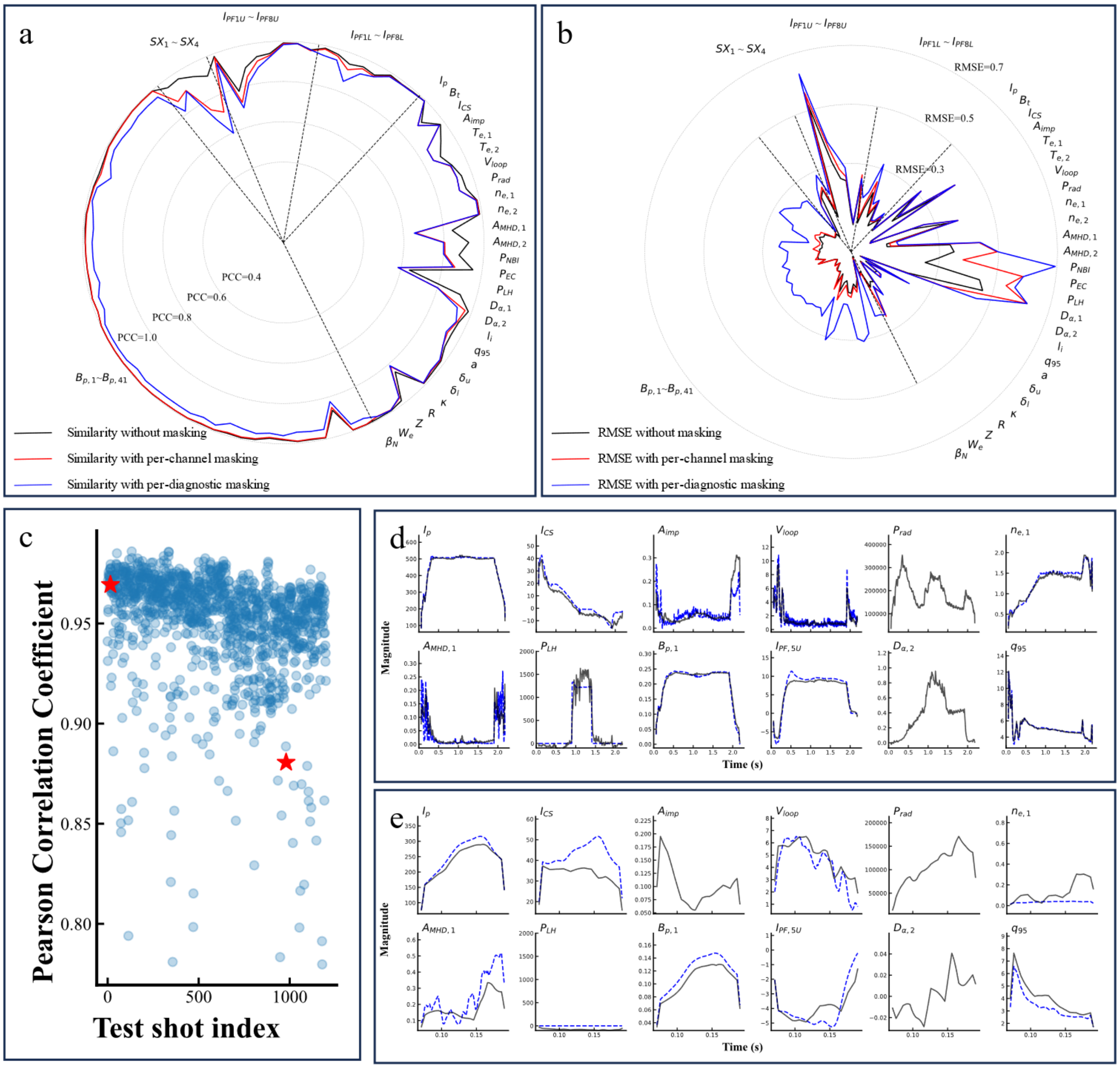
Supplementary Note 1: application of FusionMAE in latest experimental campaign of HL-3

In the first half of 2025, HL-3 carried out a new experimental campaign. FusionMAE exhibits consistent performance on this newly acquired dataset comprising 1199 shots, achieving average PCC of 97.1%, 95.0%, and 92.7% under no intentional masking, single-channel masking, and per-diagnostic masking, respectively. Detailed evaluation results on this dataset are presented in Supplementary Fig.1.



