nature portfolio

Peer Review File

FusionMAE, a self-supervised pretrained model to optimize and simplify diagnostic and control of fusion plasma

Corresponding Author: Dr Wulyu Zhong

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript proposes the use of large-scale models, such as Transformers, for nuclear fusion data processing. Transformer models are the algorithms running many of the models known as Large Language Models (LLM), and the authors show that if trained with a large unlabeled database of a device, they can extrapolate significant knowledge on the physics and operations of the device, for instance, recovering missing data. Moreover, the model embeddings can be used to train other models to address specific tasks.
The approach is interesting because it is a low-processing yet powerful method for handling complex fusion datasets. However, some aspects of the work are not fully contextualised within the existing of nuclear fusion research. For example, concepts such as mapping experimental conditions have been addressed in the literature, and the contribution of this study is best framed as an extension enabled by high-capacity models. Furthermore, the demonstration examples would benefit from deeper analysis and clearer comparison with established approaches.
Here are my comments addressing specific points of the paper and suggestions:
1. The work should be framed better within the relevant fusion literature, especially when referring to the “emerging capabilities” of the model. For example, on page 10: “Analogous to word embeddings, the plasma embeddings of physically similar discharges are expected to cluster in latent space. This structure is confirmed via Uniform Manifold Approximation and Projection (UMAP) projection of embeddings from the test set. Notably, despite uniform treatment of all 88 input channels during training, FusionMAE autonomously identifies and organises embeddings along key plasma parameters (Bt, Ip, ne and βN), indicating a nontrivial, learned understanding of plasma physics.” The clustering of common plasma conditions has been studied with dimensionality reduction methods and by interpreting the embeddings of deep neural network models [1–5].

2. The term “emerging capabilities” may be misleading, as it suggests the model itself directly performs downstream tasks. A clearer phrasing, such as “enhancement of downstream tasks”, would better reflect the role of the model in data recovery and cleaning (including handling of missing or potentially corrupted diagnostics).

3. The references on disruption prediction [38–41] are sparse given the maturity of the field. For instance, the authors should frame the methodology applied in the supervised learning methodology [6] (the model needs labeled disruptive samples to be trained) in contrast with unsupervised and anomaly detection approaches which may not need a labelling of the pulses or may also not need disruptive pulses at all [7].
Recent examples from other devices such as JET, AUG, DIII-D could be included, as the ones provided in the recent relevant Chapter of ITER Physics Basis [8].

4. The figure is difficult to interpret in its current form. Consider reducing the number of reconstructed signals shown, highlighting the best- and worst-reconstructed cases. This would make the uncertainty bands more visible. Please clarify whether reconstructions are based on masked inputs or full data, since the legend currently introduces ambiguity.

5. The performance metric is not clearly labelled. If this is the AUC described in the Methods section, it should be explicitly reported on the axes and in the caption. Please also specify the number of pulses used for training and discuss whether all 88 diagnostic channels are available in real-time during HL-3 operation.

6. In general, I suggest splitting Fig. 5 into several subfigures to easily identify the quality of the reconstructions from FusionMAE and to better detail each downstream task.
7. The manuscript notes similarity to prior disruption prediction work on HL-3 [27]. Please clarify how the present architecture differs and provide quantitative comparisons (e.g., false alarm rate, disruption warning time, or improvements from reconstructed inputs) with the previous one. Moreover, are there examples where the reconstructed measurements improve the model results?
8. A similar discussion on the novelty of the approach with respect to previous work should also be made for Equilibrium Fitting [28] and Plasma Prediction Evolution [29], discussing the changes performed and their impact on the result shown.

9. What is the inference time of FusionMAE? Can it be deployed in real-time control systems, or is its use currently limited to offline analysis?

10. Has the robustness of FusionMAE to corrupted channels been tested? Can the model improve corrupted inputs, or must these be explicitly masked during inference?

11. As a comment, it seems that the reconstructed signals always improve the performance of the model carrying out the downstream task.

12. The applicability of FusionMAE to devices beyond HL-3 should be discussed. While fine-tuning on additional pulses is likely necessary for new machines, differences in operational space and diagnostic sets may further increase data requirements. In this sense, labelling FusionMAE as a “foundation model” may be overstated. A more measured framing would position it as a promising framework for data-driven analysis in fusion research.

13. There are a few typos: “chanllenges” on Page 5, “whose detailed hyper parameters keeps the same as in [27,28]”, where I suggest using “are” in place of “keeps”.

1. Pau, A. et al. A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET. Nucl. Fusion 59, 106017 (2019).
2. Aymerich, E. et al. A statistical approach for the automatic identification of the start of the chain of events leading to the disruptions at JET. Nucl. Fusion 61, 036013 (2021).
3. Zhu, J. et al. Scenario adaptive disruption prediction study for next generation burning-plasma tokamaks. Nucl. Fusion 61, 114005 (2021).
4. Zanini, M. et al. Automated W7-X sawtooth crashes detection and characterisation. Nucl. Fusion (2024) doi:10.1088/1741-4326/ad490b.
5. Wei, Y., et al. A dimensionality reduction algorithm for mapping tokamak operational regimes using a variational autoencoder (VAE) neural network. Nucl. Fusion 61, 126063 (2021).
6. Vega, J., et al Disruption prediction with artificial intelligence techniques in tokamak plasmas. Nat. Phys. 18, 741–750 (2022).
7. Aledda, R. et ak. Improvements in disruption prediction at ASDEX Upgrade. Fusion Engineering and Design 96–97, 698–702 (2015).
8. Bandyopadhyay, I. et al. MHD, disruptions and control physics: Chapter 4 of the special issue: on the path to tokamak burning plasma operation. Nucl. Fusion 65, 103001 (2025).

Reviewer #2

(Remarks to the Author)
Summary:
The paper proposes FusionMAE, a transformer-based masked autoencoder that compresses 88 HL-3 tokamak diagnostic signals into a 256-dimensional “plasma status embedding” and claims three “emergent” capabilities—automatic analysis of secondary signals, a universal interface for downstream tasks, and robustness to missing diagnostics—together with a “virtual backup diagnosis” mode with ~0.97 Pearson correlation for reconstructed channels. The idea is timely and potentially useful, but several methodological choices and the strength of evidence fall short of the paper’s “foundation model” framing and its more ambitious claims.

Major claims
1. Unified embedding across diagnostics. A masked-autoencoding framework compresses 88 signals (12 systems) sampled at 1 kHz over 10 ms windows into a 256-D vector, from which the decoder reconstructs original and missing channels.
2. “Virtual backup diagnosis.” With 25% channel masking in training, the model reconstructs masked channels with average PCC of 0.967 (described as 96.7% reliability), and compress-then-reconstruct PCC of 0.986.
3. Emergent capabilities.
a. Automatic analysis of secondary signals (e.g., EFIT-derived shape parameters) by reconstructing them when masked.
b. All-purpose vector that improves or matches three downstream tasks—disruption prediction (DPR), equilibrium surrogate (EFIT-NN), and plasma evolution prediction (PPR)—relative to baselines trained on raw signals.
c. Robustness to missing diagnostics (stable outputs with 5–20% masked inputs).
4. Practical impact claims: The embedding could simplify interfaces, reduce the number of diagnostics, and enhance control performance, positioning FusionMAE as a domain-specific “foundation model” for fusion.

Novelty and Broader applicability:
1. Bringing masked-autoencoding to multi-diagnostic fusion time series and explicitly framing the output as an all-purpose embedding for control/analysis is a useful systems-level idea that goes beyond one-task predictors prevalent in fusion ML. The demonstration spans three representative tasks, which is broader than much prior single-task works.
2. The architecture is a generic MAE-style encoder-decoder with masking from, previous works. Against the current wave of foundation models for time series and long-context sequence models, a 256-D latent trained on 10 ms windows over a single device is modest in scale and scope. The “emergent capabilities” largely follow from the objective (masked channel reconstruction), so characterizing them as “emergence” overstates what is essentially learned cross-channel regression.
3. The fusion community will likely be interested in imputation/robustness to missing diagnostics and a single learned interface that can feed multiple modules. Broader ML interest is more limited unless the paper offers stronger methodological innovations (architecture, training recipes, scaling laws) or cross-device generalization demonstrating out-of-domain transfer.

Review:
1. Reporting a single PCC per channel averaged across shots is encour, but PCC is generally not equivalent to reliability in an operational sense. There is no calibration analysis, per-channel error distribution, or failure analysis under realistic missing-data patterns (most fusion diagnostics are missing not at random).
2. The training/eval masking is channel-level and random, while actual outages have structure (e.g., all channels are missing for a time interval). Stress tests with structured masking (by subsystem or time interval) and campaign-wise OOD splits would make the claim more credible.
3. The experiment masks all secondary channels then reconstructs them. That shows the model can regress secondary variables from primaries. This is valuable, but because those same secondary channels are part of the pretraining target, the result is not out-of-distribution; it is a standard masked-target recovery. A stronger test would:
a. Exclude specific secondary channels from the reconstruction loss during pretraining and evaluate zero-shot recovery;
b. Compare against supervised surrogates trained only on primaries.
4. The summary bars suggest improvements when using the embedding or completed data, but no confidence intervals or statistical tests are provided. This would be crucial to judge if gains are significant.
5. The baselines are not fully matched for capacity/regularization/data preprocessing, and there is no comparison to simpler representation baselines.
a. Provide: 95% Credible Intervals; tests for AUC differences in DPR; paired permutation tests for PCC.
b. Stronger baselines (e.g., a TCN/Transformer/VAE embedding without masking; PCA).
c. Ablations on mask ratio, latent size, depth/heads
d. The main text says the model is trained to minimize MSE , but Methods Eq. (3) uses L1 with weights. Which objective was actually used? Provide ablations for L1 vs. L2 and for the W schedule.
6. Uncertainty bands by randomly masking 5–20% of channels is informative but it would be more informative to consider deep ensembles, MC-dropout, heteroscedastic/quantile heads, and calibration curves to produce calibrated predictive intervals
7. All experiments are on HL-3. To support claims about influencing reactor design (e.g., fewer diagnostics), cross-tokamak transfer (e.g., zero-/few-shot to another tokomak) and closed-loop control studies (in simulation first) would be essential.
8. Given the scope of the results, the authors should consider tempering the “foundation” framing and present FusionMAE as a practical multi-diagnostic representation that improves fault tolerance to missing sensors and simplifies interfaces across analytics and control modules.
9. Tokenization: The methods section say that each column is treated as a patch, encapsulating plasma information from a single diagnostic channel even though the input is 88×10 (channels × time) and earlier text states each channel is projected. Please clarify whether tokens correspond to channels (88 tokens) or time slices (10 tokens), and how time information beyond 10 ms is handled.
10. Most transformer variants for sequences use pre-norm LayerNorm/RMSNorm; BatchNorm can be brittle for variable-length sequences and distribution shift. Please justify this design and report comparisons to LayerNorm.
11. Please include the justification for choosing the 10ms window as this might miss slower physics, and might be too short for many control relevant processes
12. Extended Table 1 shows low health rates for some key channels (e.g., ne, Te, Prad). Please evaluate FusionMAE under device-realistic outage patterns (e.g., “all interferometers down”) rather than random MCAR masks
13. Number of epochs/steps, training wall-, hardware, and parameter count are not reported; early stopping is 20 epochs but the total training regime is unclear. Please provide full reproducibility metadata.
a. parameter count, training steps/epochs, and FLOPs;
b. Provide scaling curves vs. latent size/depth and mask ratio to justify the choice of 256-D latent and 25% masking
c. Include throughput/latency measurements to support control-room use
14. Using PCC for reconstruction and for EFIT-NN/PPR is standard but insufficient: PCC is insensitive to scale/offset. Add NRMSE/MAE and relative error vs. dynamic range per channel. For DPR, also report precision–recall AUC.
15. The dataset spans four campaigns (2022–2025) with a single random split (1,950/495 shots). To assess robustness to drift, add chronological splits (train on early campaigns, test on later campaigns).
16. Signals are resampled to 1 kHz, but normalization procedures (per-channel z-score? robust scaling?) are not fully specified. Extended Table 1 lists means/SDs, yet preprocessing pipelines (filtering, detrending, synchronization and handling physically “a.u.” channels) should be described step-by-step.
17. Code/data are “available on request” rather than open, which limits reproducibility and community use.
18. A few typos (“chanllenges”, etc.) need to be corrected as well as ensure consistent units and symbols

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have replied thoroughly to my comments, providing strong evidence of the added value of the FusionMAE for fusion data processing.
However, for the added part of the SUNIST-2 data and the latest HL-3 campaign, I suggest adding a table similar to Extended Table 1 to present a clear overview of the data analysed in this work.
Moreover, I attach here a marked-up version with some minor typos and rephrasing.

Reviewer #2

(Remarks to the Author)
Dear Authors,
Thank you for the substantial additional experimentation and thoughtful revisions in response to my feedback. It has significantly strengthened the manuscript. However, a few minor things remain:
1. The “reliability” is used at a few places like abstract and Fig 3 caption. Since the metrics reported in the work are primarily correlation-style and no explicit calibration analysis or worst-case scenario has been done to tie to the notion of operational reliability. I would suggest to use “accuracy” instead.
2. The norm of the error is not clear in equation 3, will be good to explicitly make it L2
3. The SUNIST-2 results show methods portability, but not cross-device or finetuning/post-training, few-shot generalization showed for FusionMAE in Figure 1. Also, there is no closed-loop control demonstration while it is mentioned in the title. Consider changing it.
4. There are still a few missed mentions of foundational model in the manuscripts, including references to finetuning and posttraining, prompt engineering in fig 1,, “foundational model for HL-3” in page 6 and explicit use of RLHF, MOE and CoT as promising directions in pg 19 but the connection on how FusionMAE can be extended is not clear.
5. Can you also test a case where a set of diagnostics are missing in a time window

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Communications Physics initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Dear Reviewer,

We sincerely thank you for taking the time to review our manuscript. Your comments demonstrate a precise understanding of our work and highlight critical issues that are essential for strengthening the paper.

Our original motivation was to explore a data-driven framework inspired by the concept of foundational models for fusion research. Given the breadth of capabilities such a framework aims to address, together with the strict length constraints of Nature-series journals, some intermediate results and supporting analyses were not sufficiently presented or emphasized in the initial submission. This may have limited the clarity and completeness of the technical justification in certain parts of the manuscript.

Both you and another reviewer raised important concerns in this regard. In response, we have substantially expanded the experimental evaluations and further refined the technical implementation. We believe that these revisions significantly strengthen the rigor and clarity of the work, and we are grateful for your constructive feedback, which has greatly improved the quality of the manuscript.

Below we provide point-by-point responses to your specific comments and suggestions. The original comments are shown in blue, our responses are shown in black, and the corresponding revisions in the manuscript are highlighted in yellow.

Overall comment:

The approach is interesting because it is a low-processing yet powerful method for handling complex fusion datasets. However, some aspects of the work are not fully contextualised within the existing body of nuclear fusion research. For example, concepts such as mapping experimental conditions have been addressed in the literature, and the contribution of this study is best framed as an extension enabled by high-capacity models. Furthermore, the demonstration examples would benefit from deeper analysis and clearer comparison with established approaches.

Response:

We thank the reviewer for this insightful comment. One challenge in the original manuscript arises from a typical characteristic of foundation-style models, namely their broad functional scope. In the initial submission, we validated FusionMAE across multiple downstream tasks, which resulted in insufficient contextualization and clarity for individual applications. In response to this concern, we have substantially revised the *Downstream Tasks* descriptions in the *Methods* section. For each downstream task, we now explicitly introduce the research background, summarize relevant prior work, and clarify the baseline methodologies, with appropriate references to existing literature to aid interpretation within the limited manuscript length.

Regarding the distinction between our approach and existing studies on mapping experimental conditions, we have added additional subfigure c,d,e in Fig. 4 to better highlight the differences. Further clarification is also provided in the responses to the specific comments below.

Here are my comments addressing specific points of the paper and suggestions:

1. The work should be framed better within the relevant fusion literature, especially when referring to the "emerging capabilities" of the model. For example, on page 10: "Analogous to word embeddings, the plasma embeddings of physically similar discharges are expected to cluster in latent space. This structure is confirmed via Uniform Manifold Approximation and Projection (UMAP) projection of embeddings from the test set. Notably, despite uniform treatment of all 88 input channels during training, FusionMAE autonomously identifies and organises embeddings along key plasma parameters (Bt, Ip, ne and βN), indicating a nontrivial, learned understanding of plasma physics." The clustering of common plasma conditions has been studied with dimensionality reduction methods and by interpreting the embeddings of deep neural network models [1–5].

Response:

Thank you for this important comment. Regarding the analysis of the UMAP visualizations and the descriptions of the downstream tasks (disruption prediction, EFIT surrogate modeling, and plasma evolution prediction), we have made the following revisions to clarify the concern:

1. **Clarification of the contribution in dimensionality reduction analysis (UMAP).** We have added a dedicated paragraph to explicitly clarify that the primary contribution of FusionMAE in this context is not the use of dimensionality reduction itself, but rather its ability to enable meaningful latent-space organization without manual feature engineering or data cleaning. By learning plasma embeddings from heterogeneous and incomplete diagnostic combinations, FusionMAE can assess the intrinsic similarity of plasma states across discharges with differing diagnostic availability.

