sawtooth cycle: the crash is marked by a rapid drop in the core emission, accompanied by a nearly simultaneous rise in the edge emission. This inversion reflects the inward propagation of the cooling front and the outward transport of hot plasma following the internal reconnection event. Figure 2(b) magnifies the 2.85–3.00 s window and resolves several complete saw-tooth cycles, unambiguously demonstrating their periodicity. Figure 2(c) further expands the temporal axis to the 2.87–2.90 s window, resolving a single complete sawtooth cycle that clearly exhibits three sequential phases: a slow ramp, a precursory serration, and a final rapid collapse. In subsequent steps we will also use these SXR-derived features to automatically label the sawtooth-oscillation intervals.

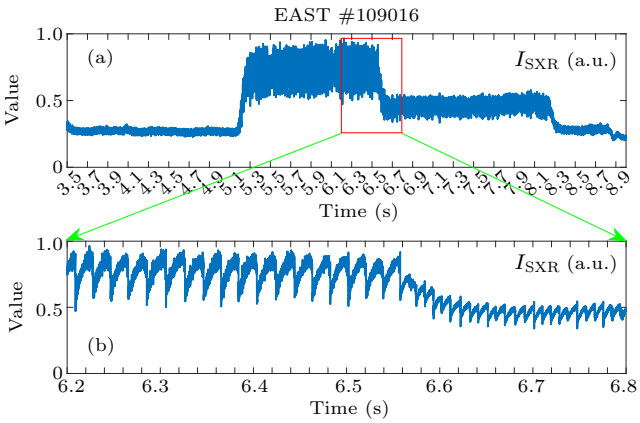


Fig. 1. Original SXR trace (upper) and a 600 ms magnified window (lower) for EAST shot #109016. The detector continuously samples the SXR emission throughout the discharge at 250 kHz. The abscissa is time t (ms) and the ordinate is the normalized intensity I/I_0. The global waveform consists of periodic 1/1 kink-driven regular compete sawtooth — slow rise followed by fast collapse — superimposed on a gently varying background. The expanded view resolves 4 μs·pt^{-1}, allowing the slope of the slow ramp, the steepness of the collapse edge and the small-scale oscillations embedded in the brief post-crash tail to be clearly distinguished.

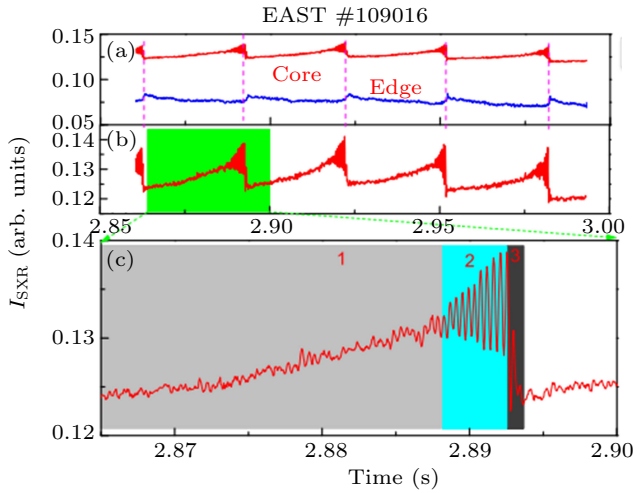


Fig. 2. Multi-channel SXR measurements for EAST shot #109016. (a) Comparison between core and edge signals: the crash first appears in the core, followed by an inverted jump in the edge, illustrating the characteristic "inversion" phase of 1/1 kink-driven regular complete sawtooth. (b) Four consecutive regular 1/1 sawtooth periods (core channel). (c) Expanded view of one period: ramp phase (gray), 1/1 sawtooth-precursor oscillation (blue), and collapse phase (black), clearly resolving the three stages of single-peak, full-reconnection event.

2.2. Dataset

Data utilized in this study originate from standard experimental discharges of the EAST tokamak. Multiple diagnostic systems acquire signals across different measurement windows; dedicated software extracts the raw experimental dataset. To construct the analysis feature set, resampling (with

SXR as the temporal baseline), temporal alignment, normalization and standardization are applied to all channels relevant to sawtooth oscillations.

In the preliminary feature-selection stage, dozens of candidate signals potentially related to sawtooth behavior are selected based on domain experience. A Pearson correlation coefficient matrix (Fig. 3) eliminates redundancy with a threshold of $|r| = 0.7$; for any pair with $|r| \geq 0.7$, only one channel is retained. This threshold is widely used in the fusion literature[27,28] to balance information loss against dimension reduction.

