PAPER • OPEN ACCESS

A machine-learning-based tool for last closed-flux surface reconstruction on tokamaks

OPEN ACCESS
IOP Publishing | International Atomic Energy Agency
Nuclear Fusion
Nucl. Fusion **63** (2023) 056019 (13pp)
https://doi.org/10.1088/1741-4326/acbfcc

A machine-learning-based tool for last closed-flux surface reconstruction on tokamaks

Chenguang Wan[1,2,*], Zhi Yu[1], Alessandro Pau[3], Olivier Sauter[3], Xiaojuan Liu[1], Qiping Yuan[1] and Jiangang Li[1,2,*]

[1] Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
[2] University of Science and Technology of China, Hefei 230026, China
[3] École Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center (SPC), CH-1015 Lausanne, Switzerland

E-mail: chenguang.wan@ipp.ac.cn and j_li@ipp.ac.cn

Received 6 December 2022, revised 30 January 2023
Accepted for publication 28 February 2023
Published 4 April 2023

Abstract

Tokamaks allow to confine fusion plasma with magnetic fields. The prediction/reconstruction of the last closed-flux surface (LCFS) is one of the primary challenges in the control of the magnetic configuration. The evolution in time of the LCFS is determined by the interaction between the actuator coils and the internal tokamak plasma. This task requires real-time capable tools to deal with high-dimensional data and high resolution at same time, where the interaction between a wide range of input actuator coils with internal plasma state responses adds an additional layer of complexity. In this work, we present the application of a novel state-of-the-art machine learning model to LCFS reconstruction in an experimental advanced superconducting tokamak (EAST) that learns automatically from the experimental data of EAST. This architecture allows not only offline simulation and testing of a particular control strategy but can also be embedded in a real-time control system for online magnetic equilibrium reconstruction and prediction. In real-time modeling tests, our approach achieves very high accuracies, with an average similarity of over 99% in the LCFS reconstruction of the entire discharge process.

Keywords: time series, magnetic reconstruction, tokamak

(Some figures may appear in colour only in the online journal)

* Authors to whom any correspondence should be addressed.

1. Introduction

One core research of tokamak physics is the regulation of the magnetic field distribution, which is needed to keep the plasma confined. Magnetic control is not trivial, in particular for advanced configurations, because the resulting distribution of the magnetic fields is determined by the interaction of complex, sometimes unpredictable plasma state evolution with a wide range of actuator inputs. Therefore, tools capable of efficiently and reliably reconstructing the evolution of magnetic fields [1–4] are paramount for the design of experiments and the development of robust control strategies. The conventional approach to this time-varying, non-linear, high-dimensional task is to solve an inverse problem to precompute a set of actuator coil (poloidal field coils, etc.) currents and voltages [3, 5, 6]. Then, the real-time estimates of the tokamak plasma equilibrium through a simulation code [4, 7, 8] allow modulating actuators' coil voltages to achieve the desired target. Although these physical simulation codes are usually effective, they require substantial effort and expertise by physicists to adapt a model whenever the tokamak magnetic configuration is changed. To overcome these bottlenecks, the fusion community has recently started investigating machine learning (ML) and artificial intelligence capabilities to reduce the complexity of models and numerical codes.

Full tokamak discharge modeling is also a critical task from a computational point of view. The typical workflow required for tokamak modeling, known as 'Integrated Modeling' [9], is computationally very expensive. For instance, a few seconds of discharge process generally takes hours to days of computation for high fidelity simulations. Moreover, the integration of the many physics processes required to describe the evolution of the plasma state adds an even further layer of complexity. In this context, a common approach is to replace high fidelity simulation codes with ML-based surrogate models. This allows us to accomplish the same task, significantly reducing computation time while preserving a reasonable level of accuracy.