   To illustrate this point concretely, we added a representative example in Fig. 4, where two discharges with nearly identical plasma trajectories but different sets of available diagnostics are analyzed. When UMAP is applied directly to the raw diagnostic data, the two shots are incorrectly separated due to differences in available signals. In contrast, UMAP applied to FusionMAE embeddings correctly clusters these discharges together, demonstrating FusionMAE's ability to recover the underlying plasma-state similarity beyond diagnostic configuration effects.

   We also explicitly clarify that the primary contribution of FusionMAE as follows:

   Compared to previous studies that employed dimensionality reduction methods like t-SNE or UMAP to analyze the scenario of discharges, FusionMAE can identify whether the underlying plasma state is consistent across different combinations of diagnostic signals.

2. **Improved contextualization of downstream tasks within existing literature.** For the downstream tasks of disruption prediction, EFIT surrogate modeling, and plasma evolution prediction, we have revised the *Downstream Tasks* descriptions in the *Methods* section to include a concise literature overview for each task. These additions summarize the scientific motivation, standard methodological approaches, and recent advances, thereby providing clearer context for the reader. We further clarify that FusionMAE primarily functions as a front-end data processing module, which is designed to be compatible with most existing optimization strategies proposed in recent studies. For example, in the section for

disruption prediction, we have:

The model architecture, data preprocessing, loss function and performance evaluation keep the same paradigm as our prior study conducted on HL-3[27], except that the dataset is extended by the new experimental campaigns. … 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 diagnostic 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.

We have also added Tables 4–6, which present a more comprehensive comparison of performance across different model variants, explicitly demonstrating the incremental improvements introduced by FusionMAE.

2. The term "emerging capabilities" may be misleading, as it suggests the model itself directly performs downstream tasks. A clearer phrasing, such as "enhancement of downstream tasks", would better reflect the role of the model in data recovery and cleaning (including handling of missing or potentially corrupted diagnostics).

Response:

Thank you for this important clarification. We agree that the term *"emerging capabilities"* may be misleading in the current context, as FusionMAE does not directly execute downstream tasks in an end-to-end manner, but instead enhances them primarily through data recovery, cleaning, and representation learning.

We therefore revised the manuscript to replace *"emerging capabilities"* with *"downstream applications"* throughout, and adjusted the positioning of FusionMAE from a *"foundation model"* to a *device-specific pretrained model*. This more accurately reflects the current scope of the work, where FusionMAE serves as a general-purpose pretraining framework that improves the performance and robustness of multiple downstream tasks rather than performing zero-shot task inference.

These revisions are intended to better align the terminology with the actual functionality of the model and to avoid overstatement of its current capabilities. While we believe this framework has the potential to evolve toward more general and transferable models in the future, the present study focuses strictly on its role as a pretraining and enhancement module.

3. The references on disruption prediction [38–41] are sparse given the maturity of the field. For instance, the authors should frame the methodology applied in the supervised learning methodology [6] (the model needs labeled disruptive samples to be trained) in contrast with unsupervised and anomaly detection approaches which may not need a labelling of the pulses or may also not need disruptive pulses at all [7].

Recent examples from other devices such as JET, AUG, DIII-D could be included, as the ones provided in the recent relevant Chapter of ITER Physics Basis [8].

Response:

Thank you for this insightful comment. We agree that disruption prediction is a mature research area and that our original manuscript did not sufficiently contextualize this downstream task within the broader fusion literature.

In response, we have substantially revised the *Methods* section by adding dedicated literature overviews for each downstream task, including disruption prediction, EFIT surrogate modeling, and plasma evolution prediction. For each task, we now explicitly summarize its physical significance, commonly adopted methodologies, and recent representative studies. We further clarify that FusionMAE primarily acts as a data-driven preprocessing and representation-learning module that can be integrated at the front end of existing task-specific pipelines, and is therefore compatible with most optimization strategies proposed in prior work.

For disruption prediction, we firstly benchmarked several commonly used supervised learning architectures (including TCN, LSTM, CNN+LSTM, and Transformer-based models) to establish a reasonable and competitive baseline for HL-3.

A comparison between LSTM, CNN+LSTM, Transformer, and the final TCN architecture is summarized in Table 4, where TCN is identified as the most suitable backbone for disruption prediction on HL-3.

We further clarify that FusionMAE primarily functions as a front-end data processing and representation module, which is designed to be compatible with most existing optimization strategies proposed in recent studies.

As demonstrated in Table 4 and Fig.6, 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.

Building upon the baseline, we then explored an extended FusionMAE-based architecture that integrates key ideas proposed in recent studies, including anomaly-detection-inspired reconstruction, multi-task learning, and physics-guided future trajectory prediction (PFNN). This integrated framework resulted in a substantial improvement in prediction performance on the HL-3 dataset.

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 Fig. 10, provides a unified formulation that naturally integrates several key concepts previously explored in disruption prediction research. … 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.

We also attempted a direct comparison with purely anomaly-detection-based disruption prediction methods. However, on the HL-3 dataset, such approaches achieved relatively limited performance (AUC ≈ 0.60). Our preliminary analysis suggests that this is mainly due to differences in dataset construction: in devices such as J-TEXT, anomaly-detection-based methods are typically applied only during the plasma flat-top phase, where reconstruction losses clearly separate disruptive and non-disruptive discharges. In contrast, the dataset used in this study includes a large proportion of ramp-up and ramp-down phases, during which reconstruction errors increase even for non-disruptive shots, thereby reducing the discriminative power of anomaly-based criteria.

For the EFIT surrogate and plasma evolution prediction tasks, we additionally expanded the comparison with previously published HL-3 models, allowing the performance improvements introduced by FusionMAE to be more clearly and quantitatively assessed.

We hope that these revisions and additional analyses adequately address your concerns and provide a clearer positioning of our work within the existing disruption prediction literature.

4. The figure is difficult to interpret in its current form. Consider reducing the number of reconstructed signals shown, highlighting the best- and worst-reconstructed cases. This would make the uncertainty bands more visible. Please clarify whether reconstructions are based on masked inputs or full data, since the legend currently introduces ambiguity.

Response:

Thank you for this helpful suggestion. We agree that the original presentation of Figure 3 made it difficult to visually assess reconstruction quality and uncertainty.

In the revised manuscript, we adjusted the figure layout while retaining the per-channel PCC radar plots, as FusionMAE involves many diagnostic channels and the relative reconstruction performance across channels provides valuable information for analysis.

At the same time, following your recommendation, we now compute the average PCC across all channels for each test discharge and present the results as a scatter plot. Based on this distribution, we select relatively better and poorly reconstructed discharges and show their reconstructed time-series waveforms. This addition makes the model's performance more directly interpretable.

We also revised the figure legend to remove ambiguity regarding the reconstruction settings. The legend has been updated to explicitly distinguish between *"without masking"*, *"with per-channel masking"*, and *"with per-diagnostic masking"*.

5. The performance metric is not clearly labelled. If this is the AUC described in the Methods section, it should be explicitly reported on the axes and in the caption. Please also specify the number of pulses used for training and discuss whether all 88 diagnostic channels are available in real-time during HL-3 operation.

Response:

Thank you for pointing this out. We have revised Figure 6 to explicitly label the performance metric on the axes and in the caption. In this figure, the reported metric is the AUC and PCC as described in the *Methods* section, and this is now clearly indicated to avoid ambiguity.

All training and testing procedures in this study, including both the self-supervised pretraining and downstream tasks, use the same dataset split described in the *Dataset Description* section of the *Methods*: 1,950 pulses for training and 495 pulses for testing. We have clarified this explicitly in the revised manuscript.

The discharges are randomly divided into a development set and a test set, containing 1,950 and 495 shots, respectively. The development set is employed for training and

validating the neural network, while the test set is used to assess algorithm performance. This dataset splitting strategy consistently applied across both the pretraining and evaluation of FusionMAE, as well as the training and testing of all downstream task models.
Regarding data availability, all 88 diagnostic channels used as model inputs are accessible in real time during HL-3 operation. This point has now been explicitly stated in the *Dataset Description* section to ensure clarity about the practical applicability of the proposed framework.
As illustrated in Fig.2a, each discharge includes 88 recorded signals that characterize the plasma state, which can be measured in real-time in HL-3 plasma control system.

6. In general, I suggest splitting Fig. 5 into several subfigures to easily identify the quality of the reconstructions from FusionMAE and to better detail each downstream task.
Response:
Thank you for this suggestion. In the revised manuscript, we have reorganized the figures to improve readability: Figures 5, 6, and 7 are now used separately to present the results for the three downstream applications, respectively.

7. The manuscript notes similarity to prior disruption prediction work on HL-3 [27]. Please clarify how the present architecture differs and provide quantitative comparisons (e.g., false alarm rate, disruption warning time, or improvements from reconstructed inputs) with the previous one. Moreover, are there examples where the reconstructed measurements improve the model results?
Response:
Thank you for raising this important point. We agree that the original description did not sufficiently clarify the relationship between the present disruption prediction architecture and our prior work on HL-3 [27].
In the revised manuscript, we explicitly clarify that FusionMAE is introduced as a data processing and representation-learning module attached to the front end of the original downstream models. The core architectures of the disruption prediction, EFIT surrogate, and plasma evolution prediction networks are not modified. Accordingly, the model structure, data preprocessing procedures, loss function design, and evaluation protocols for these downstream tasks remain identical to those used in the corresponding prior studies, as now clearly stated in the relevant sections.
As demonstrated in Table 4 and Fig.6, 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.
Regarding quantitative comparisons with previous HL-3 results, we note that the datasets used in this study differ from those in earlier works due to the inclusion of new experimental campaigns and updated data curation. As a result, a direct comparison of absolute performance metrics (e.g., false alarm rate or warning time) with previously reported values is not strictly feasible. Nevertheless, the baseline models employed in this study achieve performance levels comparable to those reported in earlier HL-3 publications[27], providing

a consistent reference point for evaluating the impact of FusionMAE.

The primary benefit of reconstructed input does not arise from introducing new physical information, but rather from a statistical effect: by mitigating the influence of missing or invalid diagnostic channels, FusionMAE reduces spurious variability in the input space, leading to a clearer separation between disruptive and non-disruptive discharges in the prediction output distribution. This effect is reflected in the overall improvement of performance metrics reported in Table 4 and Figure 6. We attempted to identify individual case studies in which the reconstruction of a specific diagnostic (e.g., density measurements near density-limit disruptions) played a decisive role in improving prediction outcomes. However, such clear-cut examples were not consistently observed in the current dataset.

We have revised the manuscript to better reflect these points and to avoid overstating the nature of the improvements introduced by reconstructed measurements.

8. A similar discussion on the novelty of the approach with respect to previous work should also be made for Equilibrium Fitting [28] and Plasma Prediction Evolution [29], discussing the changes performed and their impact on the result shown.

Thank you for this suggestion. We agree that a similar clarification regarding novelty and architectural changes is necessary for the equilibrium fitting and plasma evolution prediction tasks.

Response:

In the revised manuscript, we explicitly state that the treatment of these two downstream tasks follows the same principle as that adopted for disruption prediction. FusionMAE is introduced solely as a data processing and representation-learning module at the front end of the original models, without modifying their core network architectures. Consequently, the model structure, data preprocessing procedures, loss function definitions, and evaluation protocols for the EFIT surrogate model [28] and the plasma evolution prediction model [79] remain fully consistent with those used in the corresponding prior studies.

We further clarify that the observed performance improvements in these tasks arise from enhanced input data quality and robustness to missing or corrupted diagnostics, rather than from changes to the task-specific model designs. These points are now explicitly described in the downstream task sections of the revised manuscript.

For EFIT-NN:

The baseline surrogate architecture, illustrated in Fig. 9b, follows the EFIT-NN framework described in [28]. 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 Table 5 and Figure 6, clearly demonstrating the performance improvements enabled by the integration of FusionMAE.

For PPR:

At the current stage, FusionMAE is not well aligned with autoregressive prediction frameworks in [29,83], 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 [27] as our baseline. The corresponding architecture is shown in Fig.9c, and the model design, data preprocessing, loss function, and evaluation protocol follow those described in [27]. Model performance is assessed using the PCC and RMSE, which quantify the agreement between predicted and experimentally measured plasma shape parameters.

Table 6 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 [83] 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.

9. What is the inference time of FusionMAE? Can it be deployed in real-time control systems, or is its use currently limited to offline analysis?

Response:

We thank the reviewer for this question. On the current hardware and implementation, the single-frame inference time of FusionMAE is approximately 4.8 ms. When combined with a 10 ms input window, this allows the model to operate in a quasi-realtime manner for online monitoring and prediction tasks. However, the present implementation does not fully meet the 1 ms control cycle requirement of plasma control systems (PCS). This clarification regarding inference latency and deployment scope has been explicitly added to the section of *Training of FusionMAE*.

Other training-related hyperparameters are summarized in Table 3. The final FusionMAE model contains approximately 14 million trainable parameters. Training requires about 10 hours on a single NVIDIA A100 GPU, while inference for a single time slice takes 4.8 ms, enabling quasi-realtime application when combined with a 10 ms input time window.

10. Has the robustness of FusionMAE to corrupted channels been tested? Can the model improve corrupted inputs, or must these be explicitly masked during inference?

Response:

We thank the reviewer for raising this important point regarding robustness to corrupted diagnostic channels. In the current implementation, FusionMAE relies on externally provided anomaly or corruption labels to identify problematic input channels, which are then explicitly masked during inference.

We have explored the possibility of using reconstruction error from FusionMAE as a signal for automatic anomaly detection. However, we note that using the same model both to identify corrupted inputs and to impute them can be engineering-wise unstable and difficult to validate, particularly in safety-critical fusion applications. For this reason, we have not claimed that FusionMAE can reliably improve corrupted inputs without external supervision. In parallel, the HL-3 program is developing independent anomaly detection systems based on statistical methods and machine learning. These systems are intended to operate upstream of FusionMAE, providing robust channel-quality flags. FusionMAE is expected to cooperate with, rather than replace, such dedicated data-quality monitoring modules.

11. As a comment, it seems that the reconstructed signals always improve the performance of the model carrying out the downstream task.

Response:

We thank the reviewer for this observation. We would like to clarify that, while reconstructed signals improve performance in most evaluated cases, this behavior is not universal. We observed an unexpected performance degradation (gray bar) in one configuration of the plasma evolution prediction (PPR) task.

We have explicitly highlighted this exception and added a discussion of this anomaly in the revised manuscript (Methods, *Downstream task: plasma evolution prediction*). This discussion analyzes possible contributing factors and clarifies the conditions under which FusionMAE may or may not be beneficial.

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. 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.

12. The applicability of FusionMAE to devices beyond HL-3 should be discussed. While fine-tuning on additional pulses is likely necessary for new machines, differences in operational space and diagnostic sets may further increase data requirements. In this sense, labelling FusionMAE as a "foundation model" may be overstated. A more measured framing would position it as a promising framework for data-driven analysis in fusion research.

Response:

We thank the reviewer for this important and constructive comment. We agree that, at its current stage, labeling FusionMAE as a *"foundation model"* would be overstated. While developing more generalizable pretrained models for fusion diagnostics is a long-term research goal, the present work does not yet meet the scope or maturity implied by that term. Accordingly, we have revised the manuscript to consistently describe FusionMAE as a device-specific pretrained model and have removed or rephrased language that could suggest broader claims.

We have also conducted preliminary exploration experiments applying FusionMAE to data from another device, SUNIST-2, to assess the generality of the framework. These results, now summarized in Extended Fig.13 and section of '*Application of FusionMAE in a spherical tokamak: SUNIST-2*', provide an initial indication that the pretraining-finetune paradigm itself may be transferable, even though performance remains strongly dependent on device-specific data. We would try few-shot and zero-shot cross-device adaptation in the future, which is much more challenging but quite valuable for current fusion community.

13. There are a few typos: “chanllenges” on Page 5, “whose detailed hyper parameters keeps the same as in [27,28]”, where I suggest using “are” in place of “keeps”.

Response:

We thank the reviewer for pointing out these typos. We have carefully checked the entire manuscript and corrected these issues accordingly in the revised version.

Dear Reviewer,

We sincerely thank you for taking the time to review our manuscript. Your comments demonstrate a precise understanding of our work and highlight critical issues that are essential for strengthening the paper.

Our original motivation was to explore a data-driven framework inspired by the concept of foundational models for fusion research. Given the breadth of capabilities such a framework aims to address, together with the strict length constraints of Nature-series journals, some intermediate results and supporting analyses were not sufficiently presented or emphasized in the initial submission. This may have limited the clarity and completeness of the technical justification in certain parts of the manuscript.

Both you and another reviewer raised important concerns in this regard. In response, we have substantially expanded the experimental evaluations and further refined the technical implementation. We believe that these revisions significantly strengthen the rigor and clarity of the work, and we are grateful for your constructive feedback, which has greatly improved the quality of the manuscript.

Below we provide point-by-point responses to your specific comments and suggestions. The original comments are shown in blue, our responses are shown in black, and the corresponding revisions in the manuscript are highlighted in yellow.

Reviewer #2 (Remarks to the Author):

Summary:

The paper proposes FusionMAE, a transformer-based masked autoencoder that compresses 88 HL-3 tokamak diagnostic signals into a 256-dimensional "plasma status embedding" and claims three "emergent" capabilities—automatic analysis of secondary signals, a universal interface for downstream tasks, and robustness to missing diagnostics—together with a "virtual backup diagnosis" mode with ~0.97 Pearson correlation for reconstructed channels. The idea is timely and potentially useful, but several methodological choices and the strength of evidence fall short of the paper's "foundation model" framing and its more ambitious claims.