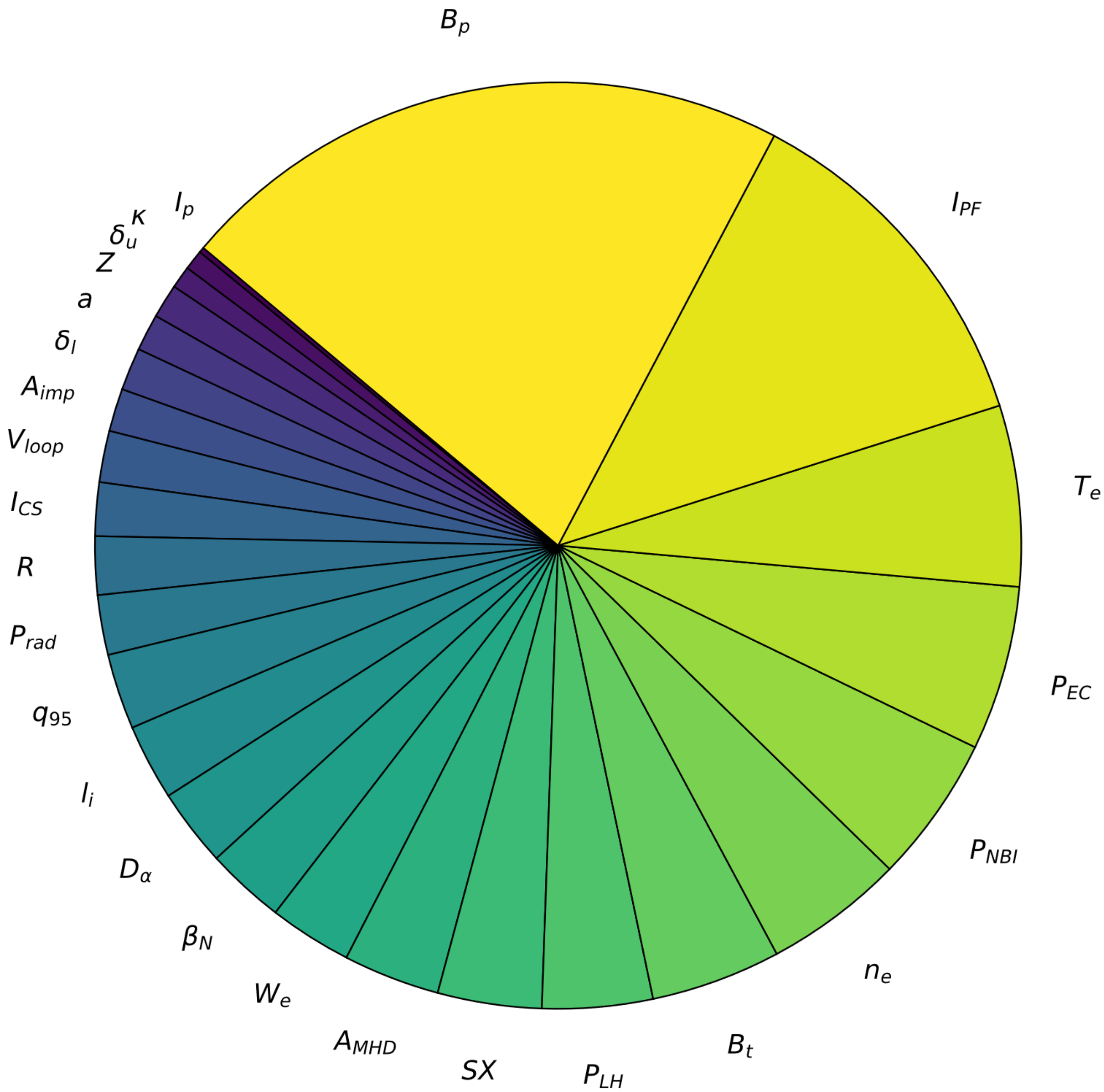
Supplementary Fig.1: Evaluation of FusionMAE's performance in latest experimental campaign of HL-3. (d) Shot 9352. (e) Shot 12891

Supplementary Note 2: contribution of each diagnostic for plasma state embedding

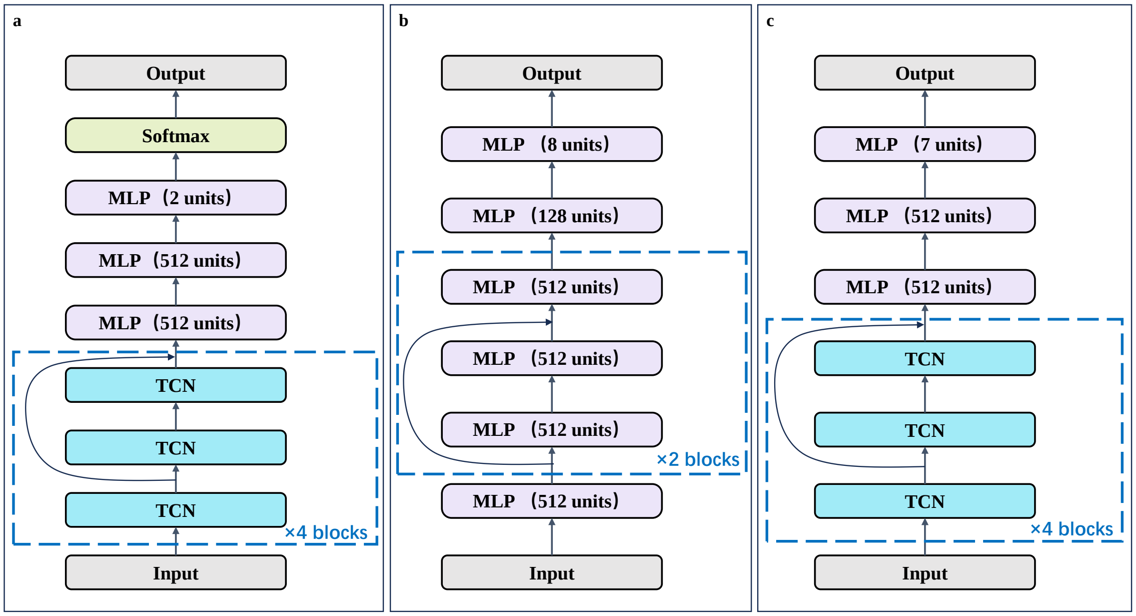
FusionMAE introduces a novel paradigm for describing plasma state. Rather than relying on an increasing number of diagnostic measurements to provide additional parameters, FusionMAE represents plasma state using a fixed-length vector. As more diagnostic signals are incorporated, the resulting vector becomes increasingly refined, enabling a more precise representation of the plasma state. Consequently, the contribution of each diagnostic signal to defining the plasma state can be quantified using Supplementary Equation (1):