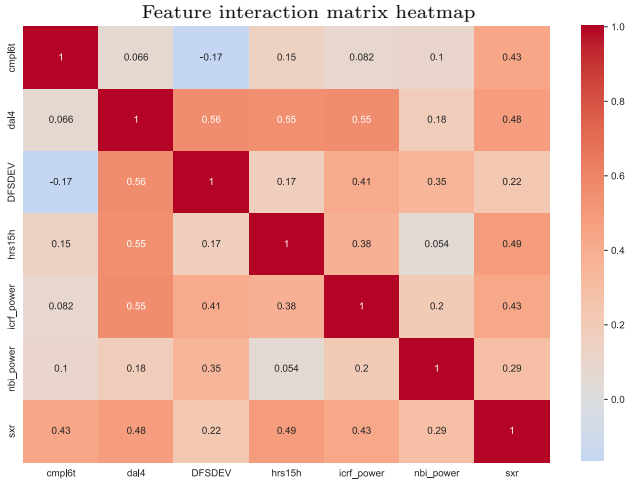


Fig. 3. Pearson correlation coefficient matrix. The displayed matrix contains the filtered coefficients, with each element representing the linear correlation between a pair of signals; higher values indicate greater redundancy. This map is therefore used to eliminate highly correlated features, achieving dimensionality reduction and a simplified dataset.

Channels surviving $|r| < 0.7$ are further screened according to the following physics criteria:

(i) Direct relevance to sawtooth crash physics: SXR (core MHD activity), magnetic probes (field perturbations), electron temperature/density (profile evolution).

(ii) Relevance to heating/particle sources: total NBI power, total ICRF power. (fast-ion drive), deuterium signal (fueling).

This procedure yields a 7-dimensional feature set, as detailed in Table 1.

Table 1. List of featured signal names and related information.

Feature	Sampling rate	Description
sxr	250 kHz	soft x ray
Hrs_15 h	1000 kHz	electron temperature
DFSDEV	275 Hz	electron density
dal4	250 kHz	deuterium signal
cmpl6t	200 kHz	magnetic probe
nbi_power	5 kHz	total NBI power
icrf_power	5 kHz	total ICRF power

For model training, 124 discharge events exhibiting clear sawtooth behavior are carefully selected from EAST experimental campaigns. A sawtooth-crash-moment identification algorithm based on dual-sliding-channels and SVD[29] first obtains approximate start/end intervals of sawtooth oscillations. These intervals are further refined through manual review and are finally assigned binary labels automatically by the program. This stringent selection and labeling process ensures a high-quality dataset and provides a solid foundation for subsequent analysis and model training/testing.

2.3. Logistic regression and SHAP (LR-SHAP)

Logistic regression models the probability of a binary outcome using the sigmoid function (the inverse log-odds transformation), producing outputs bounded between 0 and 1. This method, which dates back to Verhulst,[30] is both statistically well-founded and computationally efficient. While the signs of the coefficients indicate the direction of feature effects, they offer only limited insight into individual feature contributions. To address this limitation, we combine an L2-regularized logistic classifier (implemented with scikit-learn v1.3) with SHAP to quantitatively assess the magnitude and direction of each predictor's influence on the "sawtooth-active" probability. The complete workflow — encompassing data preparation, model training, performance evaluation, and interpretability analysis — is summarized in Fig. 4.

