Towards Open Machine Learning Benchmarks for Tokamak Event Prediction from MAST

P. Sharma[1]*, S. Jackson[1], N. Cummings[1], C. J. Ham[1], J. Hodson[1], A. Kirk[1], K. Lawal[1], D. Ryan[1], S. Pamela[1], E.d.D Zapata-Cornejo[2] and The MAST Team

[1] United Kingdom Atomic Energy Authority, Culham Campus, United Kingdom

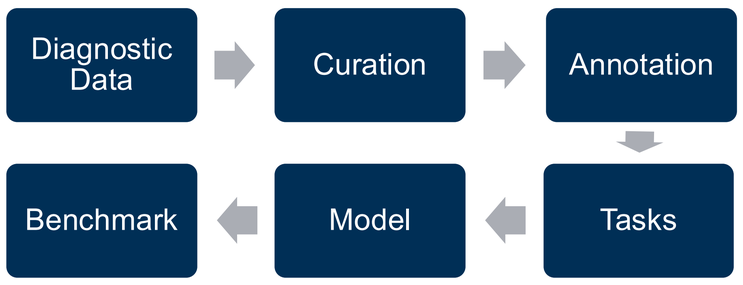
[2] Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, United States

Abstract

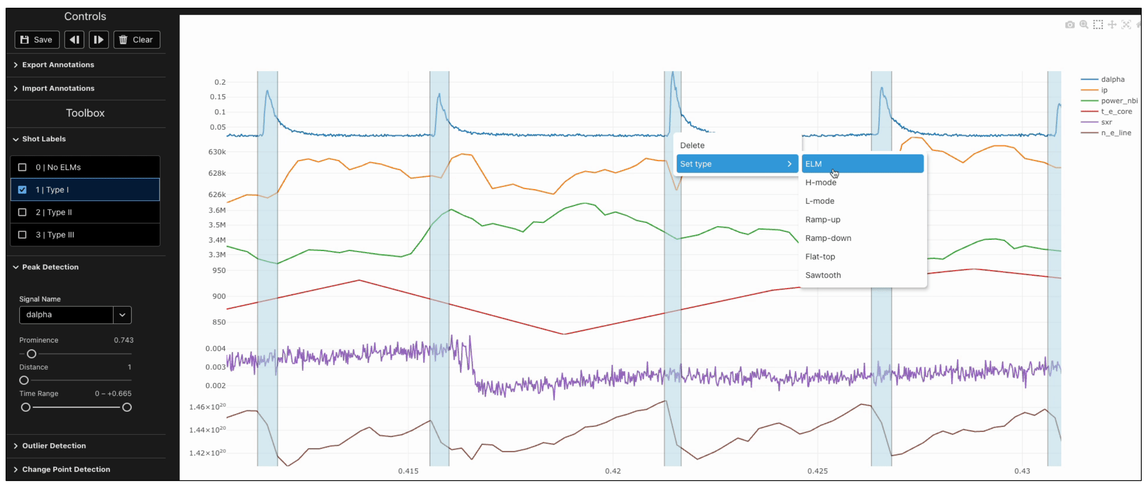
As fusion research advances toward sustainable energy production, reliable prediction of key plasma events is essential for safe tokamak operation [1]. Machine learning (ML) has emerged as a promising approach for such tasks, leveraging large volumes of diagnostic data [2–5]. While fusion facilities are beginning to endorse open data [6, 7, 8], and several closed databases of tokamak event data have been curated [9, 10, 11], the lack of standardized, open benchmarks and data currently impedes reproducibility and the systematic comparison of machine learning algorithms in fusion research. To address this gap, we present ongoing work towards curating event annotations and baseline ML models for four representative tasks: disruption prediction, MHD segmentation, confinement-mode classification, and ELM detection, providing reproducible reference implementations for future data-driven studies.