Major claims

1. Unified embedding across diagnostics. A masked-autoencoding framework compresses 88 signals (12 systems) sampled at 1 kHz over 10 ms windows into a 256-D vector, from which the decoder reconstructs original and missing channels.

2. "Virtual backup diagnosis." With 25% channel masking in training, the model reconstructs masked channels with average PCC of 0.967 (described as 96.7% reliability), and compress-then-reconstruct PCC of 0.986.

3. Emergent capabilities.

a. Automatic analysis of secondary signals (e.g., EFIT-derived shape parameters) by reconstructing them when masked.

b. All-purpose vector that improves or matches three downstream tasks—disruption

prediction (DPR), equilibrium surrogate (EFIT-NN), and plasma evolution prediction (PPR)—relative to baselines trained on raw signals.
c. Robustness to missing diagnostics (stable outputs with 5–20% masked inputs).
4. Practical impact claims: The embedding could simplify interfaces, reduce the number of diagnostics, and enhance control performance, positioning FusionMAE as a domain-specific “foundation model” for fusion.

Novelty and Broader applicability:
1. Bringing masked-autoencoding to multi-diagnostic fusion time series and explicitly framing the output as an all-purpose embedding for control/analysis is a useful systems-level idea that goes beyond one-task predictors prevalent in fusion ML. The demonstration spans three representative tasks, which is broader than much prior single-task works.
2. The architecture is a generic MAE-style encoder-decoder with masking from, previous works. Against the current wave of foundation models for time series and long-context sequence models, a 256-D latent trained on 10 ms windows over a single device is modest in scale and scope. The “emergent capabilities” largely follow from the objective (masked channel reconstruction), so characterizing them as “emergence” overstates what is essentially learned cross-channel regression.
3. The fusion community will likely be interested in imputation/robustness to missing diagnostics and a single learned interface that can feed multiple modules. Broader ML interest is more limited unless the paper offers stronger methodological innovations (architecture, training recipes, scaling laws) or cross-device generalization demonstrating out-of-domain transfer.
Response:
We thank the reviewer for the positive assessment and insightful understanding of the technical architecture and core ideas of this work. Regarding the question of how this study may stimulate broader interest from the machine learning community, we view this work as a starting point rather than a conclusive answer.
Prior AI-for-fusion studies have rarely introduced the pretraining-finetune paradigm as a systematic framework within the fusion domain, and the adoption of recent advances from the machine learning community, especially on the topic of LLMs. In this work, we draw inspiration from early large-scale representation learning approaches, such as those exemplified by GPT-1 and BERT, and adapt these ideas to fusion diagnostics in a device-specific and task-driven manner. The encouraging performance observed across multiple downstream fusion tasks suggests that this paradigm can be practically effective in this domain.
From this perspective, we believe the work may help lower the barrier for incorporating more advanced machine learning techniques into fusion research, and thereby provide a point of engagement for the broader ML community, without claiming generality beyond the demonstrated scope.
We are also actively exploring extensions along several directions, including cross-device adaptation, multimodal pretraining, and autoregressive modeling. These directions involve substantial technical challenges and are therefore left for future work.

1. Reporting a single PCC per channel averaged across shots is encouraging, but PCC is generally not equivalent to reliability in an operational sense. There is no calibration analysis, per-channel error distribution, or failure analysis under realistic missing-data patterns (most fusion diagnostics are missing not at random).

Response:

We thank the reviewer for this important clarification. We have revised the evaluation throughout the manuscript to adopt a combined strategy, reporting both PCC and RMSE to jointly reflect correlation structure and absolute error. These updates have been applied consistently in Figures 3, 12, and 13, as well as Tables 4, 5, and 6, to provide a more balanced and informative assessment of model performance.

To better probe robustness under missing or corrupted data, we expanded the data-completion evaluation to include two complementary masking scenarios: single-channel masking, and masking of an entire diagnostic group. These tests are intended to approximate different classes of realistic failure modes and to go beyond idealized random masking assumptions.

2. The training/eval masking is channel-level and random, while actual outages have structure (e.g., all channels are missing for a time interval). Stress tests with structured masking (by subsystem or time interval) and campaign-wise OOD splits would make the claim more credible.

Response:

We thank the reviewer for this suggestion. To move beyond purely random, channel-level masking, we have added structured masking tests in Figure 3, including single-channel masking and masking of entire diagnostic subsystems. These tests yield informative contrasts: for example, magnetic diagnostics, which contain many channels and exhibit high internal redundancy, are well reconstructed under single-channel masking, but their advantage largely disappears when the entire diagnostic is masked, leading to performance comparable to other subsystems.

In addition, Figure 12 reports evaluation results on data from the most recent HL-3 experimental campaign, providing a campaign-wise out-of-distribution (OOD) test. As expected, this setting introduces a performance degradation, but the model remains at a practically usable level.

3. The experiment masks all secondary channels then reconstructs them. That shows the model can regress secondary variables from primaries. This is valuable, but because those same secondary channels are part of the pretraining target, the result is not out-of-distribution; it is a standard masked-target recovery. A stronger test would:

a. Exclude specific secondary channels from the reconstruction loss during pretraining and evaluate zero-shot recovery;

b. Compare against supervised surrogates trained only on primaries.

Response:

We thank the reviewer for this clarification. We agree that describing this behavior as an “emerging capability” was inappropriate, and we have consistently revised the word to refer to these results as downstream applications of the pretrained model.

Regarding point (a), if the reconstruction loss corresponding to specific secondary channels is removed during pretraining, the model has no remaining supervisory signal from which to learn the mapping from primary diagnostics to those secondary variables. Under this setting, zero-shot recovery of the excluded secondary channels is therefore not feasible in principle.

Regarding point (b), we have added a direct comparison with supervised surrogate models trained only on primary diagnostics in Table 5.

4. The summary bars suggest improvements when using the embedding or completed data, but no confidence intervals or statistical tests are provided. This would be crucial to judge if gains are significant.

Response:

We thank the reviewer for this suggestion. To better quantify the robustness of the observed improvements, we extended the downstream evaluations using a model ensemble strategy and repeated each experiment five times. The standard deviation across runs is now reported as error bars in Figure 6.

5. The baselines are not fully matched for capacity/regularization/data preprocessing, and there is no comparison to simpler representation baselines.

a. Provide: 95% Credible Intervals; tests for AUC differences in DPR; paired permutation tests for PCC.

b. Stronger baselines (e.g., a TCN/Transformer/VAE embedding without masking; PCA).

c. Ablations on mask ratio, latent size, depth/heads

d. The main text says the model is trained to minimize MSE , but Methods Eq. (3) uses L1 with weights. Which objective was actually used? Provide ablations for L1 vs. L2 and for the W schedule.

Response:

We thank the reviewer for this detailed set of suggestions. In response to concerns raised by this reviewer and another, we substantially expanded the experimental analysis, running over 300 additional experiments to better probe robustness, baselines, and hyperparameter sensitivity.

a. For the DPR, EFIT-NN, and PPR downstream tasks, we repeated each experiment five times and reported the standard deviation as error bars in the corresponding figures. These results support the original conclusions while providing a clearer view of performance variability.
b. Regarding baselines, we have added more explicit descriptions of all downstream models and their configurations. Importantly, FusionMAE is introduced only as a replacement for the input representation; the downstream architectures, training procedures, and data preprocessing are kept identical to those used in prior work. This design choice is now clearly stated in the manuscript to ensure fair comparison.
c. We have also added ablation studies on mask ratio, latent size, and model

depth/heads, summarized in Figure 8, along with a discussion of the observed trends. These results provide additional insight into the behavior and sensitivity of the pretrained model.

d. Finally, we note that the discrepancy between the loss description in the main text and Eq. (3) was a typo. The model is trained using an L2 (MSE) loss, and this has been corrected throughout the manuscript. Figure 8 also provides the result of different settings for W to support the final design of model hyper-parameters.

We have revised the paper accordingly to clarify these points and strengthen the experimental evidence.

6. Uncertainty bands by randomly masking 5–20% of channels is informative but it would be more informative to consider deep ensembles, MC-dropout, heteroscedastic/quantile heads, and calibration curves to produce calibrated predictive intervals

Response:

We thank the reviewer for this suggestion. In the revised manuscript, we have incorporated deep ensemble-based error bars for most statistical performance evaluations to better reflect model variability.

However, for the single-shot prediction examples in Figure 7, the uncertainty bands generated by masking 5–20% of input channels are intended to illustrate robustness under partial diagnostic loss, rather than calibrated predictive uncertainty. Combining this visualization with deep-ensemble uncertainty would likely conflate these two distinct effects and reduce interpretability.

We have clarified this distinction in the revised text to avoid any implication that the masking-based bands represent calibrated uncertainty estimates.

7. All experiments are on HL-3. To support claims about influencing reactor design (e.g., fewer diagnostics), cross-tokamak transfer (e.g., zero-/few-shot to another tokomak) and closed-loop control studies (in simulation first) would be essential.

Response:

We thank the reviewer for this important clarification. We agree that our original wording regarding “influencing reactor design” was insufficiently precise, and we have revised it accordingly.

In the revised manuscript, we clarify that the intended implication is not the elimination of diagnostics through cross-device generalization, but rather the use of FusionMAE as a virtual backup diagnostic when existing channels temporarily fail. In this scenario, the goal is to reduce the need for additional hardware redundancy, rather than to rely on zero-/few-shot transfer across tokamaks. Under this application setting, cross-device transfer is not a prerequisite.

For example, in the section of Discussion:

Notably, FusionMAE achieves high-fidelity data imputation by leveraging its learned understanding of inter-parameter physical relationships, which could be used to relax the redundancy requirement for diagnostic systems for future fusion reactors.

We still agree that zero-/few-shot cross-device adaptation is an important and challenging research direction. We consider this a long-term objective and are actively exploring it, but we do not claim success in this regime in the present work. We have also conducted preliminary experiments on SUNIST-2, using the same code framework with device-specific retraining, to verify the generality of the proposed methodology and to lay groundwork for future cross-device studies.

Regarding closed-loop control, HL-3 currently lacks a mature, control-oriented simulation platform. Historically, control studies have therefore relied on data-driven simulation models. In this work, such models are included as downstream tasks, and we validate their performance when driven by FusionMAE-reconstructed signals. Future work will explore more direct control-in-the-loop studies, including experimental proposals to intentionally disable selected diagnostics and rely on FusionMAE-assisted operation, subject to machine scheduling constraints.

We have revised the manuscript to reflect this more precise and appropriately scoped interpretation.

8. Given the scope of the results, the authors should consider tempering the “foundation” framing and present FusionMAE as a practical multi-diagnostic representation that improves fault tolerance to missing sensors and simplifies interfaces across analytics and control modules.

Response:

We thank the reviewer for this suggestion and agree with this assessment. We acknowledge that the term *“foundation model”* overstates the current scope of FusionMAE. In the revised manuscript, we consistently reframe FusionMAE as a device-specific pretrained model that provides a practical multi-diagnostic representation, improving robustness to missing sensors and simplifying data interfaces across downstream analytics.

We note that our current positioning is closer in spirit to early pretrained representation models (e.g., GPT-1 or BERT), rather than models exhibiting broad zero-shot generalization. References to “foundation” capabilities have therefore been removed. Extending FusionMAE toward a more general foundation-style model remains a longer-term research direction and is not claimed in this work.

9. Tokenization: The methods section say that each column is treated as a patch, encapsulating plasma information from a single diagnostic channel even though the input is 88×10 (channels × time) and earlier text states each channel is projected. Please clarify whether tokens correspond to channels (88 tokens) or time slices (10 tokens), and how time information beyond 10 ms is handled.

Response:

We apologize for the confusion in the manuscript. Each token corresponds to a channel (88 tokens), and we have corrected the description accordingly.

The revised version is:

It processes an input consisting of an 88×10 array, where each row, a 10 ms time window of signals sampled at 1 kHz, corresponds to one of the 88 diagnostic channels. Within the Transformer framework, each row is treated as a patch, encapsulating plasma information from a single diagnostic channel.

In the current model, temporal dependencies beyond the 10 ms window are not directly modeled; instead, the plasma state is segmented into independent time windows. This design does impose limitations on the model's temporal scope, as reflected in the patch-length analysis shown in Fig.8. Nevertheless, the model already demonstrates strong representational capability within each window. Extending the framework to handle longer temporal sequences would likely require a shift to an autoregressive pretraining paradigm, which represents a substantial undertaking and will be explored in future work. We have appended an analysis after Fig.8:

In its current form, FusionMAE does not incorporate a dedicated architectural treatment for temporal dynamics. As a result, model performance degrades noticeably when the patch length becomes too long. The experiments in Fig.8g indicate that this transition occurs at approximately 10 ms. Addressing this limitation would likely require, on the one hand, augmenting the base Transformer with mechanisms tailored for time-series modeling, and on the other hand, potentially reformulating the pretraining framework in an autoregressive manner.

10. Most transformer variants for sequences use pre-norm LayerNorm/RMSNorm; BatchNorm can be brittle for variable-length sequences and distribution shift. Please justify this design and report comparisons to LayerNorm.

Response:

Thank you for the comment. Unlike typical NLP tasks, our sequences have a fixed length of 88 tokens, which makes using BatchNorm feasible. We conducted experiments comparing BatchNorm and LayerNorm and observed comparable performance. To align with common practices in large-scale transformer models, we have adopted LayerNorm in the revised version.

11. Please include the justification for choosing the 10ms window as this might miss slower physics, and might be too short for many control relevant processes

Response:

Thank you for the comment. In the current model, temporal dependencies beyond 10 ms are not explicitly modeled. Longer physical events are therefore segmented into multiple independent samples, processed separately, and subsequently reassembled into a continuous waveform. This approach is sufficient for most scenarios considered in this work, but it does limit the model's ability to capture long-timescale physical phenomena. Consistently, the parameter scan in Fig. 8 shows that simply increasing the window length does not substantially improve long-sequence modeling for FusionMAE, which lacks an explicit temporal inductive bias. We expect that a more effective treatment of long temporal dependencies would require reformulating the pretraining strategy in an autoregressive manner.

We have appended an analysis after Fig.8:

In its current form, FusionMAE does not incorporate a dedicated architectural treatment for temporal dynamics. As a result, model performance degrades noticeably when the patch length becomes too long. The experiments in Fig.8g indicate that this transition occurs at

approximately 10 ms. Addressing this limitation would likely require, on the one hand, augmenting the base Transformer with mechanisms tailored for time-series modeling, and on the other hand, potentially reformulating the pretraining framework in an autoregressive manner.

12. Extended Table 1 shows low health rates for some key channels (e.g., ne, Te, Prad). Please evaluate FusionMAE under device-realistic outage patterns (e.g., "all interferometers down") rather than random MCAR masks

Response:

Thank you for the suggestion. We have added a per-diagnostic masking mode in Fig. 3 to evaluate FusionMAE under more device-realistic outage patterns. As expected, this scenario is more challenging than per-channel masking, but the model maintains usable performance across the evaluated tasks.

13. Number of epochs/steps, training wall-clock, hardware, and parameter count are not reported; early stopping is 20 epochs but the total training regime is unclear. Please provide full reproducibility metadata.

a. parameter count, training steps/epochs, and FLOPs;

b. Provide scaling curves vs. latent size/depth and mask ratio to justify the choice of 256-D latent and 25% masking

c. Include throughput/latency measurements to support control-room use

Response:

Thank you for the detailed request regarding reproducibility. We give the information in the revised version of '*Training of FusionMAE*' section:

The final FusionMAE model contains approximately 14 million trainable parameters. Training requires about 10 hours on a single NVIDIA A100 GPU, while inference for a single time slice takes 4.8 ms, enabling quasi-realtime application when combined with a 10 ms input time window.

In addition, Fig. 8 now includes a systematic parameter scan over latent dimension, model depth, and masking ratio, which motivates the choice of a 256-dimensional latent space and 25% masking.

The number of training epochs was not fixed across all runs but varied depending on initialization and early-stopping behavior. We therefore report the typical range of epochs observed across repeated experiments in Table 3 to provide a reference for reproducibility.

14. Using PCC for reconstruction and for EFIT-NN/PPR is standard but insufficient: PCC is insensitive to scale/offset. Add NRMSE/MAE and relative error vs. dynamic range per channel. For DPR, also report precision–recall AUC.

Response:

Thank you for the suggestion. We have expanded the evaluation metrics accordingly. For the EFIT surrogate (EFIT-NN) and plasma profile reconstruction (PPR) tasks, we now report both PCC and RMSE to complement correlation-based assessment with scale-sensitive error measures, as in Table 5 and 6. For disruption prediction (DPR), we additionally report TPR, TNR, and AUC to provide a more complete characterization of classification performance, as in Table 4.

15. The dataset spans four campaigns (2022–2025) with a single random split (1,950/495 shots). To assess robustness to drift, add chronological splits (train on early campaigns, test on later campaigns).

Response:

Thank you for the suggestion. We agree that chronological splits are important for assessing robustness to temporal drift. To avoid extensive reorganization of the dataset at this stage, we instead collected data from the most recent HL-3 experimental campaign and constructed an additional held-out test set. The corresponding results are reported in Fig.12.

16. Signals are resampled to 1 kHz, but normalization procedures (per-channel z-score? robust scaling?) are not fully specified. Extended Table 1 lists means/SDs, yet preprocessing pipelines (filtering, detrending, synchronization and handling physically “a.u.” channels) should be described step-by-step.