$$\mathrm{contribution}(i)=\frac{1\text{-PCC}^{*}(i)}{\sum_{j=1}^{88}\left(1\text{-PCC}^{*}(j)\right)} \qquad (1)$$

where $\mathrm{PCC}^{*}(i)$ denotes the PCC between the embedding vector computed from the complete set of input channels and the embedding vector obtained when masking the i_{th} channel. Supplementary Fig.2 illustrates the contributions of the 88 diagnostic signals in determining plasma state, providing critical insights into the relative importance of different subsystems for future fusion reactors. For most input channels, this sequential distribution aligns well with physical intuition. The description of plasma state is primarily determined by the most critical physical quantities such as B_p, T_e, n_e and B_t. Secondly, it is directly governed by the dominant terms among external control inputs like I_{PF} and auxiliary heating power. The information about plasma magnetic configuration and current appears somewhat anomalously positioned at the end, but this can be understood as their information being entirely replaceable by B_p and I_{PF}. This analysis helps us quantify the contribution of each diagnostic to plasma observation, thereby guiding subsequent optimization of diagnostic systems and the design of fusion reactor diagnostics.



Supplementary Fig.2: The contribution of different channels in the determination of plasma state embedding.



Supplementary Fig.3: The neural network structure used in three downstream tasks: (a) disruption prediction, (b) equilibrium fitting surrogate model, (c) plasma parameter prediction. TCN refers to a temporal convolutional neural network, whose detailed hyper-parameters keeps the same as in [1, 2]

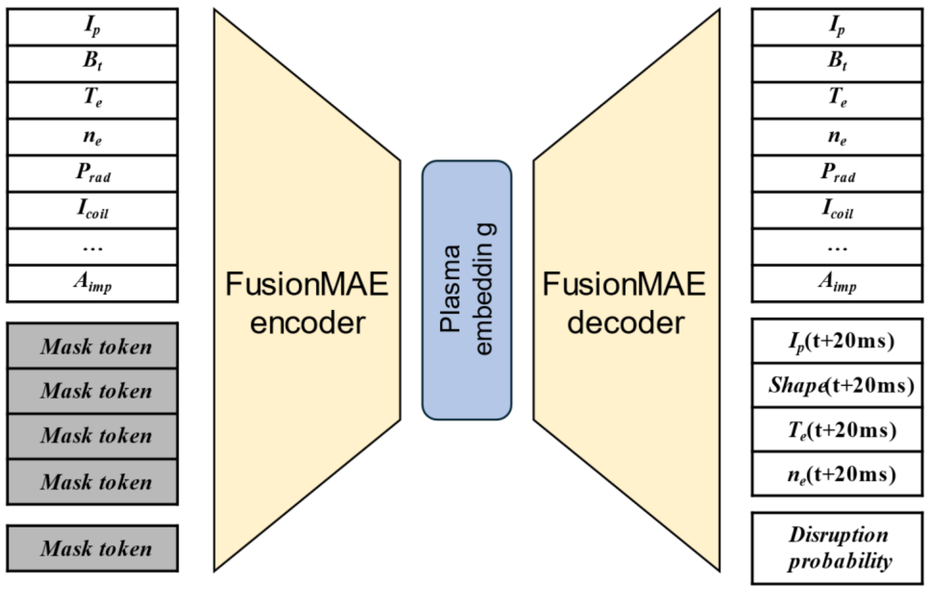
Supplementary Note 3: the downstream task of disruption prediction

Disruption is an anomalous event that occurs during tokamak discharges, characterized by the sudden loss of plasma confinement. This phenomenon can lead to severe consequences, including intense electromagnetic forces, thermal loads, and the generation of runaway electrons, all of which may cause significant damage to critical components of the device. The primary objective of disruption prediction is to identify precursors of disruptions, such as loss of plasma control and the development of magneto-hydrodynamic (MHD) instabilities, thereby enabling the activation of mitigation strategies to minimize potential damage [3]. Over the past decades, disruption prediction has primarily followed two technical paradigms: physics-based criteria and machine learning approaches. Both have achieved substantial progress and, more recently, have increasingly informed and complemented each other. Given the methodological compatibility with data-driven modeling, this study focuses on machine learning–based disruption prediction.