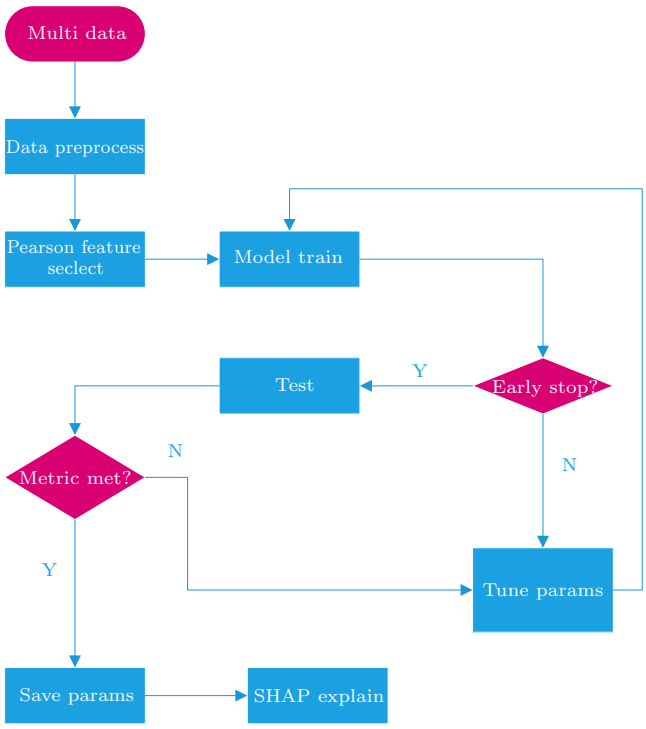


Fig. 4. LR-SHAP sawtooth-prediction model flowchart. Data preprocessing, feature selection, model training and testing, and final SHAP interpretation of the optimal model.

2.4. Model training and testing

After stratified sampling of the "sawtooth/non-sawtooth" classes from the 2020–2024 dataset (124 discharges, $\approx 2.25 \times 10^8$ instances, original ratio non-sawtooth : sawtooth $> 10{:}1$), 1% ($\approx 2.25 \times 10^6$) was reserved for training and 0.2% ($\approx$

4.5×10^5) for validation; 6.75×10^6 instances randomly drawn from the remaining 98.8% served as an independent test set with a stratified ratio of non-sawtooth : sawtooth $\approx 2:1$, unseen during training. The training set was balanced to approximately 1:1 to avoid class-imbalance bias. A distributed stochastic-gradient-descent logistic-regression model with L2 regularization ($C = 1.0$) was trained for up to 15 epochs (mini-batch size 2.56×10^5, initial learning rate 0.1); convergence was monitored via binary-cross-entropy loss on the validation set, with early stopping triggered if the loss reduction remained below 1×10^{-4} for three consecutive epochs. The final weight vector $w \in \mathbb{R}^7$ and bias b were exported for on-line inference and subsequent SHAP analysis.

3. Result

3.1. Comparative experiment

Under a fixed training/validation/test split, the proposed LR-SHAP framework was systematically benchmarked against two widely used baseline classifiers — decision tree (DT)[31] and random forest (RF).[32] Table 2 summarizes the comprehensive evaluation metrics: LR-SHAP attained an accuracy of 95%, a macro-averaged F1-score of 0.92, and balanced per-class recall values of 0.95 and 0.94 for the negative and positive categories, respectively, all of which surpass the corresponding scores of DT (accuracy 77.0%, F1 0.77, recalls 0.79 and 0.74) and RF (accuracy 84.5%, F1 0.84, recalls 0.92 and 0.76). The substantial improvements in both overall and class-specific performance indicators demonstrate the superior discriminative capacity and robustness of the LR-SHAP approach. Comparative experiments show that both decision trees and random forests lag behind logistic regression (LR) in performance, and the reasons can be summarized as follows:

(1) Over-fitting trap of complex models

DT chases training noise and shot-specific artifacts at deep splits, causing a large validation-test gap. RF still splits to minimal leaves despite bagging, thereby capturing high-dimensional sparse noise and degrading generalization.

(2) Robust advantage of logistic regression

The L_2 regularization coupled with a linear hypothesis naturally limits model capacity, conferring superior robustness to high-dimensional, sparse-feature data.

Consequently, all subsequent analyses and interpretations were carried out exclusively with the LR-SHAP model.

Table 2. Experimental comparison results.

	LR-SHAP	Decision tree	Random forest
Accuracy	95%	76.97%	84.5%
F1-score	0.92	0.77	0.84
Recall-0	0.95	0.79	0.92
Recall-1	0.94	0.74	0.76

3.2. Experimental results and error analysis

On the test set, the trained LR-SHAP model achieved 95% accuracy, macro-F1 = 0.94 and per-class recall ≥ 0.94, demonstrating good generalization to long-pulse discharge conditions unseen during training. Specific results are summarized in Fig. 5: non-sawtooth (quiet) accuracy = 97%, sawtooth accuracy = 90%, non-sawtooth recall = 0.95, sawtooth recall = 0.94, macro-F1 = 0.94.