Response:

Thank you for the comment. The preprocessing pipeline is very simple. After resampling all signals to 1 kHz, we apply per-channel Z-score normalization using the mean and standard deviation reported in Extended Table 1. No additional filtering, detrending, or robust scaling is performed, including for channels reported in arbitrary units. We have clarified this in the Dataset section of the revised manuscript.

Finally, Z-score normalization is adopted for data preprocessing in FusionMAE, where each input channel is normalized using the mean and standard deviation estimated from the training set.

17. Code/data are “available on request” rather than open, which limits reproducibility and community use.

Response:

We acknowledge this concern. Full public release of the experimental data is constrained by institutional and management policies. At present, the code can be made available through direct communication and discussion, subject to these constraints, while broader data release remains challenging.

On the other hand, we have described the technical methodology, model architecture, training procedure, and evaluation protocols in detail, and the approach has been validated across multiple devices. We believe this level of methodological transparency supports reproducibility of the proposed framework, even where direct data sharing is limited.

18. A few typos (“chanllenges”, etc.) need to be corrected as well as ensure consistent units and symbols

Response:

We thank the reviewer for pointing out these typos. We have carefully checked the entire manuscript and corrected these issues accordingly in the revised version.

FusionMAE: self-supervised pretrained model to optimize and simplify diagnostic and control of fusion plasma

Author Information

Affiliations

1. **Southwestern Institute of Physics, Chengdu, China**
2. **Tsinghua University, Beijing, China**
3. **Nankai University, Tianjin, China**
4. **Zhejiang University, Hangzhou, China**

Zongyu Yang[1], Zhenghao Yang[1&2], Wenjing Tian[1&2], Jiyuan Li[1], Xiang Sun[1], Guohui Zheng[1&3], Songfen Liu[3], Niannian Wu[1&4], Rongpeng Li[4], Zhaohe Xu[1], Bo Li[1], Zhongbing Shi[1], Zhe Gao[2], Wei Chen[1], Xiaoquan Ji[1], Min Xu[1] & Wulyu Zhong[1]

Contributions

Zongyu Yang contributed by proposing the idea, designing the algorithm, conducting coding and testing, exploring the downstream tasks, discussing the experimental results, adapting the model to SUNIST-2, writing and revising the manuscript.

Zhenghao Yang contributed by performing coding and testing and preparing key result figures.

Wenjing Tian contributed by performing coding and testing, reorganizing the experimental results from perspective of large-scale pretrained model, adapting the model to SUNIST-2, and writing the manuscript.

Jiyuan Li contributed by exploring the downstream task of plasma evolution prediction and adapting the model to SUNIST-2.

Xiang Sun contributed by exploring the application of FusionMAE on accessing data quality.

Guohui Zheng contributed by exploring the downstream task of plasma equilibrium fitting.

Songfen Liu contributed by exploring the downstream task of plasma equilibrium fitting.

Niannian Wu contributed by exploring the downstream task of plasma evolution prediction.

Rongpeng Li contributed by exploring the downstream task of plasma evolution prediction.

Zhaohe Xu contributed by exploring the downstream task of disruption prediction.

Bo Li contributed by mentor the development of three downstream algorithms.

Zhongbing Shi contributed to diagnostic systems.

Zhe Gao contributed by writing the manuscript and discussing the experimental results.

Wei Chen contributed to the analysis.

Xiaoquan Ji contributed to the experiments.

Min Xu contributed by writing the manuscript and discussing the experimental results.

Wulyu Zhong contributed by writing the manuscript, discussing the experimental results, and managing project progress.

Corresponding author

Correspondence to: Zhe Gao(gaozhe@tsinghua.edu.cn), Wulyu Zhong(zhongwl@swip.ac.cn)

Abstract

In magnetically confined fusion device, the complex, multiscale, and nonlinear dynamics of plasmas necessitate the integration of extensive diagnostic systems to effectively monitor and control plasma behaviour. The complexity and uncertainty arising from these extensive systems and their tangled interrelations has long posed a significant obstacle to the acceleration of fusion energy development. In this work, a self-supervised model, fusion masked auto-encoder (FusionMAE) is pre-trained to compress the information from 88 diagnostic signals into a concrete embedding, to provide a unified interface between diagnostic systems and control actuators. Two mechanisms are proposed to ensure a meaningful embedding: compression-reduction and missing-signal reconstruction. Upon completion of pre-training, the model acquires the capability for 'virtual backup diagnosis', enabling the inference of missing diagnostic data with 97.2% reliability. Furthermore, the model demonstrates multiple downstream applications: automatic data analysis, universal control-diagnosis interface, and enhancement of control performance on multiple tasks. This work pioneers large-scale AI model integration in fusion energy, demonstrating how pre-trained embeddings can simplify the system interface, reduce requirement for redundant diagnostic systems and optimize operation performance for future fusion reactors.

Introduction

Fusion energy has long been regarded as a clean and sustainable power source, with the potential to meet the rapidly growing global energy demand while addressing carbon neutrality. Among the various concepts, tokamak stands out as one of the most promising designs for realizing the first commercial fusion reactor, having achieved remarkable milestones in recent years. For instance, the Joint European Torus set a world record by producing 59 megajoules of fusion energy over a 5-second pulse[1]. The Korea Superconducting Tokamak Advanced Research sustained plasma with ion temperatures exceeding 100 million kelvin for 30 seconds[2]. And the EAST and WEST tokamak realized a steady state plasma lasting up to 1,000 and 1320 seconds, respectively[3]. Meanwhile, the International Thermonuclear

Experimental Reactor (ITER), one of the world's largest science projects, involving the collaboration of more than 30 nations, is currently under assembly to demonstrate the scientific feasibility of tokamak reactors[4].

Despite the remarkable achievements, the path to a Fusion Power Plant (FPP) is still hindered by several physical and engineering challenges[5]. Controlling and optimizing fusion plasma, a non-linear magnetohydrodynamic system, requires the integration of highly intricate, multidisciplinary subsystems[6]. A wide range of diagnostic systems have been developed to measure plasma properties, including current, temperature, density, rotation, and radiation [7]. Additionally, numerous simulation modules have been created to model plasma behaviour across different time scales and spatial regions[8]. Furthermore, a significant number of control actuators have been designed to regulate plasma properties[9]. For instance, ITER incorporates roughly 50 distinct plasma diagnostic systems and 70 sets of control actuators[10]. Moreover, the highly complex Integrated Modelling & Analysis Suite (IMAS) system has been established to support experimental data analysis and simulation program execution[11]. In parallel, various control loops are formed through combining different modules, working in coordination to achieve the overall operation of the fusion reactor[12]. Therefore, while the scientific feasibility of fusion energy has been validated, the engineering complexity of fusion power plant remains prohibitively high, as reflected in the extremely long timeline of ITER[13].

In recent years, artificial intelligence (AI) technologies have been widely applied in computer vision, natural language processing(NLP), and diverse scientific fields such as biology, materials science, and meteorology[14-18]. AI has proven its capability in providing end-to-end, streamlined solutions for complex problems. In the fusion domain, AI has also demonstrated numerous successful applications, including predicting major plasma disruptions[19, 20], automating the realization of complex plasma magnetic configurations[21], and achieving kinetic control to avoid plasma instabilities[22]. While these specialized models have made valuable contributions to specific challenges in fusion research, particularly in localized decision-making and control of certain plasma characteristics, their application scope remains limited. These models primarily focus on understanding specific aspects of plasma physics related to the downstream tasks, which may not provide a comprehensive grasp of the underlying principles. In contrast, a new paradigm has been demonstrated by large language models (LLMs). By developing a pretrained model that comprehensively understand plasma physics and use fine-tune, post-train or prompt engineering techniques to turn the pretrained model into task-specific scenarios, the performance on downstream tasks could be further ~~promoted.~~ Furthermore, within the base model-downstream task framework, solutions to various downstream tasks can be unified into a common algorithmic workflow. This approach would significantly simplify the control systems of fusion devices and allow for more efficient utilization of diagnostic information. This unified framework could substantially improve the overall efficiency of fusion device operation, enhancing both control precision and the optimization of key plasma characteristics.

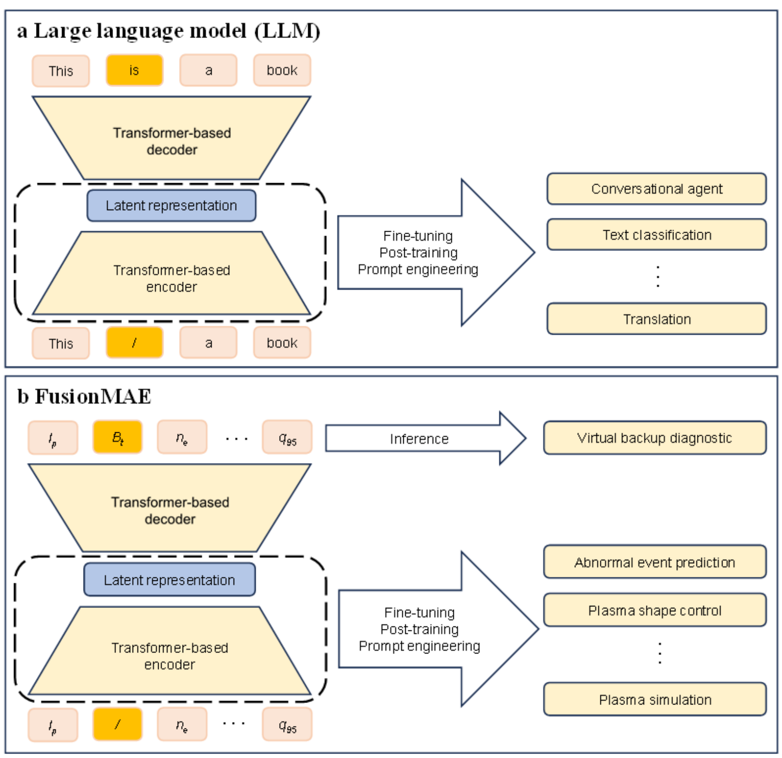


Fig.1: Comparison between the pre-training and downstream application frameworks of FusionMAE and natural language processing (NLP) large-scale models.

In this work, we present a pretrained model named the Fusion Masked Auto-Encoder (FusionMAE), which compresses data from multiple diagnostic systems into a unified plasma status embedding, based on a dataset from the largest tokamak currently operated in China, HL-3[23, 24]. This embedding allows a Transformer [25] decoder to restore all the raw diagnostic data and reconstruct possible missing data from it, showing a successful~~ly~~ integration of various diagnostic signals. Analogous to word embeddings in NLP[26], the plasma state embedding naturally clusters discharges according to similarities in operational scenarios. FusionMAE demonstrates extensive utility as a foundational model for HL-3, including leveraging its original capacity to impute missing diagnostic data for the development of virtual backup diagnostics, as well as adapting its masking mechanism to enable automated secondary data analysis. Consistent with our expectations for a pretrained model, it can support multiple downstream tasks, disruption prediction (DPR), equilibrium reconstruction (EFIT-NN) and plasma evolution prediction (PPR) [27-29]. The performance of these task-specific models can all benefit from FusionMAE and keep steady even with the failure of several input diagnostics. Fig.1 compares the pre-training and downstream application frameworks of FusionMAE and LLMs. It demonstrates that FusionMAE is positioned as a device-specific pretrained model designed for fusion experimental data, holding significant potential to evolve into a general-purpose agent tailored to the fusion research.

Results

FusionMAE: a Transformer-based model pretrained by self-supervised learning task

FusionMAE is developed to compress 88 signals from 12 systems of HL-3, as shown in Fig. 2a, into a unified embedding. The input consists of plasma current (I_p), plasma shape and position (R, Z, a, κ, δ_u, δ_l), toroidal field (B_t), poloidal field (B_p), loop voltage (V_{loop}), electron density (n_e), electron temperature (T_e), stored energy (W_e), radiation power (P_{rad}), soft-X radiation (SX), impurity radiation(A_{imp}), D-α spectrum (D_α), amplitude of magnetohydrodynamic instability modes ($A_{MHD,\,1}$, $A_{MHD,\,2}$), normalized beta(β_N), safety factor(q_{95}), inertial inductive(l_i), poloidal field coil current (I_{PF}) and auxiliary heating power (P_{NBI}, P_{EC}, P_{LH}). Detailed descriptions of these signals are provided in the Methods section. These systems span a broad spectrum of physical parameters and can be considered as a representative subset of the systems required by future fusion reactors like ITER[10].

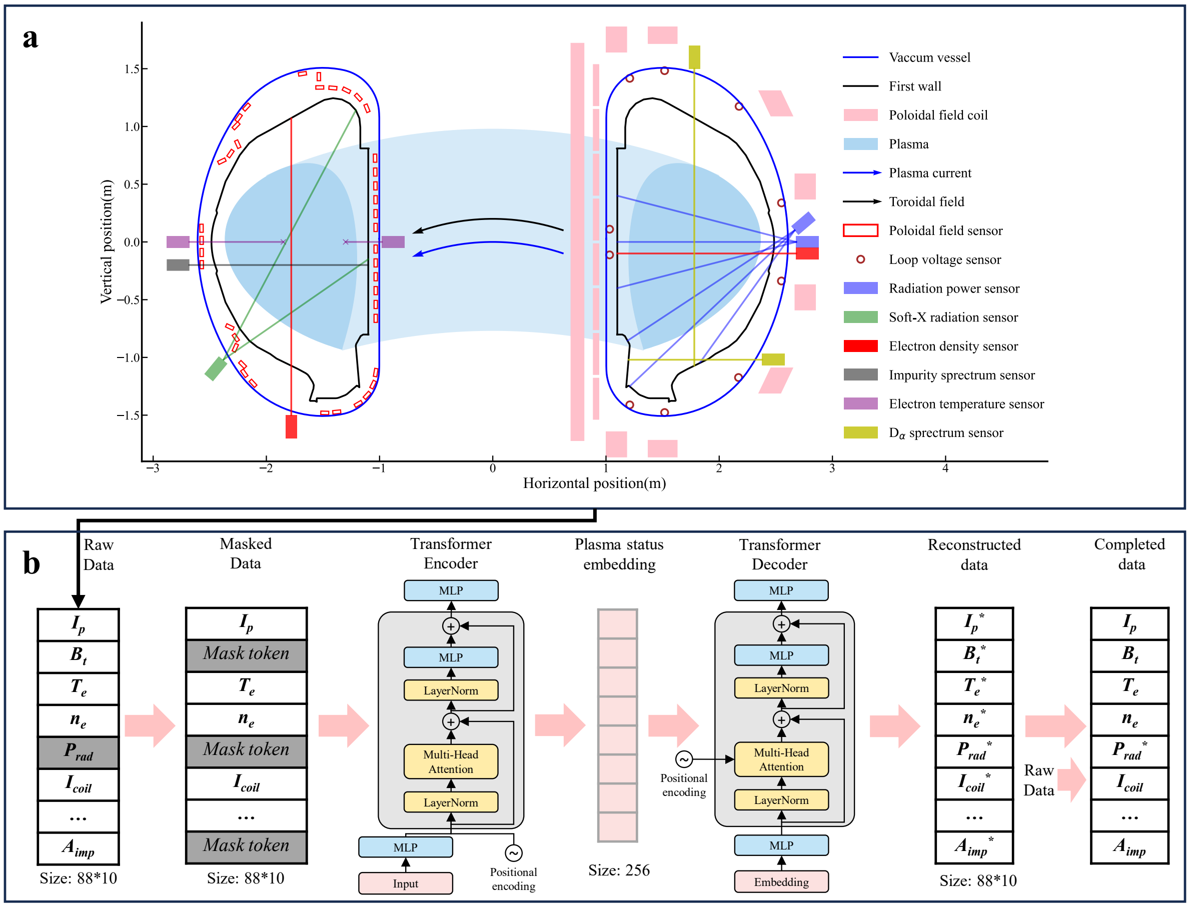


Fig.2: Overall architecture of FusionMAE.

a Diagram of the diagnostic systems employed in this study. The auxiliary heating systems are omitted due to its complexity, while the details can be found in [30]. Additionally, some secondary parameters, like A_{MHD}, W_e, β_N, q_{95} and l_i are computed from a set of primary data using dedicated programs, as described in [31, 32].

b Workflow and neural network structure of FusionMAE.

Fig.2b illustrates the workflow and neural network architecture of FusionMAE. It takes a 10ms time window of the 88 signals sampled at 1kHz as input, forming an 88×10 array. Firstly, each channel of the array is projected into a 128-dimensional space using a multilayer perceptron (MLP), and a positional encoding vector[33] is added to encode the channel index. Secondly, the resulting 88×128 array is fed to an encoder composed of several Transformer blocks, each consisting of multi-head self-attention layers[34], MLPs, layer-normalizations[35], and residual connections[36]. The encoder takes the main responsibility to learn underlying patterns of plasma status. The output of the Transformer encoder is further embedded by an MLP layer into a vector of length 256, which is the expected plasma status embedding. The decoder has a mirrored structure, where several Transformer blocks reconstruct the original data from the embedded vector. The only difference is that the positional embedding is not added and instead serves as the query vector in the final multi-head attention layer. The entire encoder-decoder structure is trained to minimize the mean squared error between the output and input signals.

Unified plasma status embedding

FusionMAE extracts a meaningful embedding through two key mechanisms. Firstly, the embedded vector is designed to be smaller than the input array to force effective compression. To preserve information, the model must capture the intrinsic low-dimensional manifold of plasma dynamics. As shown in Fig.3a, a black radar plot visualizes the Pearson correlation coefficients (PCC) between the reconstructed and original signals across all 88 channels. The averaged PCC for all channels is 98.7%, indicating that the embedding retains nearly all information contained in the raw diagnostic data, while providing a more compact and structured representation.