Extensive investigations have been carried out on devices such as JET, DIII-D, HL-3, KSTAR, and EAST, employing a wide range of algorithms from ensemble methods to deep learning, and yielding mature, device-specific prediction experience [4-11]. Recent studies can be broadly categorized into four directions: (i) multi-task learning frameworks that jointly model disruption prediction with related tasks to improve warning accuracy [12, 13]; (ii) anomaly-detection-based approaches that reduce dependence on disruptive samples and achieve competitive performance using predominantly non-disruptive data [14–16]; (iii) interpretability analyses that provide physical insights to guide disruption avoidance and mechanism studies [17–19]; and (iv) enhancement of cross-parameter and cross-device generalization to meet the requirements of future fusion reactors[20,21].

The neural network architecture of the disruption prediction algorithms proposed in this study is presented in Supplementary Fig.3a. The model architecture, data preprocessing, loss function and performance evaluation keep the same paradigm as our prior study conducted on HL-3[1], except that the dataset is extended by the new experimental campaigns. A comparison between LSTM, CNN+LSTM, Transformer, and the final TCN architecture is summarized in Supplementary Table 1, where TCN is identified as the most suitable backbone for disruption prediction on HL-3.

FusionMAE is further introduced as an independent, dataset-level optimization module that enhances the baseline TCN model. Specifically, two FusionMAE-assisted variants are considered: in the TCN + FusionMAE embedding model, the original 88-channel time series are replaced by sequences of 256-dimensional plasma state embeddings; in the TCN + data restored by FusionMAE model, the input remains the original 88 diagnostic signals, while missing or invalid channels are reconstructed using FusionMAE. As demonstrated in Fig.6 of main text and Supplementary Table 1, incorporating FusionMAE leads to a substantial improvement in disruption prediction performance. Importantly, this optimization strategy can be implemented as a front-end preprocessing module for downstream neural networks, making it readily compatible with most existing model architectures and optimization techniques reported in the literature.



Supplementary Fig.4: The structure of FusionMAE to directly output disruption probability as an additional diagnostic channel.

Furthermore, with appropriate modifications, FusionMAE can be directly extended to perform disruption prediction. This is achieved by introducing disruption probability and future plasma parameters as additional channels, masking these channels at the input stage and reconstructing them at the output. The resulting architecture, illustrated in Supplementary Fig.4, provides a unified formulation that naturally integrates several key concepts previously explored in disruption prediction research.

- **Anomaly detection.** FusionMAE's encoder–decoder structure aligns well with anomaly detection-based approaches, in which an increase in reconstruction error serves as a disruption precursor. As FusionMAE is pretrained via a reconstruction task, it is inherently compatible with this paradigm and can learn disruption-relevant patterns under appropriate supervision.

- **Multi-task learning**: Previous work has shown that that incorporating auxiliary tasks like instability recognition can significantly enhance accuracy of disruption prediction. In FusionMAE, the task of disruption prediction is naturally integrated into a multi-task learning framework, jointly optimized with tasks including missing signal imputation and secondary parameter reconstruction.
- **Predict-first neural network (PFNN)**: In our earlier state-of-the-art model in HL-3[1], the algorithm gets better performance by predicting the future plasma trajectories at first and then predicting disruptions by comparing the experimental outcomes and the predicted ones. This idea can be integrated into FusionMAE by treating future plasma parameters as masked channels to be reconstructed.

Supplementary Table 1: Performances of different versions of disruption prediction algorithms

Model	AUC	True Positive Rate (TPR)	True Negative Rate (TNR)
TCN+data restored by FusionMAE	0.925	0.875	0.862
TCN+FusionMAE embedding	0.924	0.873	0.840
TCN(baseline)	0.915	0.876	0.819
LSTM	0.906	0.835	0.824
CNN+LSTM	0.900	0.816	0.821
Transformer	0.821	0.764	0.739
FusionMAE with channel of disruption probability	0.975	0.935	0.944

With these modifications, FusionMAE achieves an AUC of 0.975 on the HL-3 disruption prediction task, significantly outperforming all previously reported results on this device. Due to the architectural complexity of FusionMAE, cross-device generalization has not yet been systematically investigated. Nevertheless, this direction appears highly promising and will be the focus of our future work, which might bring breakthrough for cross-tokamak disruption prediction.

Supplementary Note 4: the downstream task of equilibrium fitting

Equilibrium fitting is a fundamental data analysis task in tokamak operation. By numerically solving the Grad-Shafranov equation under the constraints of diagnostic signals—such as the poloidal magnetic field surrounding the plasma—the plasma's flux and current density distribution can be obtained. Based on these distributions, the plasma shape can be determined. The resulting equilibrium solution serves as a critical input for plasma shape and position control, laying the foundation for stable plasma discharge.

In this study, the second downstream task involves training a neural network to serve as a surrogate for the equilibrium fitting program (EFIT). Similar approaches have been investigated on multiple devices, including KSTAR and DIII-D, demonstrating their fundamental feasibility [22, 23]. Recent developments have further improved performance by incorporating auxiliary constraints such as physics-informed loss functions and multi-task learning strategies [2, 24, 25]. In our implementation, the baseline model takes I_{PF}, I_{CS}, B_p and I_p as inputs to predict plasma shape parameters. Two additional versions replace the raw diagnostics with either FusionMAE embeddings or FusionMAE-reconstructed inputs while performing the same prediction task.

Supplementary Table 2: Performances of different versions of EFIT surrogate models.