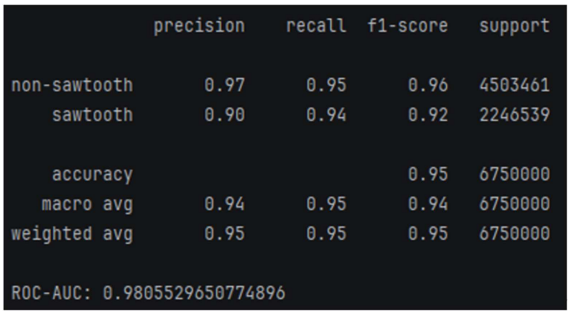


Fig. 5. The LR-SHAP model's classification performance on the independent test set (6.75×10^6 windows, quiet: sawtooth $\approx 2:1$), displaying prediction precision for each class. The macro-F1 is unaffected by sample imbalance, validating the model's robustness and calibration under a real-world imbalanced scenario.

Analysis of class-wise prediction accuracy disparities is given as follows.

(1) **Boundary blurring** "Precursor-recovery" boundary windows appear ambiguous in SXR, leading to boundary misclassifications mainly counted as FP for sawtooth, pulling its precision down to 90%.

(2) **Recall-first strategy** The 1:1 training set plus F1-maximisation allows more FP, ensuring sawtooth recall ≥ 0.94 (only 6% crashes missed), satisfying the engineering requirement "better false alarm than miss".

(3) **Macro-F1 balance** The 97% versus 90% precision difference is consolidated by macro-F1 = 0.94, remains uninfluenced by sample imbalance, directly reveals the minority-class recall, and thereby confirms the model's robustness under realistically imbalanced conditions.

Model stability was assessed on the same test set. Table 3 shows LR-SHAP continuous-probability errors of MSE = 0.0405, RMSE = 0.2012, MAE = 0.0800; MSE < 0.05 indicates minimal squared deviation between predicted probabilities and true binary labels, demonstrating good calibration. The confusion matrix in Fig. 6 corroborates these results with TP $\approx 2.1 \times 10^6$, TN $\approx 4.3 \times 10^6$, FN $\approx 2.3 \times 10^5$, FP $\approx 1.3 \times 10^5$, and per-class recall ≥ 0.94. These low error values and the tightly concentrated probability distribution jointly indicate that the LR-SHAP framework possesses both high discriminative ability and reliable probabilistic output.

Table 3. LR-SHAP error-analysis table.

	MSE	RMSE	MAE
LR-SHAP	0.0405	0.2012	0.0800

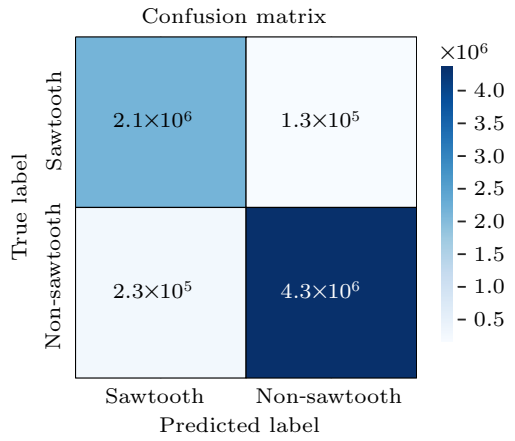


Fig. 6. Confusion matrix of the classification results based on the logistic-regression model, showing the predictive performance for the two labels "sawtooth" and "non-sawtooth"; diagonal elements indicate the numbers of correctly classified samples, while off-diagonal elements reflect misclassifications, with the overall result used to evaluate the model's classification accuracy and misclassification tendency.

4. Model interpretation — SHAP

4.1. Global interpretability

Traditional logistic-regression coefficients merely provide a global average slope, whereas SHAP yields instance-level attributions, captures nonlinear couplings, and mitigates multicollinearity-induced distortions. We therefore adopt SHAP-based global importance as our primary interpretative metric. Figure 7 shows that 1/1 kink-driven regular complete sawtooth onset is governed chiefly by two features: SXR intensity and total NBI power. Briefly, low SXR signals a stable state, whereas high NBI power flags an elevated trigger probability. Moreover, both high SXR and high NBI power correlate with elevated central pressure; under comparable confinement, higher pressure steepens the pressure gradient at the $q = 1$ surface and thereby intensifies regular 1/1 sawtooth activity. This stabilizing effect is dynamically coupled to the energy balance of the plasma core and is most clearly interpreted by distinguishing two distinct phases of the sawtooth cycle. Throughout we employ the relation

$$I_{\mathrm{SXR}} \propto n_{\mathrm{e}}^2 T_{\mathrm{e}}^{1/2} Z_{\mathrm{eff}}\,. \tag{1}$$

In our database, the discharge-averaged Z_{eff} varies by $\leq 15\%$, so I_{SXR} is dominated by the core electron pressure $n_{\mathrm{e}}T_{\mathrm{e}}$. A negative SHAP value therefore flags periods when this pressure is too low to supply sufficient free energy to the ideal internal kink.