Recently, various applications in magnetic confinement fusion research have relied on ML approaches to solve a variety of problems, such as disruption prediction [10–16], electron temperature profile estimation [17], surrogate model [18–20], plasma tomography [21], radiated power estimation [22], discharge estimation [23, 24], identifying instabilities [25], neutral beam effects estimation [26], classifying confinement regimes [27], the determination of scaling laws [28, 29], filament detection [30], coil current prediction with the heat load pattern [31], equilibrium reconstruction [17, 32–36] and equilibrium solver [37], control plasma [38–43], physics-informed ML [44], and reinforcement-learning-informed magnetic field control [3]. In particular, the use of reinforcement learning for magnetic field control work has a different target from our work, which is the design of a controller for magnetic control during the flat-top of the plasma current. The conventional controller should take over in the ramp-up and ramp-down phases.

Modeling the entire tokamak discharge process by leveraging ML approaches is challenging from both technical and computational viewpoints. The duration of a plasma discharge in an experimental advanced superconducting tokamak (EAST) [45] can be of the order of thousands of seconds, with a resulting sequence length that exceeds 1×10^6 if the sampling rate is 1 kHz. There are different classes of ML models to deal with sequence problems, recurrent neural networks (RNNs) [46], transformers [47] based on the self-attention mechanism, and several variants. In ML, attention is a technique designed to simulate cognitive attention. The result is that some parts of the input data are enhanced, whereas others are diminished. This is done so that the neural network should exert more attention on the small but important parts of the data. The self-attention mechanism allows input data to interact with each other ('self') and find out where they should pay more attention to ('attention'). For traditional RNN algorithms, training and inference time on long sequences are usually slow. The sequential nature of RNN models prevents in general achieving a high level of parallelization in computations. From an ML perspective, the processing of long time sequences characterized by short- and long-term dependencies is still an outstanding challenge. In a plethora of deep learning models, transformers are a novel architecture, which allows overcoming some of the aforementioned issues, thanks to the multi-head attention mechanism. Nevertheless, also the use of transformers for modeling long sequences presents some limitations because of their computational complexity $O(n^2 d)$, where n is the sequence length. In practice, when the sequence length is of the order of thousands of samples, and we are dealing with high-dimensional data, training and inference times start to become unacceptable for most of the applications.

Magnetic field reconstruction has two research paradigms: physics-driven and data-driven approaches. Physics-driven approaches in magnetic field reconstruction have been studied in the last decades, resulting in the development of various simulation codes, such as equilibrium fitting (EFIT) [48–50], LIUQE [51], and RAPTOR [52]. The adaptation of these codes to new target plasma configurations or to new machines requires a non-negligible effort. This aspect, together with the aspect of computational efficiency, has recently brought the fusion community to leverage more and more data-driven methods to solve tasks at different levels of complexity. However, magnetic reconstruction is far behind other applications in fusion. To the best of our knowledge, only a few works, such as [3], have actually been deployed and successfully tested in a real environment.

In this paper two different variants of 1D-shifted windows transformer model (1D-SWIN transformer) have been proposed for real-time and offline magnetic reconstruction of the last closed-flux surface (LCFS), respectively. In the case of the 1D-SWIN Transformer, the model's computational

complexity depends linearly on the sequence length n. Moreover, these models can take advantage of a high level of parallelization, thanks to the attention mechanism and the non-sequential nature of the algorithm. The models presented in this work are trained only on experimental data and can be used for the estimation of the magnetic field evolution for the entire length of the tokamak discharge, including the ramp-up and the ramp-down phases of the plasma current.

As far as the real-time estimation of the magnetic fields' evolution is concerned, the model is not directly used to control the magnetic field. It is able to predict the evolution of the magnetic field one step ahead in the future, allowing the design of more effective feedback control strategies. The real-time model can be integrated within the plasma control system (PCS) to assist robust magnetic control by predicting the magnetic field in the subsequent time step. The offline model remarkably reduces the execution time required to simulate the evolution of the magnetic field for the entire discharge. Moreover, when coupled to other ML-based surrogate models for the prediction of 0D quantities like in [24], it allows to simulate the evolution of various quantities of interest, supporting the experimental design and the optimization of the target plasmas. Compared to the model described in [3], our model does not rely on a physics simulation code, whose computational complexity cannot be ignored. Furthermore, given the regression task, the training of our model is in general more efficient than the training of a model based on reinforcement learning. Furthermore, the reinforcement learning model in magnetic control involves an agent exploring the unknown control space to achieve fine control over the tokamak's magnetic field surface. The convergence of reinforcement learning is difficult. The reinforcement learning model must interact with the simulation code for each control command generated for the magnetic field control task, which further reduces the efficiency of the model training. The regression task has clear targets and inputs, allowing the model to be trained more efficiently. Another non-negligible aspect, which is of increasing importance in fusion and in many other fields of science, is that transformers have become particularly successful when used in the context of transfer learning. The key concept is that the model has the capability to learn the underlying dynamics characterizing the evolution of the magnetic field in a tokamak, encoding this knowledge in a reduced latent space representation that can be 'easily' adapted to new devices. Such a perspective is extremely attractive and would allow to significantly optimize the exploitation of fusion devices for more and more advanced studies.