Parameter	MLP + Data restored by FusionMAE		MLP + FusionMAE embedding		MLP (baseline)		Reconstructed by FusionMAE	
	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE
R(m)	0.953	0.00944	0.949	0.0100	0.943	0.0117	0.960	0.00897
Z(m)	0.927	0.0118	0.919	0.0129	0.900	0.0155	0.924	0.0119
a(cm)	0.941	0.746	0.936	0.828	0.927	0.990	0.939	0.805
κ	0.990	0.0138	0.988	0.0154	0.983	0.0198	0.986	0.165
δ_u	0.971	0.0198	0.965	0.0219	0.954	0.0262	0.980	0.0183
δ_l	0.985	0.0226	0.984	0.0258	0.973	0.0339	0.984	0.0239
l_i	0.967	0.00649	0.961	0.00678	0.945	0.00864	0.940	0.00930
q_{95}	0.954	0.214	0.954	0.220	0.940	0.300	0.967	0.202
Average	0.961	-	0.957	-	0.946	-	0.959	-

The baseline architecture, illustrated in Supplementary Fig.3b, keeps the EFIT-NN framework described in [2]. Model performance is evaluated by PCC and RMSE, which quantifies the agreement between predicted plasma shape parameters and EFIT-derived ground truth. A comprehensive comparison of different model variants is presented in Fig.6 of main text and Supplementary Table 2, clearly demonstrating the performance improvements enabled by the integration of FusionMAE.

Supplementary Note 5: the downstream task of plasma evolution prediction

Plasma evolution prediction serves as the foundation for controller design in tokamak operation. The prediction algorithm can be represented as a function, as described in Supplementary Equation (2).

$$s(t+1)=F(s(t),\, a(t)) \qquad (2)$$

where $s(t)$ and $s(t+1)$ denote the present and future plasma parameters, respectively, while $a(t)$ represents the control action. The function F encapsulates the plasma response to various control actions. Researchers design controllers based on the prediction algorithm to regulate plasma behavior and achieve the desired target parameters. In this study, $s(t)$ and $s(t+1)$ include I_p, R, Z, a, κ, δ_u, δ_l, while the control actions $a(t)$ are represented by I_{PF} and I_{CS}.

The objective of this downstream task is to train a neural network on historical data to approximate the function F and predict plasma current and shape evolution. Data-driven plasma response models have been validated on devices such as DIII-D, KSTAR, NSTX, and EAST, covering the prediction of profiles, instabilities, and global plasma parameters [26–29]. In HL-3, we have previously developed a series of data-driven simulators for magnetic control, including a single-step predictor mapping coil currents to plasma response [1], as well as two autoregressive models based on coil currents and power supply voltages, respectively [30,31].

At the current stage, FusionMAE is not well aligned with autoregressive prediction frameworks in [29,79], as such approaches require repeated model inference over multiple time steps. While this is feasible for lightweight models, it would lead to prohibitive training and inference costs when applied together with FusionMAE. Therefore, in this study we adopted the single-step prediction model in [1] as our baseline. The corresponding architecture is shown in Supplementary Fig.3c, and the model design, data preprocessing, loss function, and evaluation protocol follow those described in [1]. Model performance is assessed using the PCC and RMSE, which quantify the agreement between predicted and experimentally measured plasma shape parameters.

Supplementary Table 3 presents a detailed comparison among different algorithm variants, demonstrating the overall performance improvement achieved by integrating FusionMAE. In addition, the WaveNet-based model proposed in [31] is adapted to make auto-regressive prediction within a 20 ms window and take the result of final step as a single-step prediction result. The results indicate that TCN and WaveNet provide comparable baseline performance for this downstream task under the single-step prediction setting.

Notably, plasma evolution prediction is the only downstream task in which FusionMAE exhibits a slight performance degradation in the absence of intentional data masking, as indicated by the gray and blue bars in Fig.6 of main text. This behavior is attributed to the distinct physical roles of the state variables $s(t)$ and the actuator signal $a(t)$, which are treated equivalently in the current FusionMAE formulation. While this simplification has negligible impact on disruption prediction and EFIT surrogate tasks, it is suboptimal for evolution prediction and results in a modest loss of accuracy. At higher masking ratios, however, this deficit is offset by the gains from FusionMAE's data imputation capability. Future work will explore autoregressive pretraining strategies that explicitly distinguish state and control variables, which is expected to further improve performance in plasma evolution prediction.

Supplementary Table 3: Performances of different versions of evolution prediction models.

Parameter	TCN + Data restored by FusionMAE		TCN + FusionMAE embedding		TCN (baseline)		WaveNet	
	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE
I_p(kA)	0.999	2.63	0.998	5.03	0.999	2.53	0.999	2.26
R(m)	0.989	0.00386	0.961	0.00807	0.988	0.00387	0.984	0.00401
Z(m)	0.981	0.00482	0.968	0.00851	0.979	0.00494	0.972	0.00610
a(cm)	0.985	0.331	0.969	0.423	0.985	0.328	0.985	0.329
κ	0.998	0.00565	0.990	0.0124	0.994	0.00681	0.998	0.00301
δ_u	0.995	0.00656	0.977	0.0135	0.994	0.00667	0.988	0.0101
δ_l	0.997	0.00869	0.989	0.0177	0.997	0.00895	0.996	0.00910
Average	0.992	-	0.978	-	0.991	-	0.989	-