The two canonical phases implied by a negative SHAP value of the SXR signal.

(1) **Post-crash recovery phase** A direct marker of core stability immediately after a sawtooth crash, energy and particles are expelled from the core, producing simultaneous drops in n_{e} and T_{e}. According to Eq. (1), the combined fall in n_{e}^2 and $T_{\mathrm{e}}^{1/2}$ causes a sharp decrease in the SXR intensity. This low-SXR "negative perturbation" indicates a stable configuration for three reasons: (i) the reduced pressure gradient weakens the drive for the $m/n = 1/1$ kink-tearing mode; (ii) the relaxed current profile increases the magnetic shear near the $q = 1$ surface, temporarily raising the instability threshold; (iii) fast-ion contributions (e.g., δW_{K}) are diminished before the core is reheated, further enhancing stability.

The model therefore learns that a strong negative SXR perturbation is a robust signature of the recent crash and the ensuing stable recovery interval, statistically implying a longer wait-time to the next crash and receiving a large negative SHAP value.

(2) **Late energy-accumulation phase** An indicator of incomplete core recovery later in the cycle, during the NBI-driven recovery, a relative negative SXR perturbation does not reflect a globally cold plasma but rather a transient setback in core-energy accumulation. Possible causes include: (i) small-scale MHD events or transient transport losses that locally reduce T_{e}; (ii) transient impurity influx or increased core radiation (higher local Z_{eff}) that enhances radiative losses in Eq. (1). Such a perturbation acts as a "brake" on the approach to the instability boundary; the model interprets it as a stabilizing factor that postpones the next crash.

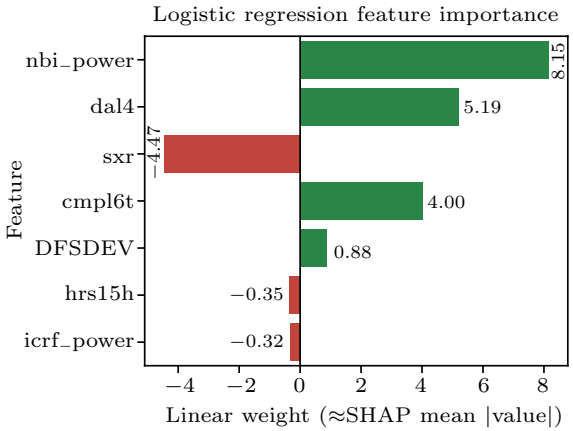


Fig. 7. Feature-weight distribution of the logistic-regression model. The bar chart shows the mean SHAP values of the linear weights for each feature; positive values indicate that the feature increases the probability of a sawtooth event, while negative values exert an inhibitory effect.

The negative SHAP value of the SXR signal and the positive SHAP value of the NBI power form a dynamically antagonistic pair: the former flags that the core electron pressure $n_{\mathrm{e}}T_{\mathrm{e}}$ is too low to supply sufficient free energy to the $m/n = 1/1$ kink mode, whereas the latter represents continuous external heating that steadily steepens $\nabla(n_{\mathrm{e}}T_{\mathrm{e}})$. In this tug-of-war, NBI increases the core-pressure gradient, while SXR marks episodes where energy is depleted (post-crash) or transiently lost (during late reheat), thereby enlarging the stability margin and statistically delaying the next crash. By simultaneously capturing the deep relaxation after a crash and the transient energy losses in the growth phase, the model moves beyond pure correlation and identifies physically meaningful precursors and stabilizers of the sawtooth cycle.