Introduction

MAST diagnostic signals are curated and pre-processed into a machine learning ready format. Curated signals are then annotated by a human labeller, producing the ground-truth for different types of plasma events. We derived four different tasks and use these annotations to create baseline models on four different tasks and gather baseline performance metrics.



Data Annotation



Our interactive diagnostic annotation tool used to label ELMs, confinement modes, MHD modes and disruptions across key signals. We use this to generate ground truth labels for each of the four tasks considered

Tasks

Four machine learning tasks were identified including 1) ELM identification, 2) confinement mode classification, 3) MHD modes segmentation, 4) and disruptions prediction.

Disruption prediction

- Sudden loss of plasma confinement; must predict ahead of time with warning time.
- Ground truth: Auto-detected from plasma current (417 shots).
- Baseline Model: Stacked BiLSTM with weighted sampling and sliding time window.
- Limitation: Ambiguous cases without sharp current drop remain unresolved.
- Early predictions when flat-top phase is unclear.
- Performance declines as lead time increases (tested 10/30/60 ms).

MHD mode segmentation

- Theory: Plasma instabilities (LLM, fishbones, NTMs, sawteeth) reduce performance and may trigger disruptions.
- Ground truth: Hand labelled using spectrogram annotation tool using Mirnov coil data (85 shots).
- Model: Mask R-CNN with ResNet-101 backbone, based on the Wavystar and Wavystar AI framework [12-14].
- Limitation: High-frequency structures hard to label; non-expert annotations.
- IoU is extremely low because LLMs are thin structures; small misalignments penalise overlap heavily.

Confinement mode classification

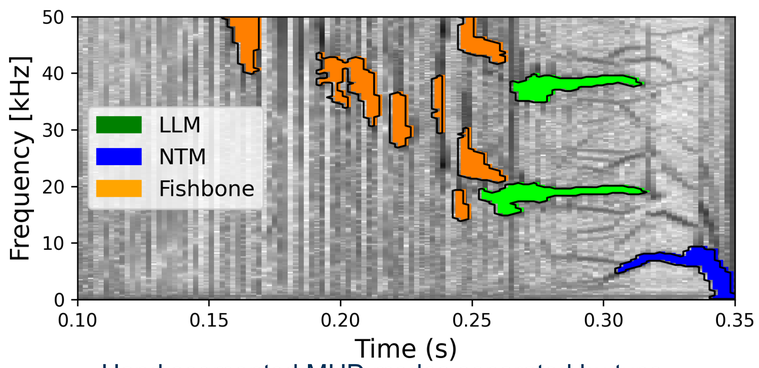
- Theory: Transition between low confinement (L) and high (H) confinement.
- Ground truth: H-mode intervals hand-labelled regions(85 shots).
- Model: 1D U-Net with sliding time window.
- Limitation: Label boundaries may misalign by tens of ms.

ELM spike identification

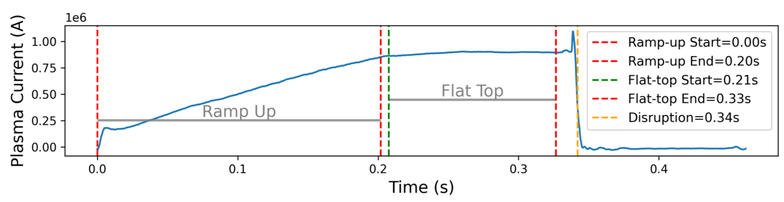
- Theory: Short bursts during H-mode, crucial for plasma–wall interaction.
- Ground truth: Thresholding on Dα + manual verification (101 shots).
- Model: 1D U-Net with sliding time window.
- Limitation: Narrow spikes → metrics sensitive to small misalignments.