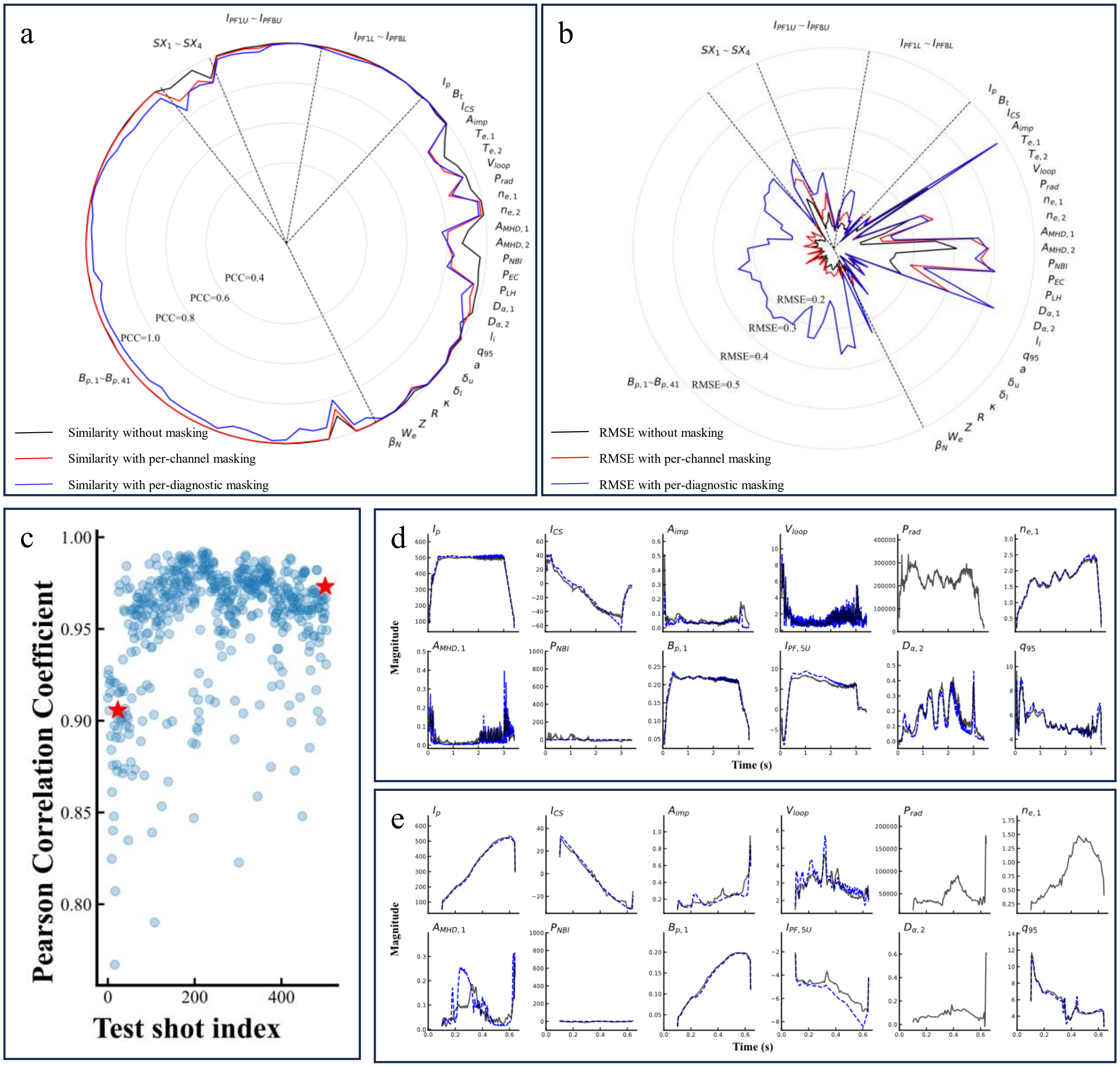


Fig.3: Evaluation of FusionMAE's performance in restoring compressed raw signals and reconstructing missing signals.

a Radar plot illustrating the PCC between original signals and FusionMAE output across 88 channels. Gray radar plot is tested without masking, and the loss of information is only brought by the compression. Red radar plot is tested by intentionally masking one of the 88 signals, which indicates FusionMAE's reliability to restore missing signals. Blue radar plot is a more ~~stressed testing~~ with all signals from a same diagnostic subsystem masked together.

b Radar plot illustrating the normalized RMSE between original signals and FusionMAE output.

c Scattered plot of averaged PCC with per-channel masking across 88 channels against the test shot index. The shot index is ordered by the date and time of discharge. Red stars are representative testing shots plotted in subfigure d and e.

d Comparison between reconstructed signals (black solid lines) and raw signals (blue dotted lines) for Shot 8909.

e Comparison between reconstructed signals (black solid lines) and raw signals (blue dotted lines) for Shot 1430.

However, when the model is solely trained to compress and restore data, it tends to memorize signals on a channel-by-channel basis and struggles to capture the interconnections between channels. To address this limitation, a more challenging task has been devised. In each training step, a random 25% of the input channels are masked, forcing the model to reconstruct the channels based on its understanding of inter-channel relations. To validate FusionMAE's performance under this situation, two testing experiments are conducted. Firstly, we intentionally mask one of the channels, calculate the PCC between restored and original signals, and repeat the process across all the 88 channels. Secondly, we mask the signals from a same diagnostic or control subsystem together, calculate the PCC and repeat the process on 12 subsystems, which is a more ~~stressed testing.~~ FusionMAE successfully achieved average PCCs of 97.2% and 94.5%, which is presented in Fig.3a by blue and red radar plots. FusionMAE shows a gradually declining performance in the three testing modes, without masking, with per-channel masking and with per-diagnostic masking, as the input information goes more incomplete. For B_p signals, the PCC is very close to 100% with per-channel masking, showing that the corresponding diagnostic is highly redundant and is not sensitive to the missing of a few signals. But there is an obviously larger performance degradation with per-diagnostic masking for B_p, compared to other diagnostics, which shows that the reconstruction of these signals will be as difficult as others when all the B_p diagnostics fail. Fig.3b shows three radar plots for Rooted Mean Squared Error (RMSE) calculated on normalized data, which shows very similar results with Fig.3a.

Fig.3c shows the averaged PCC across the 88 channels for each test shot. For most of the shots, the averaged PCC is higher than 95%. Shots in early campaigns have a relatively lower PCC distribution, as the control system of HL-3 is still under development and the discharge is not very stable in this stage. Most shots with PCC lower than 90% ~~is~~ those that terminate~~s~~ early in ramp-up phase and ~~doesn't~~ establish a steady plasma status, which may challenge the feasibility of FusionMAE. Fig.3d and Fig.3e shows examples of data reconstruction for Shot 8909 and Shot 1430, as representative cases with relatively lower and higher performance.

The plasma state embedding extracted by FusionMAE plays a role analogous to word embeddings in NLP. Represented by a 256-length vector, it encapsulates the intrinsic physical state of the plasma. Each diagnostic system provides a partial observation of this state, which is transformed into output either through the decoder in digital space, or by physical diagnostic hardware in real experiments. Notably, this state should remain unchanged regardless of the number of diagnostic signals used, while more signals only refine the estimation, and fewer signals result in a coarser approximation. This is demonstrated in Fig.4a, when the diagnostic signals are randomly reduced by 5% and 20%, the extracted embeddings will only introduce limited oscillation around the black line.

Analogous to word embeddings, the plasma embeddings of physically similar discharges are expected to cluster in latent space. This structure is confirmed via Uniform Manifold Approximation and Projection (UMAP) projection of embeddings from the test set as shown in Fig.4b. Notably, despite uniform treatment of all 88 input channels during training, FusionMAE autonomously identifies and organizes embeddings along key plasma parameters (B_t, I_p, n_e and β_N), indicating a nontrivial, learned understanding of plasma physics.

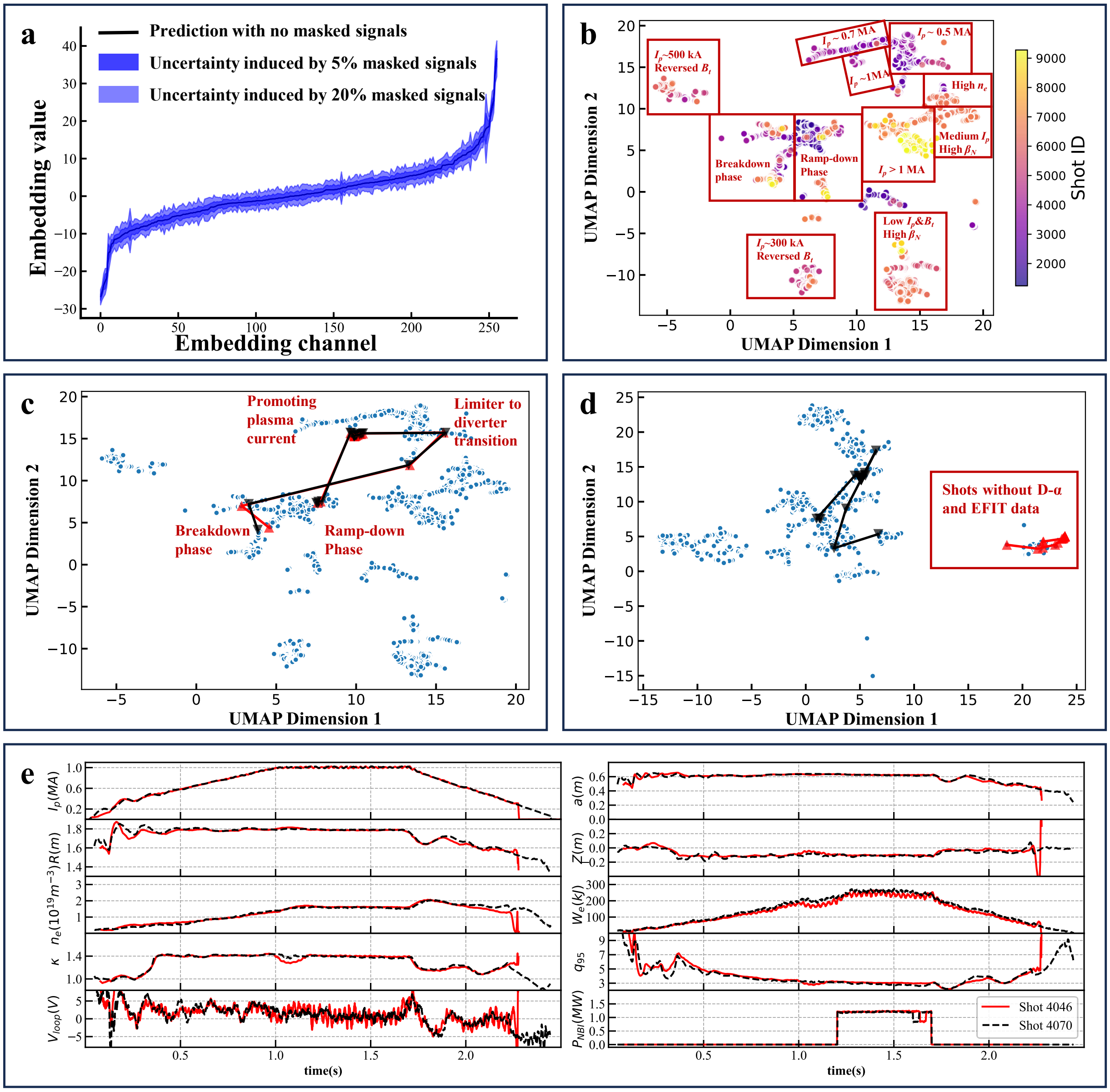


Fig.4: Demonstration of a plasma status embedding and its similar characteristic as word embedding.

a Plasma state embedding generated by FusionMAE during an HL-3 discharge. Blue bands indicate the uncertainty induced by randomly masking 5% and 20% of the input signals, showing that the embedding remains stable under partial input dropout. Embedding dimensions are sorted by value extracted from complete 88 signals for clarity.

b UMAP projection of averaged embeddings from all 100 ms trajectories in the test set, revealing clustering aligned with key physical parameters, B_t, I_p, n_e and β_N

c Trajectories of Shot 4046 and Shot 4070 in the UMAP plot generated from the embeddings from FusionMAE, which are very similar and show clear organized structures related to the operation stage.

d Trajectories of the two shots in the UMAP plot generated from raw data. Two shots are grouped into different clusters since Shot 4046 lacks valid EFIT and D-α data.

e Main plasma parameters during the two shots. Here the plasma shape parameters (R, Z, κ, ...) for Shot 4046 are restored by FusionMAE.

Compared to previous studies that employed dimensionality reduction methods like t-SNE or UMAP to analyze the scenario of discharges[37-40], FusionMAE can identify whether the underlying plasma state is consistent through different combinations of diagnostic signals. As a result, the dimensionality reduction analysis no longer depends on manual feature selection or data cleaning, leading to more robust and physically meaningful representations. This is further demonstrated in Fig.4c–4e. Two discharges, shot 4046 and shot 4070, are selected from the testing set. These two shots are operated under the same template setting, resulting in highly similar plasma trajectories, as shown in Fig.4e. However, for shot 4046, the EFIT code fails to converge during the automatic intra-shot reconstruction process and the data are therefore missing from the dataset. The D-α signal is also unavailable because the diagnostic system was not activated. Under such circumstances, when UMAP is directly applied to the raw diagnostic signals, the trajectories of the two shots are separated due to the discrepancy in the available signal sets, as illustrated in Fig.4d. In contrast, the UMAP projection based on the embeddings learned by FusionMAE (Fig.4c) correctly captures the similarity between the two discharges, despite the differences in diagnostic availability.

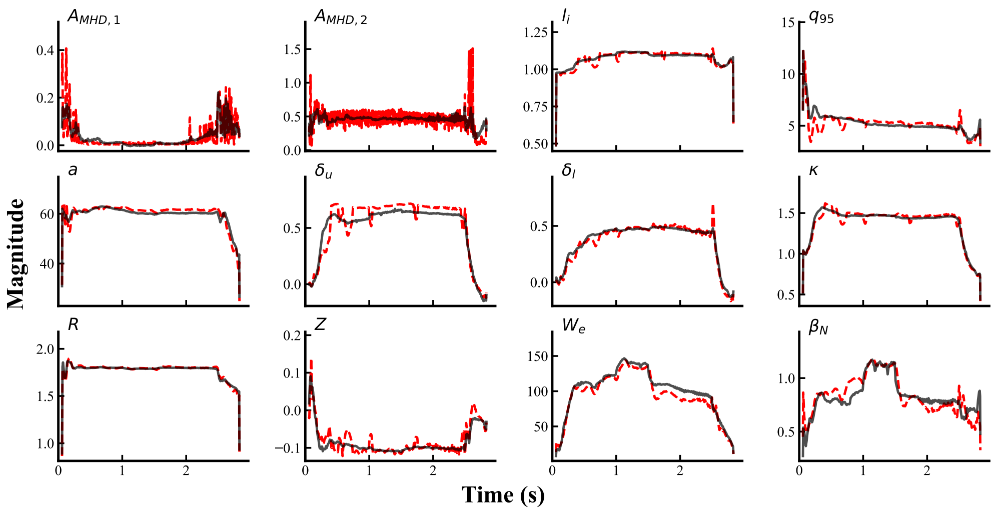


Fig.5: Demonstration of FusionMAE for the automatically ~~analyse~~ of secondary data. Black lines denote the reconstructed secondary signals when all of them are masked in the input. Red dashed lines represent the ground truth given by conventional analysis tools, like EFIT[41].

Downstream application: automatica~~lly~~ data analysis for secondary data

While prior analyses have primarily emphasized FusionMAE's pretraining performance, its true value as a large-scale pretrained model lies in its broad application on downstream tasks. Some of the applications are already empirically observed and systematically characterized in this and the following two subsections.

Firstly, the capability of signal reconstruction can be reframed as a form of automated data analysis. Among the 88 input channels, some are secondary data generated by data analysis programs. For example, the plasma shape parameters (R, Z, κ, ...) are calculated by the EFIT code based on the poloidal magnetic field and coil currents. An extreme test case was constructed where all secondary data were masked and the result is shown in Fig.5a. FusionMAE successfully reconstructed these masked secondary parameters with high fidelity, demonstrating its capacity to infer underlying physical relationships and approximate the functionality of traditional analysis tools. The statistical evaluation of FusionMAE's performance on this application is also presented in Fig.3a and 3b by the blue radar plots. More detailed statistical result can be found in the column of 'Reconstructed by FusionMAE' in Extended Table 5.

Downstream application: all-purpose vector for multiple control and simulation tasks

Since the plasma status embedding has integrated all the information from the 88 input signals, it is natural to be considered as an 'all-purpose vector' to support various tasks related to tokamak operation. To demonstrate this capability, three downstream tasks are selected. The first task, disruption prediction (DPR), involves identifying disruption precursors hidden in plasma parameters and exemplifies pattern recognition and anomaly detection [42-45]. The second task, EFIT-NN, is to calculate the plasma shape based on the poloidal magnetic field and coil currents, representing a data analysis challenge[46]. Instead of relying solely on FusionMAE's inherent data analysis capabilities as in Fig.5, the shape parameters in the 88*10 input array are masked, and an individual neural network is trained to compute them based on the plasma status embedding. The third task, plasma evolution prediction (PPR), is to forecast the future plasma parameters based on the previous parameters and planned coil currents, illustrating predictive simulation[47, 48]. All tasks adhere to the same problem definitions and baseline solutions as described in [27-29], except for certain modifications that conflict with FusionMAE. Performance on DPR is evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC)[49], while EFIT-NN and PPR are measured by the PCC between the predicted outputs and the ground truth[50]. Both AUC and PCC are bounded by 1, with values closer to 1 indicating better performance.