4.2. Individual-sample interpretability

To elucidate the decision logic for individual samples, we plotted SHAP force plots for tens of thousands of spatiotemporal points. In the plots, red rightward arrows increase the probability of 1/1 kink-driven regular complete sawtooth occurrence, while blue leftward arrows suppress the probability; the baseline output accumulates the SHAP contributions of each feature sequentially. Figure 8 presents three typical time points for shot 116510 (sawtooth oscillation interval [2.8598 s, 6.2494 s]): $T = 2.5$ s, 2.86 s, and 4.5 s correspond to Figs. 8(a), 8(b), and 8(c), respectively, and are labeled "before sawtooth oscillation," "early sawtooth oscillation," and "mid-oscillation". Figure 8(a) is located 0.36 s before the first collapse, during the"absolute baseline" phase: the core exhibits low temperature and low density, auxiliary heating is relatively weak, and all MHD activities remain below threshold. The model output is only 2%, fully consistent with the physical scenario of "not yet building up a gradient.". Figure 8(b) sits exactly at the entrance to the oscillation interval. As the NBI power rises and the instability strengthens, the model assigns this instant a $\sim 60\%$ probability of "entering the sawtooth band"; closer to the interval edges the likelihood drops relative to points deep inside the oscillation. This "boundary-blur" effect is the main reason why the sawtooth class is recognized slightly less accurately than the non-sawtooth class, and it reflects an inherent uncertainty common to all temporal-event-detection models. Figure 8(c) is situated in the middle of the oscillation. The injected NBI power dominates the weighted contribution, serving as the core energy source driving each sawtooth cycle. The model assigns an extremely high probability ($\approx 95\%$) that this point is within the active sawtooth region, and the interpretation aligns completely with the observed dense collapses in the experiment.

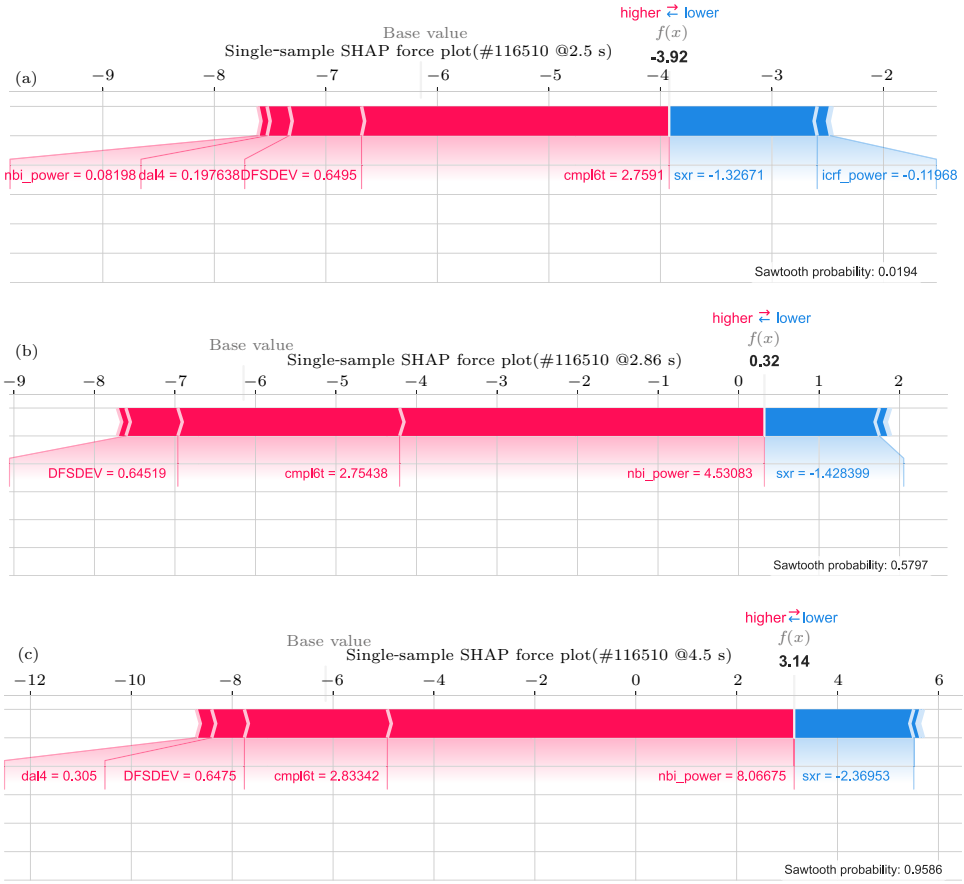


Fig. 8. Local SHAP force plots visualizing feature contributions for individual samples from shot #116510, whose sawtooth oscillations are confined to [2.8598 s, 6.2494 s]: (a) #116510 @ 2.5 s, the "absolute baseline" immediately before the oscillatory window; (b) #116510 @ 2.86 s, the instant of the first crash; (c) #116510 @ 4.5 s, the mid-sequence period of dense sawtooth. Red segments push the model output toward the "sawtooth" class, whereas blue segments suppress it. The horizontal axis denotes SHAP values; larger absolute magnitudes indicate stronger contributions to the prediction for that point. The predicted probability of being classified as a sawtooth event is annotated in the lower-right corner.