Supplementary Table 4: Hyperparameter settings used for training FusionMAE

Hyper-parameter	Explanation	Value
Gaussian noise	A Gaussian-distributed noise signal is added to the input data to implement data augmentation during the training phase.	μ=0, σ=0.003
Optimization method	The function used to map the partial derivatives of the prediction error with respect to the neural network weights to the adjustments made to the neural network weights.	Adam (adaptive moment estimation), with learning rate = $0.001 \times (0.1)^{i/10000}$, in the i_{th} step of training
Weight decay	The technique of reducing the absolute values of network weights with a relative ratio after each training step to mitigate overfitting by penalizing excessively large weights.	$0.0003 \times (0.1)^{i/20000}$, in the i_{th} step of training
Batch size	The number of samples used to compute the prediction error and update the network weights simultaneously during a training step.	256
Early stopping	The number of epochs that the training framework will wait for improved performance before halting training and reverting to the network's best weights.	20
Maximum training epoch	The maximum number of training epochs. The training process will be terminated after the final epoch.	500, the training usually finishes between 80~200 epochs

Supplementary Note 6: application of FusionMAE in a spherical tokamak: SUNIST-2

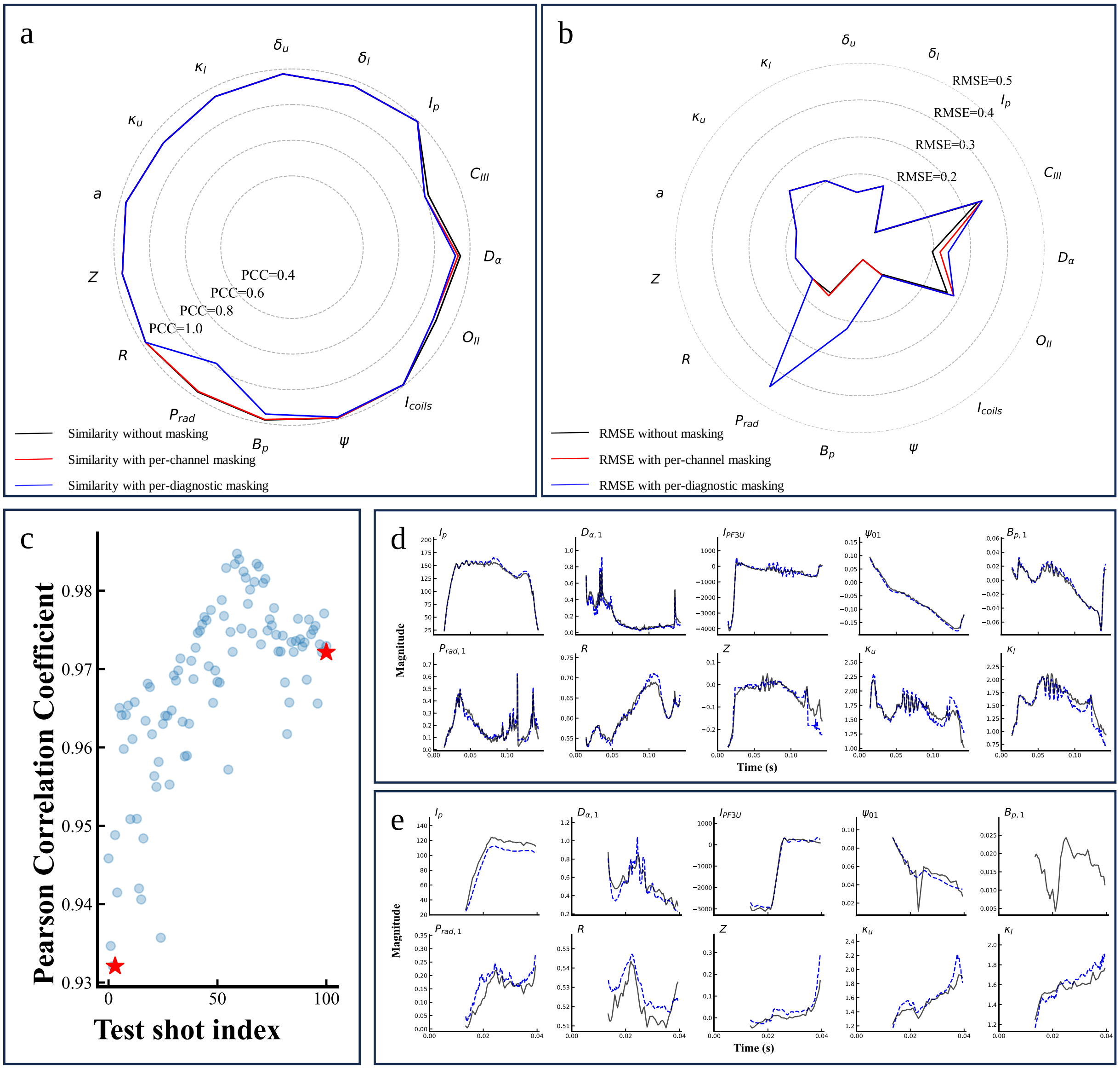
Owing to the inherent complexity of the FusionMAE framework, its application to cross-device scenarios has not yet been systematically investigated. Nevertheless, we have conducted preliminary adaptations of the algorithm to several new devices, providing initial evidence of its general applicability. Here, we present FusionMAE's performance on SUNIST-2, a medium-sized spherical tokamak jointly operated by Star Torus Fusion and Tsinghua University [32].