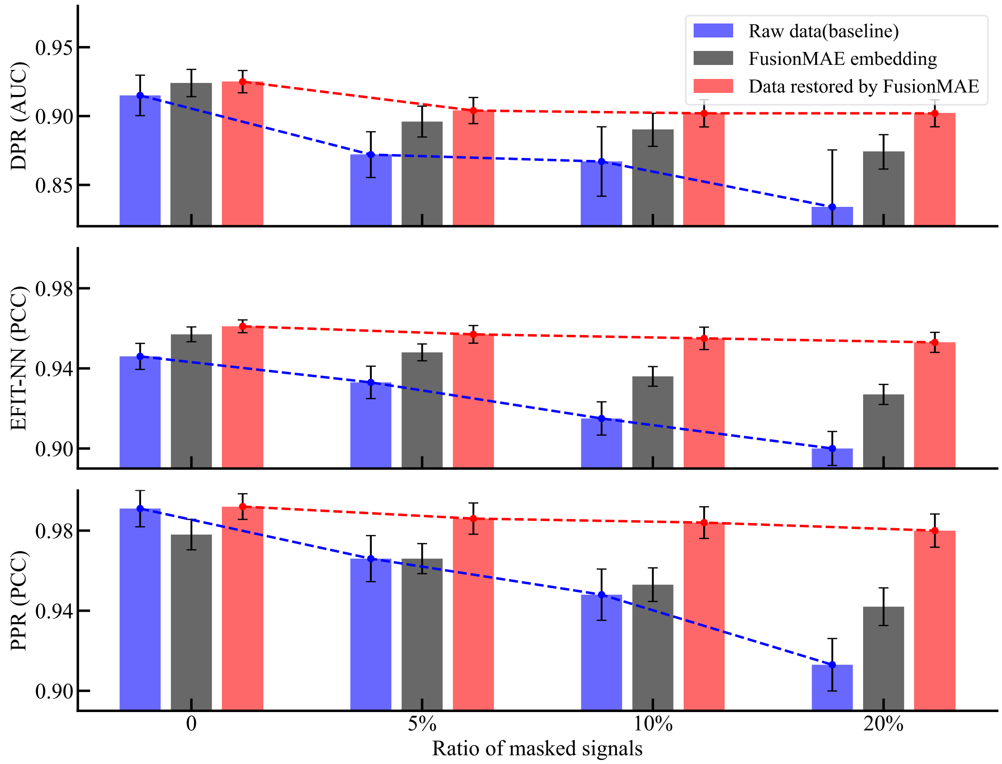


Fig.6: The performance of DPR, EFIT-NN and PPR based on models utilizing the plasma status embedding and data completed by FusionMAE, compared with ordinary models that are developed on the raw data. Error bars are obtained by repeating the training and testing for 5 times and calculating the standard deviation of performance.

The results of comparison experiments are shown in Fig.6 by the first 3 bars in each subfigure. In these experiments, the performances are tested for three algorithm versions using different inputs, i.e., raw data, plasma status embedding, and data with missing signals completed by FusionMAE. Impressively, both the versions using embedding and completed data achieve similar or even higher performance compared to the version using raw data. These results reveal that FusionMAE provide a precise, self-consistent representation of the plasma and benefits the downstream tasks. The superior performance of the third version over the second one likely stem from the information loss during the generation of the plasma status embedding. By combining the raw data with the reconstructed signals, the third version compensates for this loss and achieves the highest performance.

Downstream application: robust control against missing input signals

FusionMAE provides another advantage by enabling robust tokamak operation even when some diagnostic systems are unavailable[51]. On one hand, missing signals can be replaced by mask tokens and subsequently reconstructed by FusionMAE. On the other hand, the embedded vector, which integrates information from multiple diagnostic and control systems, remains stable when part of the input signals is unavailable.

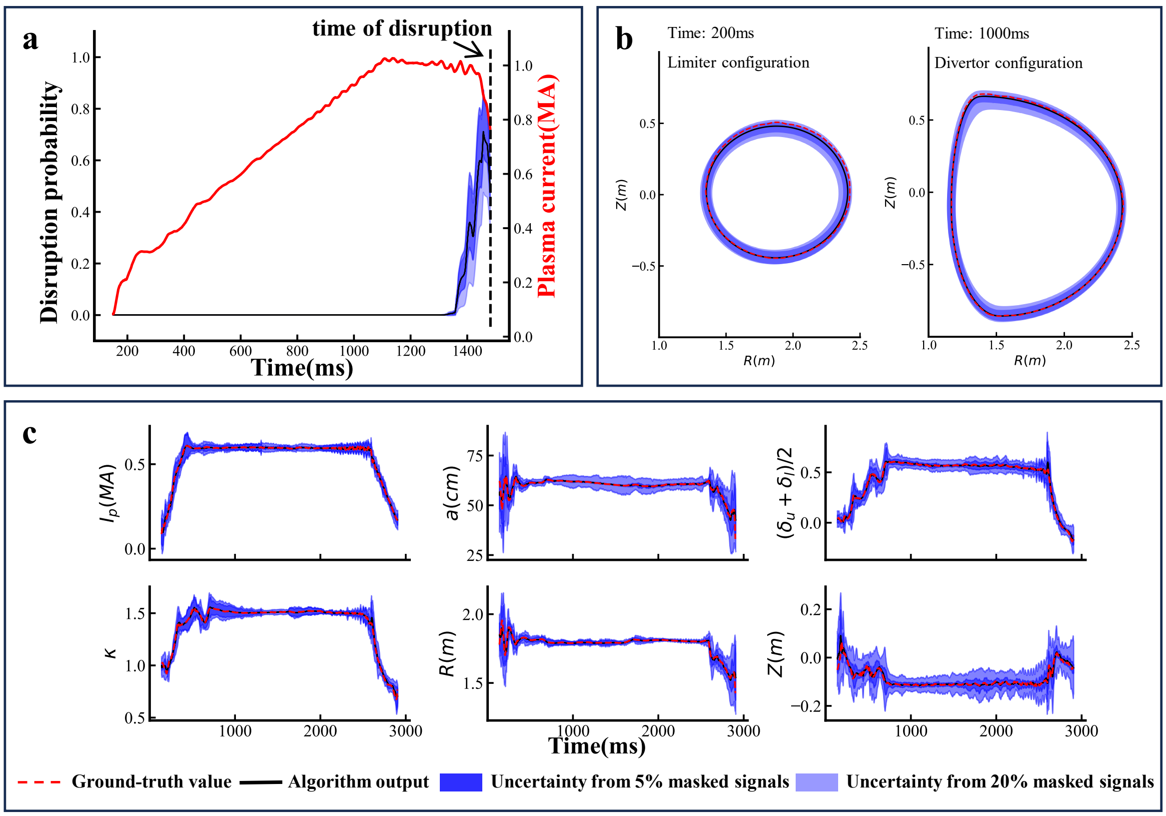


Fig.7: Output of DPR(a), EFIT-NN(b), PPR(c) model during an HL-3 discharge using the plasma state embedding. Blue bands indicate uncertainties introduced by randomly masking 5% and 20% of input signals, demonstrating robustness to missing and invalid diagnostics. The uncertainties are multiplied by 5 to get more visible.

Fig.7a, 7b, and 7c present the outputs of the DPR, EFIT-NN, and PPR tasks, respectively, during an HL-3 discharge, with the shaded area representing the uncertainty when 5% and 20% of the input is masked.

All outputs remain relatively stable, exhibiting only minor oscillations when some input signals are masked. Fig.6 further reports the statistical performance of the DPR, EFIT-NN, and PPR tasks with 5%, 10%, and 20% of the input masked. As expected, the performance of the baseline algorithm degrades rapidly as more signals are masked, whereas the degradation for algorithms supported by FusionMAE is considerably slower. Consequently, FusionMAE establishes a robust operational framework for tokamak reactors while demonstrating potential to minimize diagnostic system dependencies, which is a critical advancement toward FPP.

Discussion

This study presents FusionMAE, a large-scale pretrained model based on the masked autoencoder (MAE) architecture[52] that captures the intrinsic correlations among diverse plasma and operational parameters in tokamaks, thereby integrating multi-source data into a unified plasma state embedding. For fusion reactor operation, this integrated embedding shows significant potential for enhancing both system integration and reliability.

Regarding system integration, the plasma state vector generated by FusionMAE synthesizes information from multiple diagnostic systems, serving as a universal representation. This vector not only offers robust support for a variety of downstream computational tasks but also exhibits superior feature extraction capabilities that systematically enhance the performance of all downstream modules. Such a unified representation could revolutionize tokamak operational architecture by unifying the interface between diagnostic and control and benefitting all downstream control tasks.

In terms of operational reliability, FusionMAE tackles a critical challenge for future fusion reactors: mitigating operational instability resulting from partial diagnostic failures[53]. The model demonstrates computational resilience when handling incomplete input data, with downstream task outputs showing only bounded fluctuations and quantifiable uncertainty. Notably, FusionMAE achieves high-fidelity data imputation by leveraging its learned understanding of inter-parameter physical relationships, which could be used to relax the redundancy requirements for diagnostic systems for future fusion reactors.

From an artificial intelligence perspective, FusionMAE represents a successful application of self-supervised pre-training paradigm [54] to fusion data science. Its emergent properties, including sophisticated feature extraction, cross-task generalization, and intelligent behaviours in data compression and imputation, confirm the applicability of foundational architectures like BERT[55] and GPT[56] to fusion data domains. This finding suggests that fusion-AI integration should transcend conventional small-model approaches (e.g., supervised learning) and embrace cutting-edge techniques from the large-scale model era. Promising directions include applying human feedback reinforcement learning, mixture-of-experts architectures, and chain-of-thought reasoning to develop foundational models for fusion physics[57-60]. The demonstrated success of FusionMAE underscores the potential of adapting state-of-the-art AI methodologies to accelerate progress toward practical fusion energy solutions.

Methods

Dataset description

The HL-3 Tokamak, formerly known as HL-2M, is an experimental fusion device developed by the Southwestern Institute of Physics in China. It is designed to enhance the understanding of key physics and engineering principles essential for the advancement of international fusion initiatives such as ITER. HL-3 is capable of sustaining a plasma current between 1.5 and 3 MA while operating under a magnetic field strength of 2.3 T. Its structural characteristics include a major radius of 1.78 m, a minor radius of 0.65 m, and an aspect ratio of approximately 2.8, reflecting its toroidal configuration.

The dataset utilized in this study encompasses 2,445 plasma discharges collected from four HL-3 experimental campaigns, spanning the years 2022 to 2025. The discharges are randomly divided into a development set and a test set, containing 1,950 and 495 shots, respectively. The development set is employed for training and validating the neural network, while the test set is used to assess algorithm performance. This dataset splitting strategy consistently applied across both the pretraining and evaluation of FusionMAE, as well as the training and testing of all downstream task models. As illustrated in Fig.2a, each discharge includes 88 recorded signals that characterize the plasma state, which can be measured in real-time in HL-3 plasma control system. To ensure alignment and streamline the FusionMAE framework, all signals are resampled at 1 kHz. The physical significance, statistical properties, and measurement methods of these signals are detailed in Extended Table 1.

For most channels, the health rate, defined as the proportion of shots in which a given signal channel has recorded valid data, is below 100%. This observation underscores the inherent challenges and complexity of fusion diagnostics, emphasizing that missing diagnostic signals are an unavoidable reality in fusion facilities. To systematically evaluate the performance of various algorithms under different levels of missing data, we constructed three extended datasets based on the original dataset. Specifically, 5%, 10%, and 20% of the signal channels were randomly replaced with Gaussian noise curves, simulating artificial data loss. These extended datasets were subsequently used for comparative experiments, as presented in Fig.4a, 6 and 7. The results demonstrate that FusionMAE effectively learns from incomplete datasets, overcoming a key limitation of conventional algorithms, which typically require fully clean and complete data for training and inference.

Finally, Z-score normalization is adopted for data preprocessing in FusionMAE, where each input channel is normalized using the mean and standard deviation estimated from the training set.

Extended Table 1: The detailed information about the 88 channels used in the development of FusionMAE, 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.	0.384	0.228	100%
R	1	m	Horizontal position of the plasma center in the poloidal cross-section, calculated using the EFIT code.	1.76	0.110	90.9%
Z	1	m	Vertical position of the plasma center in the poloidal cross-section, calculated using the EFIT code.	-0.0608	0.113	90.9%
a	1	cm	Minor radius of the plasma in the poloidal cross-section, calculated using the EFIT code.	58.7	7.68	90.9%
κ	1	-	Elongation ratio of the plasma in the vertical direction, calculated using the EFIT code.	1.22	0.268	90.9%
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.	0.195	0.225	90.9%
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.	0.361	0.405	90.9%
B_t	1	T	Toroidal magnetic field strength at the geometric center of the vacuum vessel.	1.31	0.296	99.7%
B_p	41	a.u.	Local poloidal magnetic field strength, measured by 41 pick-up coils surrounding the plasma.	0.0701	0.0777	99.8%
V_{loop}	1	V	Toroidal loop voltage of the plasma.	2.35	3.14	100%
n_e	2	$10^{19}/m^3$	Line-averaged electron density, measured using Faraday interferometers.	0.943	0.985	64.4%
T_e	2	keV	Electron temperature, measured using electron cyclotron emission (ECE) diagnostics.	0.428	0.454	57.3%
W_e	1	kJ	Plasma stored energy.	67.7	67.8	90.9%
P_{rad}	1	MW	Total thermal radiation power emitted by the plasma.	0.122	0.194	63.2%

SX	4	a.u.	Line-integrated soft X-ray intensity from the plasma.	0.304	0.360	78.7%
D_α	2	a.u.	Intensity of the D-α spectrum.	0.225	0.525	74.3%
A_{MHD}	2	a.u.	Amplitude of magnetohydrodynamic (MHD) instabilities with a toroidal mode number n=1, derived from raw signals recorded by Mirnov probes.	0.173	0.292	93.1%
A_{imp}	1	a.u.	Intensity of the C III impurity spectrum.	0.141	0.270	64.7%
β_N	1	-	Normalized beta, representing the normalized plasma pressure.	0.655	0.525	90.9%
q_{95}	1	-	Safety factor at the 95% flux surface.	5.31	2.74	90.9%
l_i	1	-	Internal inductance of the plasma.	1.03	0.068	90.9%
I_{PF}	16	kA	Current in the poloidal field coils of the tokamak.	-0.618	1.70	99.9%
I_{CS}	1	kA	Current in the central solenoid coil of the tokamak.	6.35	36.7	99.9%
P_{NBI}	1	kW	Power of neutral beam injection.	55.1	220	99.9%
P_{EC}	1	kW	Power of electron cyclotron resonant heating.	19.2	109	98.7%
P_{LH}	1	kW	Power of lower hybrid resonant heating.	5.10	30.0	98.6%

Model architecture of FusionMAE

FusionMAE is a Transformer-based masked autoencoder, with its architecture illustrated in Fig.2b. The model design is inspired by the masked autoencoder for vision Transformers proposed in [52]. It processes an input consisting of an 88×10 array, where each row, a 10 ms time window of signals sampled at 1 kHz, corresponds to one of the 88 diagnostic channels. Within the Transformer framework, each row is treated as a patch, encapsulating plasma information from a single diagnostic channel.

To handle missing and invalid channels, a mask token is introduced in the neural network, replacing unavailable data. This mask token is a randomly initialized trainable vector, which is updated throughout the training process and gradually converges into an optimal representation that encodes missing information. In addition to these inherently missing or invalid channels, 25% of the valid channels are intentionally masked during training. These masked channels serve as the primary targets for reconstruction, enabling FusionMAE to learn to reconstruct missing information effectively, as detailed in the next section.

Afterwards, each row of input is projected into a 128-dimensional space using a multilayer perceptron (MLP), followed by the addition of a positional encoding ($\boldsymbol{PE}$) vector [33] to encode the channel index and inform the encoder about the diagnostic source of each patch. The positional encoding is computed using Equations (1) and (2), where *idx* represents the index of the input channel, and $\boldsymbol{2i}$ and $\boldsymbol{2i+1}$ refer to the indices of the even and odd elements of the positional encoding vector, respectively.

$$PE(idx, 2i) = \sin(idx/88^{2i/128}) \qquad i = 0,1,2,\ldots,63 \tag{1}$$

$$PE(idx, 2i+1) = \sin(idx/88^{(2i+1)/128}) \qquad i = 0,1,2,\ldots,63 \tag{2}$$

The resulting 88×128 array is then passed through an encoder composed of multiple Transformer blocks, each incorporating multi-head self-attention layers [34], MLPs, layer normalizations [35], and residual connections [36]. This Transformer block architecture is widely adopted in large-scale models such as BERT and GPT, demonstrating its effectiveness in learning intricate dependencies within sequential data.

The encoder primarily captures underlying patterns in plasma behaviour. Its output maintains the same shape as the input (88×128) and is subsequently flattened into a 11264-dimensional vector. This vector is further compressed by an MLP into a 256-dimensional embedded representation.