Ground truth examples

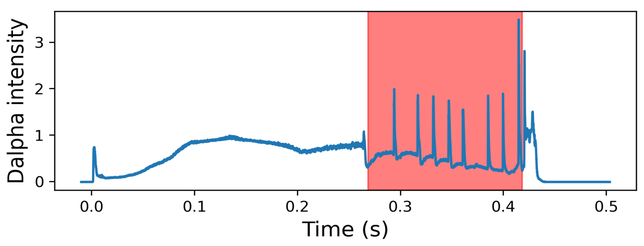
Example annotations curated from the data annotation stage



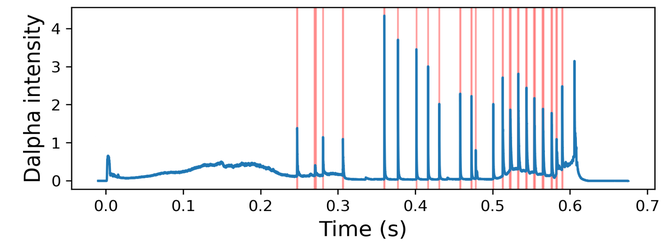
Hand segmented MHD modes separated by type



Identified disruption time point & pulse phases

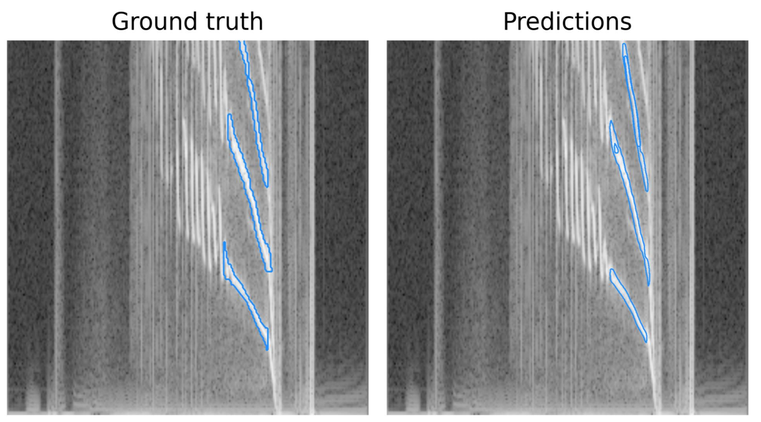


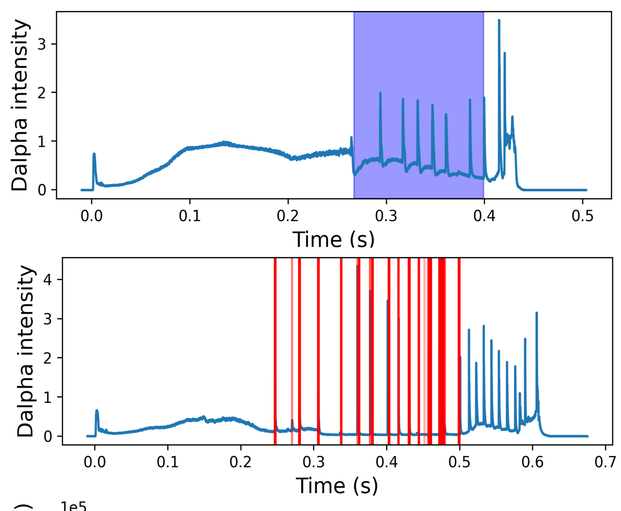
Hand labelled L/H confinement mode classification