According to the main quantities required for magnetic field control [2], the data used to build the ML model are primarily magnetic signals and references for control, namely magnetic surface probe data, in-vessel currents, poloidal field coils data, flux loop data, plasma current, and shape references. For the real-time version of the model, the average similarity is over 99%, and the inference time is 0.7 ms (<1 ms in accordance with the typical control system requirements). For the offline version of the model, the average similarity is over 93%, and the average inference time is ∼0.22 s for a sequence length of 1×10^6, which is lower than the real-time model because of different settings (as it will be discussed in the following sections).

Our contributions are summarized as follows:

(i) We propose a generalized 1D shifted windows transformer architecture that can compute long time series.
(ii) One of the models can be integrated into tokamak control for estimating the real-time magnetic field in advance.
(iii) One of the models can also be combined with a 0D proposal estimation model to give a complete prediction of experimental proposal results.
(iv) The validity of the proposed model is demonstrated over a large experimental dataset of the EAST tokamak.

This paper is organized as follows. First, section 2 describes our ML model, used dataset, and the model training. Then, section 3 presents our model results and a short analysis. Finally, section 4 provides a brief discussion and conclusion.

2. Methods

Our main workflow is shown in figure 1. In this section, we will introduce details in terms of machine learning model design, dataset selection, and model training.

2.1. ML model

The general architecture of our ML models is shown in figure 2. Our architecture uses a customized 1D-shifting window attention mechanism inspired by the SWIN transformer [53] to model long-term dependencies and interactions between inputs and outputs. We stack self-attention blocks to build the ML model.

In the framework of deep learning, there are four main candidate architectures for modeling such long-time sequences: convolutional neural network (CNN), RNN, such as long-short term memory, gated recurrent unit, transformer, and our customized 1D SWIN transformer. In addition, some critical quantitative criteria should be considered for modeling tokamak magnetic probe data: computational complexity, number of sequential operations, and maximum path length [54]. Table 1 demonstrates that 1D-shifted window attention has roughly as many sequential operations and computational complexities as CNNs. Generally, the attention mechanism can achieve superior performance with respect to CNN in numerous time sequence tasks, such as natural language processing (NLP) [47, 55].

Generally speaking, some differences should be present between the real-time and offline model-building strategies.