5. Conclusion and perspectives

To address the critical need for real-time control of 1/1 kink-driven regular complete sawtooth oscillations in tokamaks, this study proposes, for the first time, an interpretable machine learning framework capable of end-to-end prediction of sawtooth-active intervals. By constructing a large-scale database containing 10^8 samples (2.25×10^8 data points), we systematically compare the classification performance of logistic regression, decision tree, and random forest models. The results demonstrate that the logistic regression model, utilizing

only seven diagnostic inputs, achieves 95% accuracy on an independent test set (6.75×10^6 samples), with per-class recall rates of 0.94 and 0.95. It significantly outperforms traditional algorithms while demonstrating strong robustness under high-dimensional, sparse, and class-imbalanced conditions. Error analysis reveals MSE = 0.0405, RMSE = 0.2012 and MAE = 0.0800 for the probability outputs, with the calibration error consistently remaining below 0.05. These metrics confirm that the model possesses both excellent discriminative capability and a reliable probability distribution, making it suitable for direct application in subsequent feedback control. Both global and individual SHAP interpretations consistently demonstrate that, under identical constraints and heating conditions, low SXR values correspond to a relatively flat core pressure gradient, representing a stable state. In contrast, an increase in NBI power amplifies the central pressure gradient and facilitates regular 1/1 sawtooth triggering, which aligns with established physical understanding. The introduction of SHAP enables multi-scale interpretability from macroscopic trends to individual cases, providing a data-driven quantitative basis for both macroscopic research and case-by-case analysis of regular 1/1 sawtooth phenomena, and establishing an algorithmic foundation for quasi-real-time control.

Future work will focus on the following directions. (1) Embedding the framework into the EAST real-time control system and utilizing FPGA-based acceleration to achieve end-to-end inference latency of less than 1 ms. (2) Introducing online incremental learning to mitigate potential concept drift during long-pulse discharges. (3) Exploring the integration of deep temporal models (e.g., temporal CNN-transformer) with SHAP to capture more complex nonlinear evolution characteristics while maintaining interpretability.

Acknowledgements

Project supported by the National Key R&D Program of China (Grant No. 2022YFE03010003) and the National Natural Science Foundation of China (Grant No. 12275309).