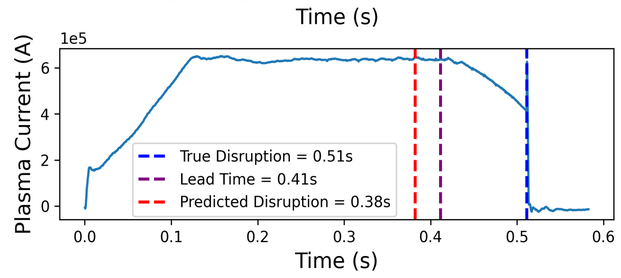


Hand labelled ELM spike identification

Results







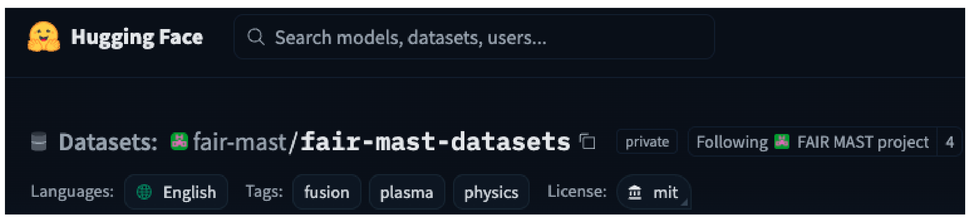
Baseline model predictions:
Top-left: MHD segmentation (predicted vs labelled contours).
Top-right: Confinement classification (predicted H-mode interval).
Middle-right: ELM detection (predicted spikes in Dα).
Bottom-right: Disruption prediction (lead time vs true disruption).

- **Classification** - *Confinement & ELMs:* Precision/Recall/F1 + ROC AUC to better reflect event detection under class imbalance. *Disruption:* Precision/Recall/F1 only; **ROC AUC omitted** (unreliable under extreme temporal imbalance + windowing).
- **Disruption early-warning:** hit-rate 0.63; median warning 26.5 ms. Alarms earlier than 50 ms are counted as premature (not true positives).
- **Segmentation (MHD):** report IoU; note it heavily penalises thin modes (LLMs).

Task	Confinement	ELMs	MHD modes	Disruption
Precision	0.82 ± 0.22	0.79 ± 0.20	0.75 ± 0.13	0.84 ± 0.07
Recall	0.83 ± 0.21	0.80 ± 0.20	0.73 ± 0.15	0.94 ± 0.06
F1-score	0.79 ± 0.25	0.78 ± 0.20	0.72 ± 0.10	0.87 ± 0.09
IoU	-	-	0.39 ± 0.01	-
ROC AUC	0.90	0.85	-	-

Conclusion & Future Work

- Baselines provide starting points but are not yet full benchmarks.
- Limitations include annotation noise and label misalignments.
- Current metrics do not fully capture thin/filamentary structures.
- Future work:
  - Improve label quality via review + model feedback.
  - Extend baselines towards an open benchmark suite with annotation and data tools.
  - Release datasets and models openly for community use.



The FAIR-MAST repository will host curated AI-ready diagnostic data and event annotations from MAST, supporting reproducible ML research for community use.

1. De Vries P.C. et al. Nucl. Fusion 51, 053018 (2011).
2. Anirudh R. et al. IEEE Trans. Plasma Sci. (2023).
3. Pavone A. et al. Plasma Phys. Control. Fusion 65, 053001 (2023).
4. Zhu J.X. et al. Nucl. Fusion 63, 046009 (2023).
5. Montes K.J. et al. Nucl. Fusion 61, 026022 (2021).
6. Academic Research Platform LHD / National Institute for Fusion Science (2025).
7. Jackson S. et al. SoftwareX 27, 101869 (2024).
8. WEST Project. Opening of WEST data and update of WEST Publication Rules (2025).
9. Eidietis N.W. et al. Nucl. Fusion 55, 063030 (2015).
10. Zhang M. et al. Fusion Eng. Des. 160, 111981 (2020).
11. Litaudon X.L. et al. Nucl. Fusion (2023).
12. Zapata-Cornejo E.D.D. et al. Plasma Phys. Control. Fusion 66, 095016 (2024).
13. Bustos A. et al. Plasma Phys. Control. Fusion 63, 095001 (2021).
14. Zapata-Cornejo E.D.D. Ph.D. Thesis, Aix-Marseille University (2024).

Acknowledgements
We thank the MAST team for providing experimental data. Computations used the Cambridge CSD3 cluster. The MHD-modes task builds upon the Wavystar-labelling codes developed at MIT (Supported by ENI).

Find out more
www.gov.uk/ukaea