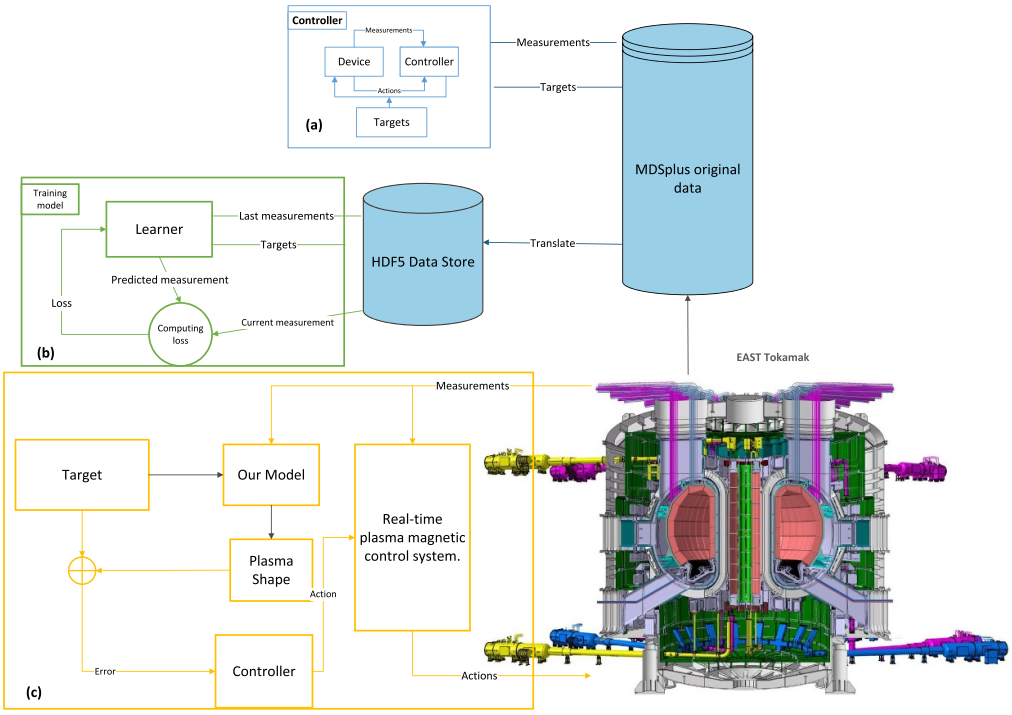

Figure 1. Representation of the components of our machine learning model design and usage. (*a*) The conventional controller working loop. The controller measurements difference between targets and the magnetic probe measurement values at the *current* time. According to the difference, the controller sends actions to the actuator coils. (*b*) Sketch of the learning loop. The learner reads the measurements and targets from the HDF5 data store, and then computes the loss between the predicted magnetic field and the target magnetic field. Finally, using the loss as the criterion to train the learner. (*c*) The online usage for the tokamak control. Our model can predict the tokamak's last closed-flux surface (LCFS), the controller reads the estimation to generate the next control action sent to magnetic coils.

The real-time model requires that the single-step inference is fast enough. That is, the one-step inference time of the model should be less than the response time required by the control system, and the actual system output of the previous step can be fed back as input to the model. According to the requirements of the EAST magnetic control system, the model inference time should be less than 1 ms. For a typical transformer model, a single-step input is complex. If the preset control commands are modified, the whole sequence needs to be recalculated, making the inference time exceed the control system requirements. In our work, we let 'window size' $=$ 1, which makes our model calculate the attention only in the channel axis, and the single-step input becomes less expensive. This design of the model results in a one-step inference time of $\sim$0.7 ms, allowing satisfying real-time constraints. For the offline model, the actual system output from the previous step should not be fed back as an input unless it is trained using the teaching force trick. The teaching force trick is a trick for training a neural network that uses ground truth as input. In our case, the ground truth is the actual tokamak responses to the control commands. The time requirement of the offline mode can be relaxed, but it should generally be within 1 h. Otherwise, the advantage of the ML model over the integrated modeling model will be diminished. If we use the teaching force, we have to recompute all the past sequences step by step, so the inference time of the entire sequence will be in the order of 1×10^5 s for the reason of the computational complexity. This paper's offline model does not use the teaching force trick because the inference time requirement is much shorter than 1 h.

2.2. Dataset

In this paper, a total of 16,609 discharges of the EAST tokamak (discharge range between #56804 and 96915) were selected to construct the total dataset. The training set, validation set, and test set are divided in chronological order. The training set has 14,732 discharges, the validation set has 200 discharges, and the test set has 1677 discharges. In the experimental range #56804–96915, there are only 30 long discharge discharges (discharge time $>$50 s), 10 of which are included in the training set, and the remaining 20 discharges are included in the test set. The validation set is relatively small because the model does not update parameters during the validation phase, and a relatively small validation set can speed up model training. As shown in table 2, we have selected the reference of plasma current , the in-vessel current IC1, the poloidal field coil current, the reference of poloidal field coils, the shape reference as the input signals, and the output signals include all magnetic probe signals of the magnetic