Supplementary Table 5: The detailed information about channels used in the development of FusionMAE in SUNIST-2, including the physical significance, statistical properties and data health rate.

Channel name	Channel number	Unit	Physical significance	Statistical properties		
				Mean value	Standard deviation	Health rate
I_p	1	MA	Toroidal plasma current, measured by Rogowski coils.	0.122	0.0367	100%
R	1	m	Horizontal position of the plasma center in the poloidal cross-section, calculated using the EFIT code.	0.583	0.0637	84.2%
Z	1	m	Vertical position of the plasma center in the poloidal cross-section, calculated using the EFIT code.	-0.0230	0.138	84.2%
a	1	m	Minor radius of the plasma in the poloidal cross-section, calculated using the EFIT code.	0.319	0.0439	84.2%
κ_u	1	-	Elongation ratio of the upper half of plasma in the vertical direction, calculated using the EFIT code.	1.68	0.296	84.2%
κ_l	1	-	Elongation ratio of the lower half of plasma in the vertical direction, calculated using the EFIT code.	1.63	0.343	84.2%
δ_u	1	-	Upper triangularity, defined as the degree to which the upper half of the plasma deviates from a perfectly circular shape towards a triangular form, calculated using the EFIT code.	0.346	0.243	84.2%
δ_l	1	-	Lower triangularity, defined as the degree to which the lower half of the plasma deviates from a perfectly circular shape towards a triangular form, calculated using the EFIT code.	0.390	0.231	84.2%
B_p	87	T	Local poloidal magnetic field strength, measured by 41 pick-up coils surrounding the plasma.	0.00239	0.184	99.2%
I_{coil}	13	kA	Current in the poloidal field coils of the tokamak.	-1.97	2.12	100%
ψ	29	Wb	Magnetic flux.	-0.173	0.182	99.2%
P_{rad}	16	a.u.	Line integrated thermal radiation power measured by AXUV diagnostic.	0.584	0.348	98.7%
A_{imp}	2	a.u.	Intensity of the C_{III}, O_{II} emission line.	0.103	0.285	99.8%
D_α	2	a.u.	Intensity of the D-α emission line.	0.197	0.213	99.8%

In SUNIST-2, the input of FusionMAE consists of 157 channel as detailed in Supplementary Table 5. The resulting dataset consists of 417 shots for training and 116 shots for validation, with an average pulse length of approximately 200 ms. Compared with HL-3, the SUNIST-2 diagnostic set exhibits lower diversity, with magnetic measurements accounting for most input channels. This high redundancy among magnetic diagnostics makes the training of FusionMAE easier. Conversely, the substantially smaller dataset size, approximately 50 times smaller than that of HL-3, poses additional challenges for model development and generalization.

The overall result of FusionMAE is comparable with the results obtained on HL-3, as shown in Supplementary Fig.5. The PCC under no intentional masking, single-channel masking and per-diagnostic masking are 98.0%, 97.8% and 93.9%, respectively. These encouraging results motivate future efforts toward applying FusionMAE across multiple devices using zero-shot or few-shot learning strategies. Although this represents a highly challenging problem, it has the potential to constitute a transformative step toward general-purpose, cross-device AI frameworks for fusion research.



Supplementary Fig.5: Evaluation of FusionMAE's performance in SUNIST-2. The result for signals from a same diagnostic has been averaged in subplot (a) and (b) as there are too many channels in SUNIST-2. Other details follow the same as in Fig.3 of main text. (d) Shot 250731022. (e) Shot 250419001

Supplementary References

1. Zongyu Yang, Wulyu Zhong, Fan Xia, et al. Implementing deep learning-based disruption prediction in a drifting data environment of new tokamak: HL-3. Nuclear Fusion, 65, 026030 (2025).

2. G.H. Zheng, Z.Y. Yang, S.F. Liu, et al. Real-time equilibrium reconstruction by multi-task learning neural network based on HL-3 tokamak. Nuclear Fusion, 64, 126041 (2024)

3. T.C. Hender, J.C. Wesley, J. Bialek, et al. Progress in the ITER physics basis: chapter 3: MHD stability, operational limits and disruptions Nuclear Fusion, 47, S128–202 (2007)

4. J. Vega, A. Murari, S. Dormido-Canto, et al. Disruption prediction with artificial intelligence techniques in tokamak plasmas. Nature physics, 18, 741-750, (2022)

5. Zongyu Yang, Fan Xia, Xianming Song, et al. A disruption predictor based on a 1.5-dimensional convolutional neural network in HL-2A. Nuclear Fusion, 60, 016017 (2020).

6. M. Tobin, S. A. Sabbagh, V. Zamkovska, et al. Vertical instability forecasting and controllability assessment of multi-device tokamak plasmas in DECAF with data-driven optimization. Plasma Physics and Controlled Fusion, 66 105020, (2024).

7. Wei Zheng, Fengming Xue, Zhongyong Chen, et al. Disruption prediction for future tokamaks using parameter-based transfer learning. Communications Physics 6, 181 (2023).