The decoder adopts a structure that closely mirrors the encoder. The 256-dimensional embedded vector is first expanded back into a 11264-dimensional vector, which is then reshaped into an 88×128 array. Subsequently, multiple Transformer blocks process this array to reconstruct the original data.

The key distinction between the encoder and decoder lies in the treatment of positional encoding. Unlike in the encoder, where positional embeddings are directly added to the input patches, in decoder, they serve as query vectors in the final multi-head attention layer, aiding in the reconstruction process.

Finally, an MLP layer projects the 88×128 array into an 88×10 array, representing the reconstructed diagnostic signals across all channels, including those that were originally missing or invalid.

Extended Table 2: The detailed information about model structure of FusionMAE

Number	Repeat Time	Layer type	Output shape	Layer parameters
1	1	Input layer	88*10	
2	1	Multi-layer perceptron	88*128	128 units, dropout rate = 0.2
3	20	LayerNorm	88*128	
4		Multi-head attention	88*128	8 heads, key dimention = 128, dropout rate = 0.2
5		LayerNorm	88*128	
6		Multi-layer perceptron	88*256	256 units, dropout rate = 0.2
7		Multi-layer perceptron	88*128	128 units, dropout rate = 0.2
8	1	Reshape	11264	
9	1	Multi-layer perceptron	256	256 units, dropout rate = 0.2
10	1	Multi-layer perceptron	11264	11264 units, dropout rate = 0.2
11	1	Reshape	88*128	
12	4	LayerNorm	88*128	
13		Multi-head attention	88*128	8 heads, key dimention = 128, dropout rate = 0.2
14		LayerNorm	88*128	
15		Multi-layer perceptron	88*256	256 units, dropout rate = 0.2
16		Multi-layer perceptron	88*128	128 units, dropout rate = 0.2
17	1	Multi-layer perceptron	88*10	10 units, dropout rate = 0.2

Training of FusionMAE

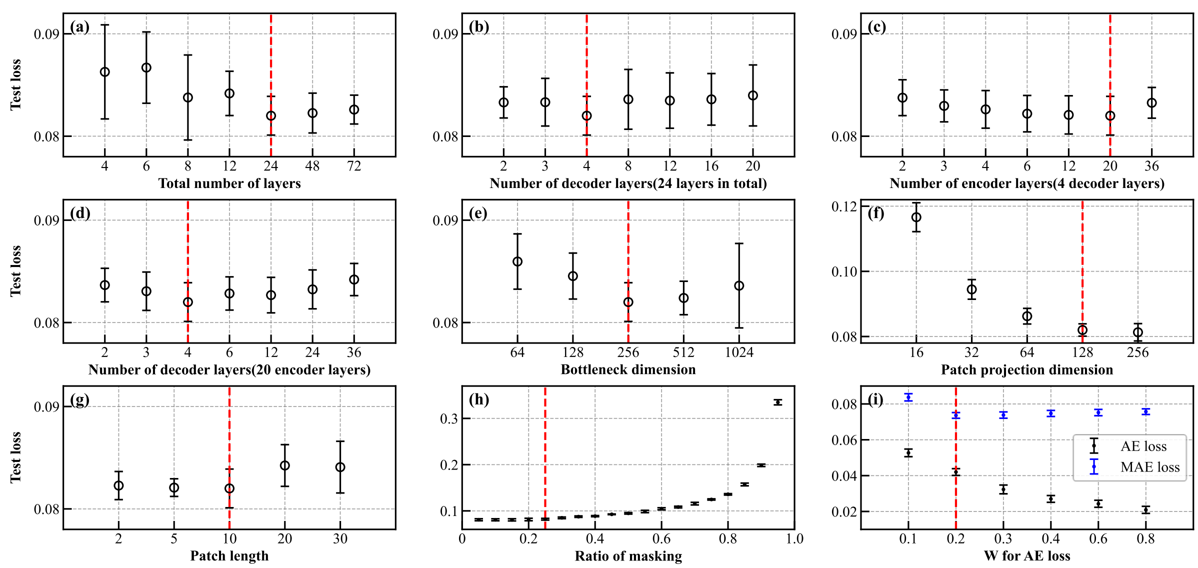
The FusionMAE model is trained using the AdamW optimizer, aiming to minimize the loss function defined in Equation (3). In this formulation, $X_{recontruct}$ represents the output of FusionMAE. $X_{original}$ denotes the original input data. Term W assigns weights to each input channel, with values set as 0.2, 0, and 1 for valid channels, invalid/missing channels, and intentionally masked channels, respectively.

$$loss = \sum_{1}^{88} W * \left\| X_{recontruct} - X_{original} \right\| \qquad (3)$$

This weighting strategy ensures that FusionMAE primarily learns to reconstruct the intentionally masked channels, while simultaneously preserving valid information during the compression process. Given the model's architecture, its capability to reconstruct intentionally masked channels naturally extends to the reconstruction of invalid and missing channels, as these are treated equivalently during training.

A systematic set of ablation experiments is conducted to determine the optimal hyperparameters of FusionMAE, with the results summarized in Fig.8.

Model depth is firstly investigated by varying the total number of Transformer blocks. As shown in Fig.8a, FusionMAE exhibits progressively improved and more stable performance as the network depth increases, with the performance gain saturating beyond 24 stacked blocks. Based on this observation, the total depth is fixed at 24 layers. We further examine the placement of the bottleneck layer in Fig.8b and observe a modest but consistent improvement when the bottleneck is shifted toward the output end, corresponding to a deeper encoder and a shallower decoder. Additional scans, including varying the encoder depth with a fixed 4-layer decoder (Fig.8c) and varying the decoder depth with a fixed 20-layer encoder (Fig.8d), yield consistent trends. These results suggest that the encoder bears a higher modeling burden, as it must accommodate diverse patterns of missing data, whereas the decoder mainly performs a relatively fixed reconstruction process based on the learned plasma embedding, which has largely disentangled the influence of missing signals.



Extended Fig.8: Result of comparison experiments used to specify the hyperparameter settings of FusionMAE

After fixing the model depth, we explore the model width. The first width-related parameter is the dimensionality of the bottleneck layer (Fig.8e), which also defines the dimension of the plasma state embedding. This dimension must be smaller than the total input dimensionality (880 channels) to prevent tricky solutions such as direct memorization of the inputs, while still retaining sufficient expressive power to represent the wide range of plasma states. Scanning this parameter from 64 to 1024 reveals that a dimension of 256 achieves the best performance on the HL-3 dataset. Another width parameter is the patch projection dimension (Fig.8f), which controls the internal width of the Transformer blocks. Increasing this dimension leads to improved and more stable performance; however, it also results in a linear increase in model size. A value of 128 is therefore selected as a balanced choice, providing strong performance with an acceptable computational cost.

In its current form, FusionMAE does not incorporate a dedicated architectural treatment for temporal dynamics. As a result, model performance degrades noticeably when the patch length becomes too long. The experiments in Fig.8g indicate that this transition occurs at approximately 10 ms. Addressing this limitation would likely require, on the one hand, augmenting the base Transformer with mechanisms tailored for time-series modeling, and on the other hand, potentially reformulating the pretraining framework in an autoregressive manner.

For the ratio of intentionally masked patches in Fig. 8h, the results provide particularly informative insights. When the masking ratio remains below a critical threshold (approximately 30%), the testing loss remains nearly constant, indicating that FusionMAE effectively exploits cross-channel correlations and reflecting a substantial level of redundancy in the diagnostic measurements. Beyond this threshold, the loss increases in a nonlinear manner, revealing a fundamental information limit where plasma state reconstruction becomes underdetermined. This behavior establishes a quantitative framework for assessing diagnostic redundancy and provides empirical guidance for identifying minimum viable diagnostic configurations, which is especially relevant for balancing measurement robustness and system complexity in future fusion reactors.

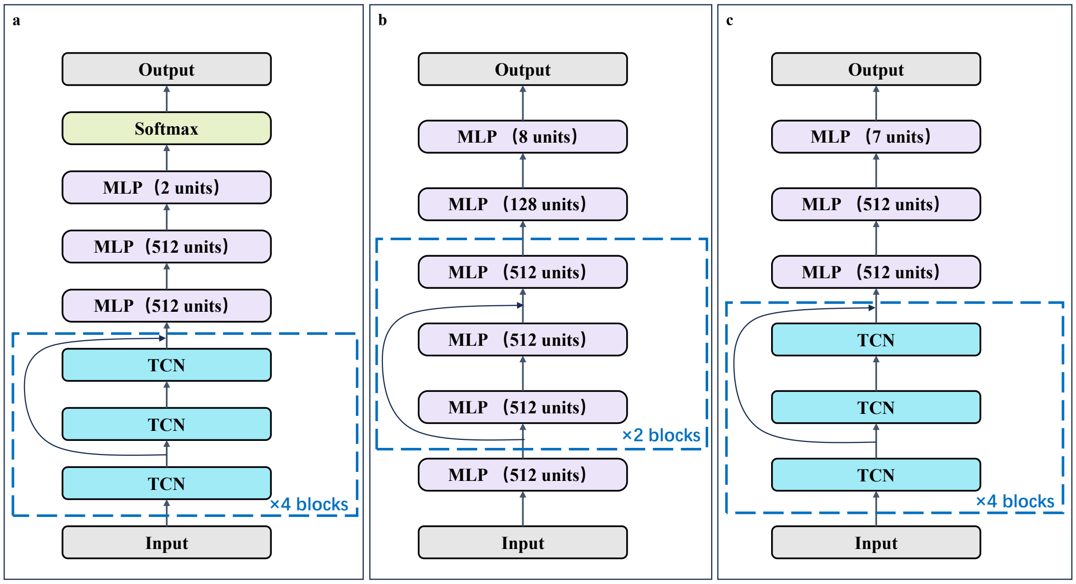
The final hyperparameter to determine is W in Equation (3). As shown in Fig.8i, increasing W steadily reduces the loss for valid data (AE loss). However, beyond $W = 0.2$, the loss for masked data (MAE loss) begins to increase. Since our study prioritizes FusionMAE's robustness to missing diagnostics and its ability to capture inter-diagnostic correlations, we selected $W = 0.2$, corresponding to the minimum MAE loss.

Other training-related hyperparameters are summarized in Table 3. The final FusionMAE model contains approximately 14 million trainable parameters. Training requires about 10 hours on a single NVIDIA A100 GPU, while inference for a single time slice takes 4.8 ms, enabling quasi-realtime application when combined with a 10 ms input time window.

Overall, the training of FusionMAE demonstrates high repeatability and low sensitivity to hyperparameter choices. This behavior stands in marked contrast to previously developed task-specific small models for HL-3 and highlights the robustness of self-supervised pretraining combined with Transformer-based architecture. These results suggest that such a framework may serve as a key foundation for next-generation AI-for-fusion solutions.

Extended Table 3: 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



Extended Fig.9: The neural network structure used in three downstream tasks: (a) disruption prediction, (b) equilibrium fitting surrogate model, (c) plasma parameter prediction. TCN refers to temporal convolutional neural network, whose detailed hyper-parameters keeps the same as in [27, 28]

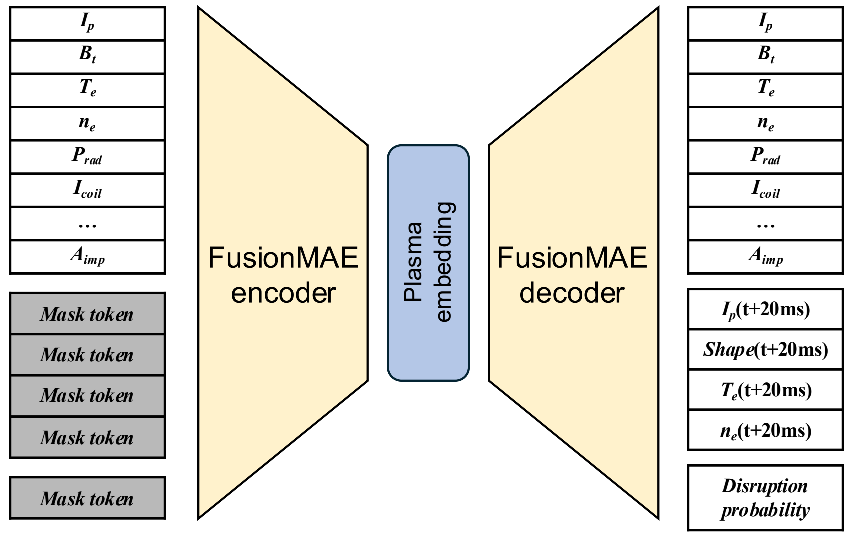
Downstream task: disruption prediction

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 magnetohydrodynamic (MHD) instabilities, thereby enabling the activation of mitigation strategies to minimize potential damage [61]. 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 their 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 [20, 42–45, 62–64]. 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 [65, 66]; (ii) anomaly-detection-based approaches that reduce dependence on disruptive samples and achieve competitive performance using predominantly non-disruptive data [67–69]; (iii) interpretability analyses that provide physical insights to guide disruption avoidance and mechanism studies [70–72]; and (iv) enhancement of cross-parameter and cross-device generalization to meet the requirements of future fusion reactors[73,74].

The neural network architecture of the disruption prediction algorithms proposed in this study is presented in Fig.9a. The model architecture, data preprocessing, loss function and performance evaluation keep the same paradigm as our prior study conducted on HL-3[27], 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 Table 4, 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 Table 4 and Fig.6, 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.



Extended Fig.10: 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 Fig.10, 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[27], 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.

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.

Extended Table 4: 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

Downstream task: 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 [75, 76]. Recent developments have further improved performance by incorporating auxiliary constraints such as physics-informed loss functions and multi-task learning strategies [28, 77, 78]. 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.

The baseline architecture, illustrated in Fig.9b, keeps the EFIT-NN framework described in [28]. 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 Table 5 and Fig.6, clearly demonstrating the performance improvements enabled by the integration of FusionMAE.

Extended Table 5: 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	-

Downstream task: 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 Equation (4).

$$s(t+1) = F(s(t), a(t)) \qquad (4)$$

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 [79–82]. 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 [27], as well as two autoregressive models based on coil currents and power supply voltages, respectively [29,83].

At the current stage, FusionMAE is not well aligned with autoregressive prediction frameworks in [29,83], 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 [27] as our baseline. The corresponding architecture is shown in Fig.9c, and the model design, data preprocessing, loss function, and evaluation protocol follow those described in [27]. Model performance is assessed using the PCC and RMSE, which quantify the agreement between predicted and experimentally measured plasma shape parameters.

Table 6 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 [83] 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. 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.

Extended Table 6: Performances of different versions of evolution prediction models.

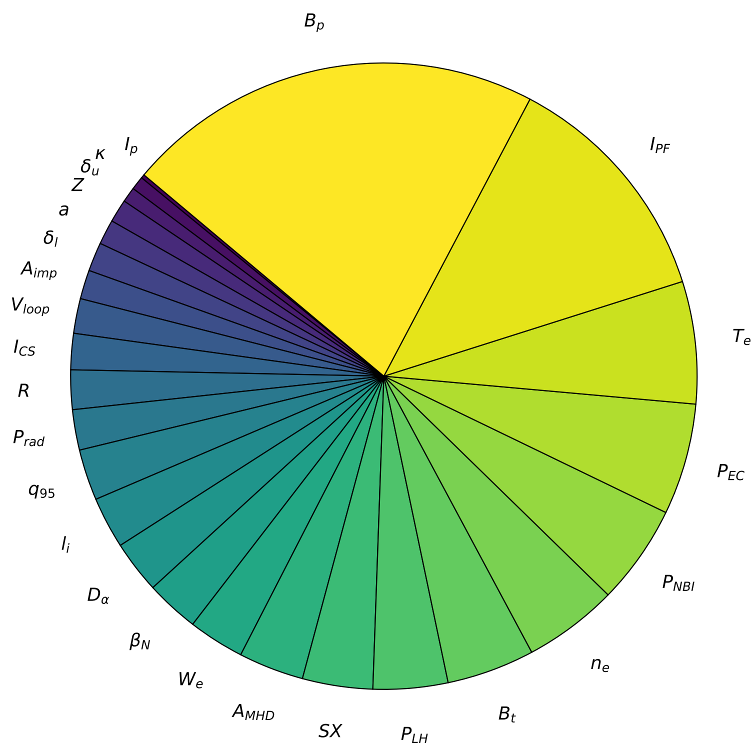
Parameter	TCN + Data restored by FusionMAE		TCN + FusionMAE embedding		TCN (baseline)		WaveNet	
	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE
I_p(kA)	0.999	2.63	0.998	5.03	0.999	2.53	0.999	2.26
R(m)	0.989	0.00386	0.961	0.00807	0.988	0.00387	0.984	0.00401
Z(m)	0.981	0.00482	0.968	0.00851	0.979	0.00494	0.972	0.00610
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	-

Contribution of each diagnostic for plasma status embedding

FusionMAE introduces a novel paradigm for describing plasma status. Rather than relying on an increasing number of diagnostic measurements to provide additional parameters, FusionMAE represents plasma status 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 status can be quantified using Equation (5):

$$contribution(i) = \frac{1 - PCC^*(i)}{\sum_{j=1}^{88}\left(1 - PCC^*(j)\right)} \qquad (5)$$