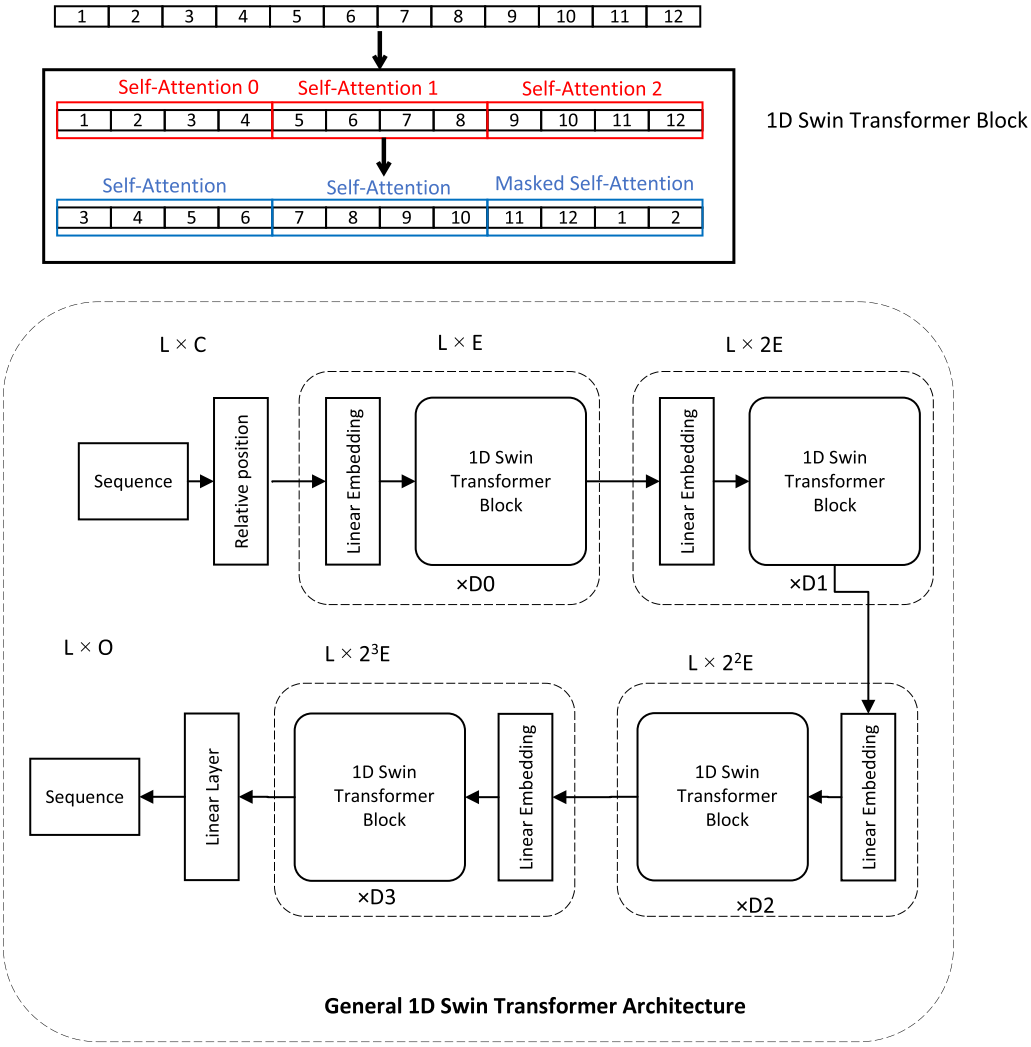

Figure 2. Our machine learning model architecture. In the figure, 'L' is sequence length, 'E' is the embedded dimension, 'C' is the input sequence channels number, and 'O' is the output sequence channels number.

Table 1. CNN, RNN, transformer, 1D SWIN transformer comparison.

Model type	Computational complexity	Sequential operation	Maximum path length
CNN	$O(knd^2)$	$O(1)$	$O(n/k)$
RNN	$O(nd^2)$	$O(n)$	$O(n)$
Transformer	$O(n^2d)$	$O(1)$	$O(1)$
1D SWIN transformer	$O(nwd)$	$O(1)$	$O(n/w)$

where k is kernel size of CNN, d is sequence dimension, n is sequence length, w is the window size of 1D SWIN transformer.

field. Because the in-vessel current IC1 could not be obtained in advance at the experimental proposal stage, the input signals of the offline model did not include IC1, and the output signals of previous step data were not input to the offline model for efficiency reasons. All data were uniformly sampled at 1 kHz for the entire length of the discharge, and all time axes were aligned to the same time-base. Data were saved to HDF5 files discharge-by-discharge. For fast and robust training, each discharge experiment was saved as a separate HDF5 file, with 209 gigabytes of original data.