References

[1] Wei L, Liu B, Wang F P, Zhang H and Duan W S 2021 *Chin. Phys. B* **30** 035201
[2] Pan S S, Duan Y M, Xu L Q, Chao Y, Zhong G Q, Sun Y W, Sheng H, Liu H Q, Chu Y Q, Lü B, Jin Y F and Hu L Q 2023 *Acta Phys. Sin.* **72** 135203 (in Chinese)
[3] Fan T S, Yuan X, Zhang X, Xie X F, Chen Z J, Peng X Y, Du T F, Hu Z M, Cui Z Q, Chen J X, Li X Q, Zhang G H and Wang X G 2013 *Sci. Sin.-Phys. Mech. Astron.* **43** 1236
[4] Xu M, Xu L Q, Zhao H L, Li Y Y, Zhong G Q, Hao B L, Ma R R, Chen W, Liu H Q, Xu G S, Hu J S, Wan B N and EAST Team 2023 *Acta Phys. Sin.* **72** 215204 (in Chinese)
[5] Bandyopadhyay I, Igochine V, Sauter O, *et al.* 2025 *Nucl. Fusion* **65** 103001
[6] Chapman I T, Hender T C, Saarelma S, Sharapov S E, Akers R J and Conway N J 2006 *Nucl. Fusion* **46** 1009
[7] Mück A, Goodman T P, Maraschek M, Pereverez G, Ryter F and Zohm H 2005 *Plasma Phys. Control. Fusion* **47** 1633
[8] Chapman I T 2011 *Plasma Phys. Control. Fusion* **53** 013001
[9] Felici F, Goodman T P, Sauter O, Canal G, Duval B P, Rossel J X and the TCV Team 2012 *Nucl. Fusion* **52** 074001
[10] Girardo J B, Sharapov S, Boom J, de Vries P, Fitzgerald M, Hawkes N, Kiptily V, Mantsinen M, Nave M F F, Schneider M and JET Contributors 2016 *Phys. Plasmas* **23** 012505
[11] Wen X D, Xu L Q, Hu L Q, Liu H Q, Duan Y M, Chu Y Q, Zhong G Q, Zhang W, Zhang X J and Mai C W 2023 *Nucl. Techn.* **46** 010502
[12] Li J C, Ding Y H, Yu Q Q, Wang N C, Li D, Zhang X Q, Han D L, Chen Z P, Yang Z J, Zhou S, Yan W, Liang Y F, Zhang X L, Lin X D, Sun H B, Gao X and Li J G 2020 *Nucl. Fusion* **60** 126002
[13] Gude A, Maraschek M, Eulenberg P, Igochine V and ASDEX Upgrade Team, Max Planck Institute for Plasma Physics, Max Planck Society 2015 *Automated sawtooth detection with multi-signal analysis* In *42nd EPS Conference on Plasma Physics* (Geneva: European Physical Society)
[14] van den Brand H, de Baar M R, van Berkel M, Blanken T C, Felici F, Westerhof E, Willensdorfer M, ASDEX Upgrade Team and EUROfusion MST1 Team 2016 *Plasma Phys. Control. Fusion* **58** 075002
[15] Isel N, Isayama A, Ishida S, Sato M, Oikawa T, Fukuda T, Nagashima A, Iwama N and JT-60 Team 2001 *Fusion Eng. Des.* **53** 213
[16] Lan G, De Carlo M, van Diggelen F, Tomczak J M, Roijers D M and Eiben A E 2020 arXiv:2001.07804 [cs.NE]
[17] Proulx M J, Eracleous T, Spencer B, Passfield A, de Sousa A and Mohammadi A 2021 arXiv:2105.13295 [cs.HC]
[18] Gramacy R B and Polson N G 2012 *Bayesian Anal.* **7** 567
[19] El-Koka A, Cha K H and Kang D K 2013 *Regularization parameter tuning optimization approach in logistic regression* In *2013 15th International Conference on Advanced Communications Technology (ICACT)*, PyeongChang, Korea (South), pp. 13–18
[20] Xie Y and Willett R "Online logistic regression on manifolds", 2013 *IEEE International Conference on Acoustics*, Speech and Signal Processing, Vancouver, BC, Canada, 2013, pp. 3367–3371
[21] Lundberg S M and Lee S I 2017 *Adv. Neural Inf. Process. Syst.* **30** 4765
[22] Kumar R, Saxena A, Kumar R, Marwaha S, Singh J and Singh G K 2022 *IEEE Access* **10** 35421
[23] von Goeler S, Stodiek W and Sauthoff W 1974 *Phys. Rev. Lett.* **33** 1201
[24] Danjo A and Koike F 1989 *Phys. Rev. Lett.* **62** 741
[25] Zuo Y, Liu Y, Zhou T, Zhao H, Zhang Y, Figini L, Ang T, Ling B and Hu L 2019 *Plasma Sci. Technol.* **21** 095103
[26] Chen K, Xu L, Hu L, Duan Y, Li X, Yuan Y, Mao S, Sheng X and Zhao J 2016 *Sci. Instrum.* **87** 063504
[27] Piccione A, Berkery J W, Sabbagh S A and Andreopoulos Y 2021 *Nucl. Fusion* **60** 046033
[28] Giacomin M, Pau A, Ricci P, Sauter O, Eich T, the ASDEX Upgrade team, JET Contributors and the TCV team 2022 *Phys. Rev. Lett.* **128** 185003
[29] Wen X D 2023 *Experimental studies on the effect of on-axis ICRF heating on the sawtooth in EAST device* [in Chinese] (Master's Thesis) (Hefei: University of Science and Technology of China) pp. 22–29
[30] Verhulst P J 1838 *Corresp. Math. Phys.* **10** 113 (in French)
[31] Roe B P, Yang H J, Zhu J, Liu Y, Stancu I and McGregor G 2005 *Nucl. Instrum. Methods A* **543** 577
[32] Kirasich K, Smith T and Sadler B 2018 *SMU Data Sci. Rev.* **1** 9