8. J.X. Zhu, C. Rea , K. Montes, et al. Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks. Nuclear Fusion 61, 026007 (2021)

9. B. H. Guo, et al. Disruption prediction using a full convolutional neural network on EAST. Plasma Physiscs and Controlled Fusion, 63, 025008(2021).

10. Jeongwon Lee, Sang-hee Hahn, Hyunsun Han, et al. Real-time data-driven disruption prediction and its mitigation of MA-plasma experiments in KSTAR with a lower carbon divertor. Nuclear Fusion, 65, 056040 (2025)

11. E. Aymerich, G. Sias, F. Pisano, et al. CNN disruption predictor at JET: Early versus late data fusion approach. Fusion Engineering and Design, 193, 113668(2023)

12. J.X. Zhu, C. Rea, R.S. Granetz, et al. Integrated deep learning framework for unstable event identification and disruption prediction of tokamak plasmas. Nuclear Fusion, 63, 046009 (2023)

13. Y. H. Jia, H. Y. Wang, W. H. Hu, et al. Plasma Instability Events Detection and Disruption Prediction in EAST Tokamak via Heterogeneous-Feature Multi-Task Learning, 30th IAEA Fusion Energy Conference, Chengdu, China, EX-S:3276 (2025)

14. W. Zheng, Q. Q. Wu, M. Zhang, et al. Disruption predictor based on neural network and anomaly detection on J-TEXT. Plasma Physics and Controlled Fusion, 62, 045012 (2020)

15. X.K. Ai, W. Zheng, M. Zhang, et al. Adaptive anomaly detection disruption prediction starting from first discharge on tokamak. Nuclear Fusion, 65, 036011(2025)

16. Vega J. et al 2016 Review of disruption predictors in nuclear fusion: classical, from scratch and anomaly detection approaches IECON 2016–42nd Annual Conf. IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016

17. D.M. Liu, X.L. Zhu, Y.S. Jiang, et al. An interpretable disruption predictor on EAST using improved XGBoost and SHAP. Nuclear Fusion, 65, 086035 (2025)

18. G.A. Ratta, J. Vega, A. Murari, et al. PHAD: a phase-oriented disruption prediction strategy for avoidance, prevention, and mitigation in JET. Nuclear Fusion, 61, 116055 (2021)

19. Zongyu Yang, Fan Xia, Xianming Song, et al. In-depth research on the interpretable disruption predictor in HL-2A. Nuclear Fusion, 61, 126042 (2021)

20. Zongyu Yang, Wulyu Zhong, Jiyuan Li, et al. Augmenting the extrapolation capability of disruption prediction to extended parameter regimes by predict-first neural network. 30th IAEA Fusion Energy Conference, Chengdu, China, TH-S:3178 (2025)

21. B.H. Guo, D.L. Chen, C. Rea, et al. Disruption prediction on EAST with different wallconditions based on a multi-scale deep hybrid neural network. Nuclear Fusion, 63, 094001 (2023)

22. L. L. Lao, S. Kruger, C. Akcay, et al. Application of machine learning and artificial intelligence to extend EFIT equilibrium reconstruction. Plasma Physics and Controlled Fusion, 64, 074001 (2022)

23. Semin Joung, Jaewook Kim, Sehyun Kwak, et al. Deep neural network Grad–Shafranov solver constrained with measured magnetic signals. Nuclear Fusion, 60, 016034 (2020)

24. S. Madireddy, C. Akçay, S. E. Kruger, et al. EFIT-Prime: Probabilistic and physics-constrained reduced-order neural network model for equilibrium reconstruction in DIII-D. Physics of Plasmas, 31, 092505 (2024)

25. Semin Joung, Y.-C.Ghim, Jaewook Kim, et al. GS-DeepNet: mastering tokamak plasma equilibria with deep neural networks and the Grad–Shafranov equation. Scientifc Reports, 13, 15799(2023)

26. J. Abbate, R. Conlin and E. Kolemen. Data-driven profile prediction for DIII-D. Nuclear Fusion 61 046027 (2021)

27. Jaemin Seo, Y.-S. Na, B. Kim, et al. Feedforward beta control in the KSTAR tokamak by deep reinforcement learning. Nuclear Fusion, 61, 106010 (2021)

28. M.D. Boyer and J. Chadwick. Prediction of electron density and pressure profile shapes on NSTX-U using neural networks. Nuclear Fusion, 61, 046024 (2021)

29. Chenguang Wan, Zhi Yu, Feng Wang et al. Experiment data-driven modeling of tokamak discharge in EAST. Nuclear Fusion, 61, 066015 (2021)

30. Niannian Wu, Zongyu Yang, Rongpeng Li, et al. High-Fidelity Data-Driven Dynamics Model for Reinforcement Learning-based Magnetic Control in HL-3 Tokamak. Communications Physics 8, 393 (2025)

31. Jiyuan Li and Zongyu Yang. Data-Driven Tokamak Plasma Magnetic Response Modelling. Sixth IAEA Technical Meeting on Fusion Data Processing, Validation and Analysis, Shanghai, China, Oral:28 (2025)

32. Tingzhi Chang, Zhicheng Lei, Menghua Yang, et al. First results of lithium coating experiments in the SUNIST-2 spherical tokamak. Plasma Physics and Controlled Fusion, 67, 075023 (2024)