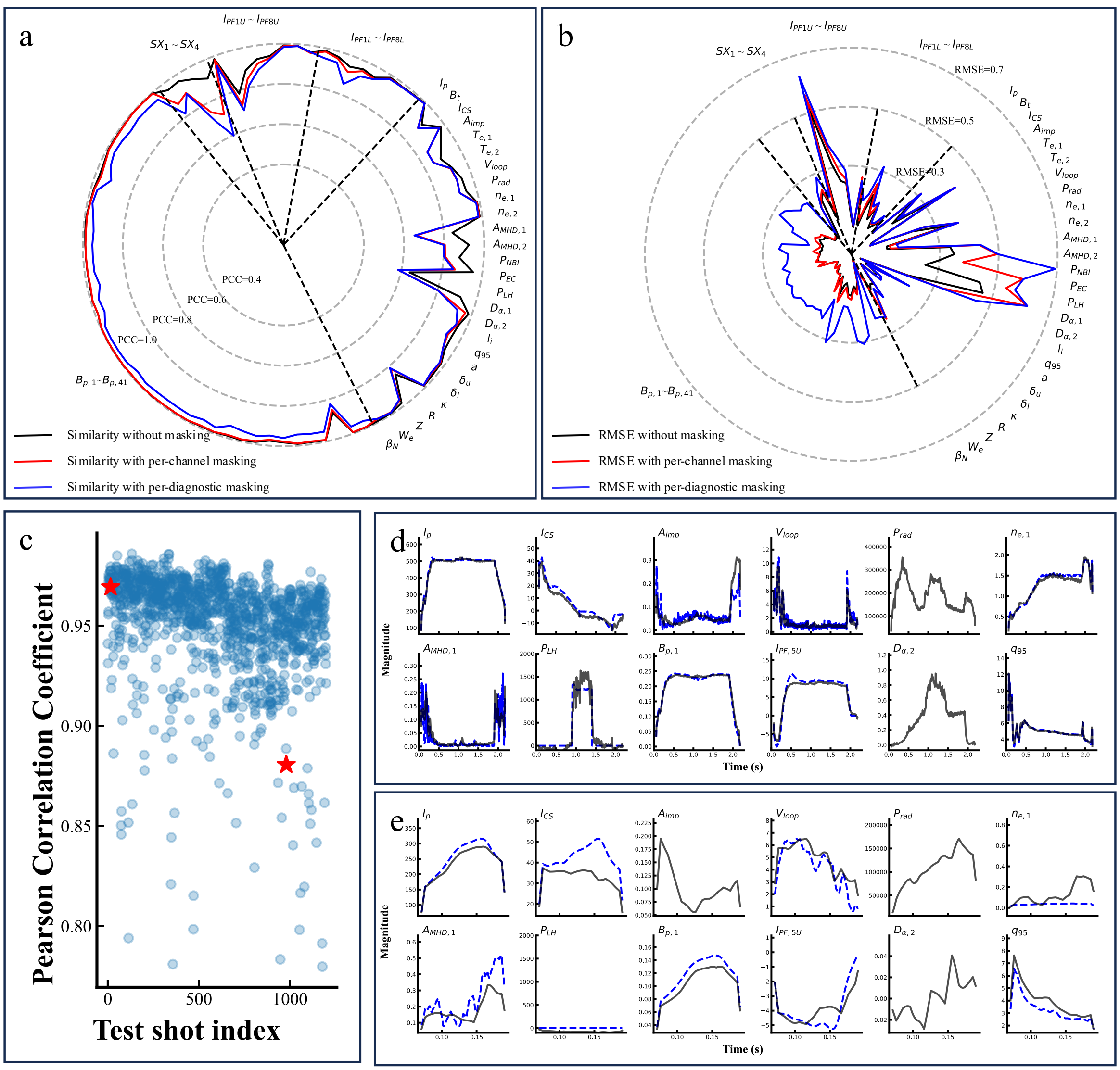
where $PCC^*(i)$ denotes the PCC between the embedding vector computed from the complete set of input channels and the embedding vector obtained when masking the i_{th} channel. Fig.11 illustrates the contributions of the 88 diagnostic signals in determining plasma status, providing critical insights 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.



Extended Fig.11: The contribution of different channels in the determination of plasma status embedding.

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 Fig.12.



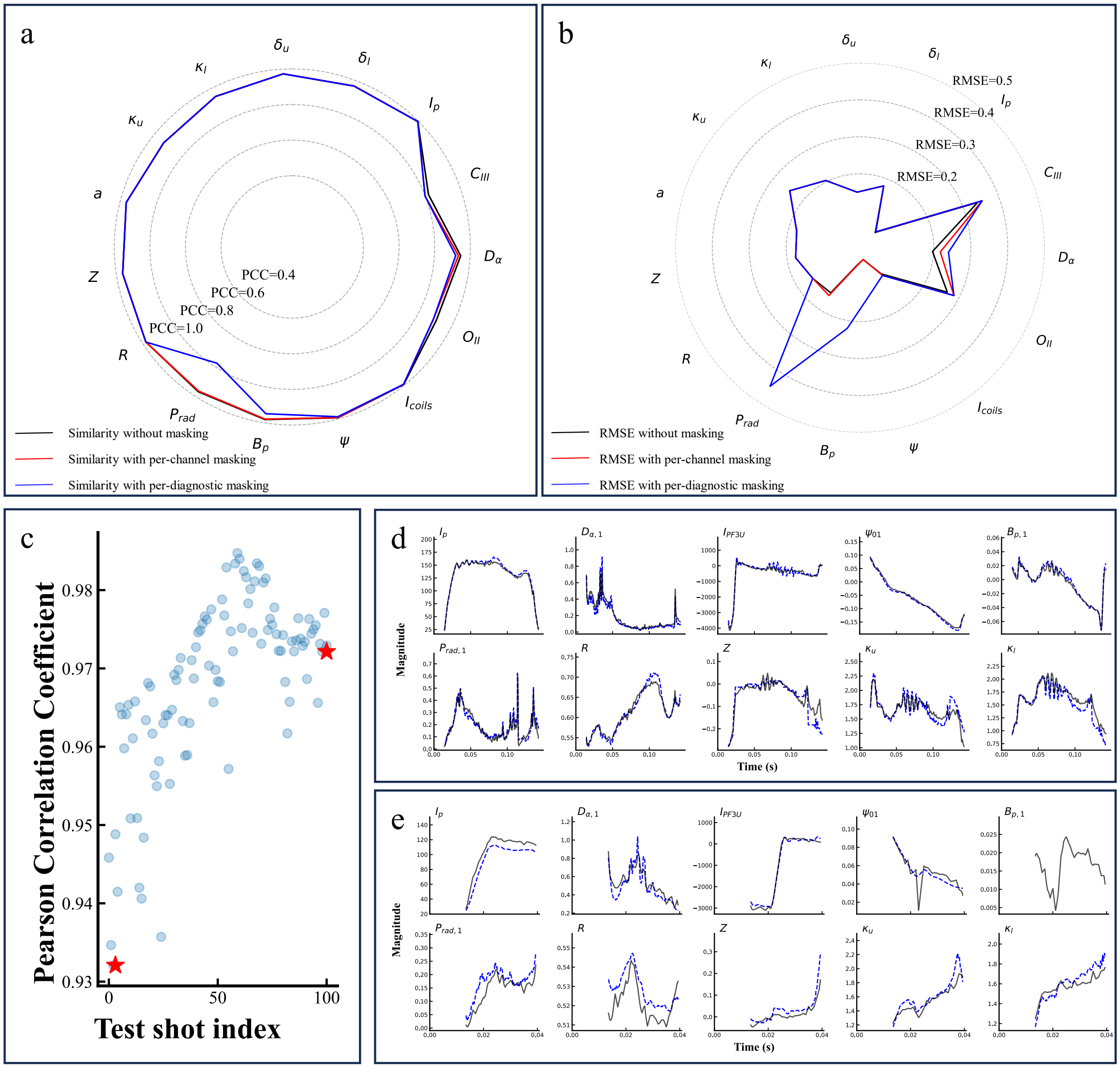
Extended Fig.12: Evaluation of FusionMAE's performance in latest experimental campaign of HL-3. (d) Shot 9352. (e) Shot 12891

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 [84].

In SUNIST-2, the input of FusionMAE consists of plasma current (I_p), plasma shape and position (R, Z, a, κ_u, κ_d, δ_u, δ_l), coil current(I_{coil}, 13 channels), toroidal field (B_t), poloidal field (B_p, 87 channels), magnetic flux (ψ, 29 channels), radiation power (P_{rad}, 16 channels), impurity radiation spectrum (C_{III}, D_α, O_{II}, 2 channels for D_α). 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 Fig.13. 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.



Extended Fig.13: 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. (d) Shot 250419001. (e) Shot 250731022

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code Availability

The code that supports the findings of this study are available from the corresponding author upon reasonable request.

References

1. C.F. Maggi, D. Abate, N. Abid, et al. Overview of T and D–T results in JET with ITER-like wall. *Nuclear Fusion* **64** 112012 (2024).

2. Han, H. et al. A sustained high-temperature fusion plasma regime facilitated by fast ions. *Nature* **609**, 269–275 (2022).

3. Song, Y. et al. Realization of thousand-second improved confinement plasma with Super I-mode in tokamak EAST. *Sci. Adv.* **9**, eabq5273 (2023).

4. B. Bigot, Preparation for assembly and commissioning of ITER. *Nucl. Fusion* **62**, 042001 (2022).

5. Ambrogio Fasoli. Essay: Overcoming the Obstacles to a Magnetic Fusion Power Plant. *Physical Review Letters*, **130**, 220001 (2023)

6. Shimada, M. et al. Progress in the ITER physics basis—chapter 1: overview and summary. *Nuclear Fusion* **47**, S1 (2007).

7. Alan E. Costley, et al. Towards Diagnostics for a Fusion Reactor. *IEEE Transactions on Plasma Science* **38**:10, 2934 – 2943, (2010).

8. A. Fasoli, S. Brunner, et al. Computational challenges in magnetic-confinement fusion physics. *Nature Physics* **12**, 411–423 (2016)

9. K. Kurihara, J.B. Lister, D.A. Humphreys, et al. Plasma control systems relevant to ITER and fusion power plants. *Fusion Engineering and Design*. **83**: 7–9, 959-970, (2008)

10. A.J.H. Donné, A.E. Costley, R. Barnsley, et al. Progress in the ITER physics basis—chapter 7: diagnostics. *Nuclear Fusion* **47**, S337 (2007).

11. F. Imbeaux, S.D. Pinches, J.B. Lister, et al. Design and first applications of the ITER integrated modelling & analysis suite. *Nuclear Fusion* **55**, 123006 (2015).

12. Y. Gribov, D. Humphreys, K. Kajiwara, et al. Progress in the ITER physics basis—chapter 8: Plasma operation and control. *Nuclear Fusion* **47**, S385 (2007).

13. Bernard Bigot. ITER construction and manufacturing progress toward first plasma. *Fusion Engineering and Design*. **146**: A, 124-129, (2019)

14. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **60**: 6 84-90, (2017)

15. Luciano Floridi, Massimo Chiriatti. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, **30**, 681–694, (2020)

16. John Jumper, Richard Evans, Alexander Pritzel, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021)

17. Jiali Li, · Kaizhuo Lim, · Haitao Yang, et al. AI Applications through the Whole Life Cycle of Material Discovery. *Matter*, **3**: 2, 393-432, (2020).

18. Kaifeng Bi, Lingxi Xie, Hengheng Zhang, et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533-538, (2023)

19. Julian Kates-Harbeck, Alexey Svyatkovskiy, William Tang. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526-531 (2019).

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

21. Jonas Degrave, Federico Felici, Jonas Buchli, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414-419 (2022).

22. Jaemin Seo, SangKyeun Kim, Azarakhsh Jalalvand, et al. Avoiding fusion plasma tearing instability with deep reinforcement learning. *Nature* **626**, 746-751 (2024).

23. X.R. Duan, M. Xu, W.L. Zhong, et al. Progress of HL-2A experiments and HL-2M program. *Nuclear Fusion*, **62**, 042020 (2022).

24. X.R. Duan, M. Xu, W.L. Zhong, et al. Recent advance progress of HL-3 experiments. *Nuclear Fusion*, **64**, 112021 (2024).

25. Kai Han, Yunhe Wang, Hanting Chen, et al. A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **45**:1, 87-110 (2023)

26. Yang Li and Tao Yang. Word embedding for understanding natural language: a survey. *Guide to big data applications*, 83-104. (2018)

27. 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).

28. 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).

29. 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)

30. Mei Huang, Jao Rao, Shaodong Song, et al. Design and Research of Electron Cyclotron Resonance Heating and Current Dive System on HL-2M Tokamak. *EPJ Web of Conferences* **147**, 04006 (2017)

31. Rui Ma, Jiaxian Li, Shuo Wang, et al. Eddy current effect study on the HL-3 plasma equilibrium reconstruction. *Fusion Engineering and Design*. **212**, 114816, (2025)

32. Wang A., Sun T. F., Chen W., et al. Final development and preliminary experiment progress of in-vessel resonant magnetic perturbation coils system on HL-3 tokamak. *Fusion Engineering and Design*, **208**, 114702 (2024).

33. Pu-Chin Chen, Henry Tsai, Srinadh Bhojanapalli, et al. A Simple and Effective Positional Encoding for Transformers. Preprint at http://arXiv.org/abs/2104.08698 (2021)

34. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in neural information processing systems*, **30**, (2017).

35. Sergey Ioffe, Christian Szegedy. *Proceedings of the 32nd International Conference on Machine Learning,* **37**, 448-456, (2015).

36. He K., Zhang X., Ren S., et al. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*.770-778, (2016).

37. A. Pau, A. Fanni, S. Carcangiu, et al. A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET. Nuclear Fusion 59, 106017, (2019).

38. E. Aymerich, A. Fanni, G. Sias, et al. A statistical approach for the automatic identification of the start of the chain of events leading to the disruptions at JET. Nuclear Fusion 61, 036013, (2021).

39. J. Zhu, C. Rea, R.S. Granetz, et al. Scenario adaptive disruption prediction study for next generation burning-plasma tokamaks. Nuclear Fusion 61, 114005, (2021).

40. Y. Wei, J.P. Levesque, C.J. Hansen, et al. A dimensionality reduction algorithm for mapping tokamak operational regimes using a variational autoencoder (VAE) neural network. Nuclear Fusion 61, 126063, (2021).

41. Sun X., Akcay C., Amara T. B., et al. Impact of various DIII-D diagnostics on the accuracy of neural network surrogates for kinetic EFIT reconstructions. *Nuclear Fusion*, **64**, 086065 (2024).

42. 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).

43. 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).

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

45. 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)

46. J.P. Qian, L.L. Lao, H.Q. Liu, et al. EAST equilibrium current profile reconstruction using polarimeter-interferometer internal measurement constraints. *Nuclear Fusion* **57**, 036008 (2016)

47. Chenguang Wan, Zhi Yu, Alessandro Pau, et al. A machine-learning-based tool for last closed-flux surface reconstruction on tokamaks. *Nuclear Fusion* **63**, 056019 (2023)

48. S H Kim, J F Artaud, V Basiuk, et al. Full tokamak discharge simulation of ITER by combining DINA-CH and CRONOS. *Plasma Physics and Controlled Fusion*, **51** 105007, (2009).

49. Viv Bewick, Liz Cheek, Jonathan Ball. Statistics review 13: Receiver operating characteristic curves. *Critical Care*, **8**, 508, (2004).

50. Cohen I., Huang Y., Chen J., et al. Pearson correlation coefficient. *Noise reduction in speech processing*, 1-4 (2009).

51. Walsh M, Andrew P, Barnsley R, et al. Integration of diagnostics on ITER. *IEEE 26th Symposium on Fusion Engineering* 1-8 (2015).

52. He K., Chen X., Xie S., et al. Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16000-16009, (2022).

53. Osamu Mitarai, Katsunori Muraoka. Analyses of Diagnostic Failure Effect and Fail-Safe Ignited Operation in a Tokamak Fusion Reactor, Fusion Technology, **36**(2): 194-211, (1999).

54. Goyal S, Kumar A, Garg S, et al. Finetune like you pretrain: Improved finetuning of zero-shot vision models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19338-19347, (2023).

55. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Preprint at http://arXiv.org/abs/1810.04805 (2019)

56. A Radford, K Narasimhan, T Salimans, I Sutskever. Improving language understanding by generative pre-training. Preprint at http://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (2018)

57. Yuntao Bai, Andy Jones, Kamal Ndousse, et al. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. Preprint at http://arXiv.org/abs/2204.05862 (2022)

58. Zhou Y, Lei T, Liu H, et al. Mixture-of-experts with expert choice routine. *Advances in Neural Information Processing Systems*, 35: 7103-7114, (2022).

59. DeepSeek-AI, Aixin Liu, Bei Feng, et al. DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model. Preprint at http://arXiv.org/abs/2405.04434 (2024)

60. DeepSeek-AI, Daya Guo, Dejian Yang, et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. Preprint at http://arXiv.org/abs/2501.12948 (2024)

61. 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)

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

63. 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)

64. 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)

65. 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)

66. 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)

67. 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)

68. 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)

69. 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

70. 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)

71. 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)

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

73. 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)

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

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

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

77. 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)

78. 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)

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

80. 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)

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

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

83. 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)

84. 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)

Acknowledgements

This work is supported by National MCF R&D program of China under Grant No. 2024YFE03240100 and National Natural Science Foundation of China under Grant No. U21A20440. The authors wish to thank Professor Yi Tan and Dr. Yuhang Luo for their support to adapt FusionMAE onto SUNIST-2. And thank all the members at Southwestern Institute of Physics, StarTorus Fusion and Tsinghua University for their contributions to the collection of dataset and development of algorithm.