2.3. Model training

Before the model is trained, each signal's mean, variance, and presence flag are calculated for each discharge, and then the data are stored in a MongoDB database. The data are then

Table 2. The input and output of signals of the models.

Signal	Physical meaning	Number of channels	Meaning of channels
Output signals		73	
BP	Equilibrium magnetic probes	38	38 magnetic probes data
FL	Flux loops	35	35 flux loops data
Input signals		57	
Ref. I_p	Reference of plasma current	1	Plasma current reference
IC1[a]	In-vessel coil no.1 current	1	In-vessel coil no.1 current
PF	Poloidal field coils voltage	12	poloidal field no.1-12 coil current
Ref. PF	Nominal current of poloidal field coils	12	Nominal current of poloidal field no. 1-12 coil
Ref. shape	Shape reference	31	20 groups of control points

[a] only used in real-time version.

Table 3. Our model hyperparameters. The model architecture can be found in figure 2.

Hyperparameter	Explanation	Best value of the real-time model	Best value of the offline model
η	Learning rate	1×10^{-4}	1.5×10^{-4}
Optimizer	Optimizer type	SGD	SGD
Loss	Loss function	MaskedMSELoss	MaskedMSELoss
Epoch	Number of epochs	40	35
Scheduler	Scheduler type	OneCycle [58]	OneCycle
Window_size	Window size	1	12
C	Input channel	130	56
E	Embedded dimension	60	36
[D0, D1, D2, D3]	Depth list for layers	[2, 2, 4, 2]	[2, 2, 4, 2]

normalized for each discharge and finally fed into the ML model for training. The input set is different for the offline model and the real-time model. As analyzed in section 2.1, the real-time model input dimension is 130, which includes the system output at the previous step and the current IC1 signal. We can use the teaching force for training, and IC1 can be obtained in real-time experiments. For the offline model, the input dimension is 56 because the IC1 and the system output at the previous step are not used.

Both versions of the model use Centos OS 7 executing on eight P100 GPU cards. During the training of our model, we used a custom masked mean square error (MSE) loss function (MaskedMSELoss).

$$l(\boldsymbol{x},\boldsymbol{y}) = L = \frac{\sum_{i=0}^{i=N}\{l_1, l_2, \ldots, l_N\}}{N}, \tag{1}$$

$$l_i = \sum_{j=0}^{j=\text{len}} f_i \cdot \left(\boldsymbol{x}_j^i - \boldsymbol{y}_j^i\right)^2, \tag{2}$$

where $\boldsymbol{x}$ is batch experimental sequence data, $\boldsymbol{y}$ is batch-predicted sequence result, $\boldsymbol{x}_j^i$, $\boldsymbol{y}_j^i$ are the jth point values of the ith experimental sequence and predicted sequence. f_i is a signal data existence vector of ith experimental sequence, f_i equals to 1 when the sequence exists and 0 otherwise. f_i is used to mask a signal that does not have original data. The $\sum_{j=0}^{j=\text{len}}$ is another mask for the invalid length of the sequence. This term prevents training on the zeros padding of the sequence. The use of existence masks and length masks can prevent models from being trained on sequences without actual target values and meaningless zeros padding tails. The zero padding tail comes from the fact that we use zeros to pad sequences within a batch to the same length. This improves the accuracy and speed of the training process, where we used the bucketing algorithm [56] for training acceleration, and the Tree of Parzen Estimator algorithm [57] for the architectural hyperparameter search. We also tested various optimizers and regulators and finally obtained the optimal set of hyperparameters as shown in table 3.

3. Results

We trained, validated, and tested real-time and offline versions of the proposed transformer-based model on the dataset during the 2016–2020 EAST campaigns with discharge numbers in

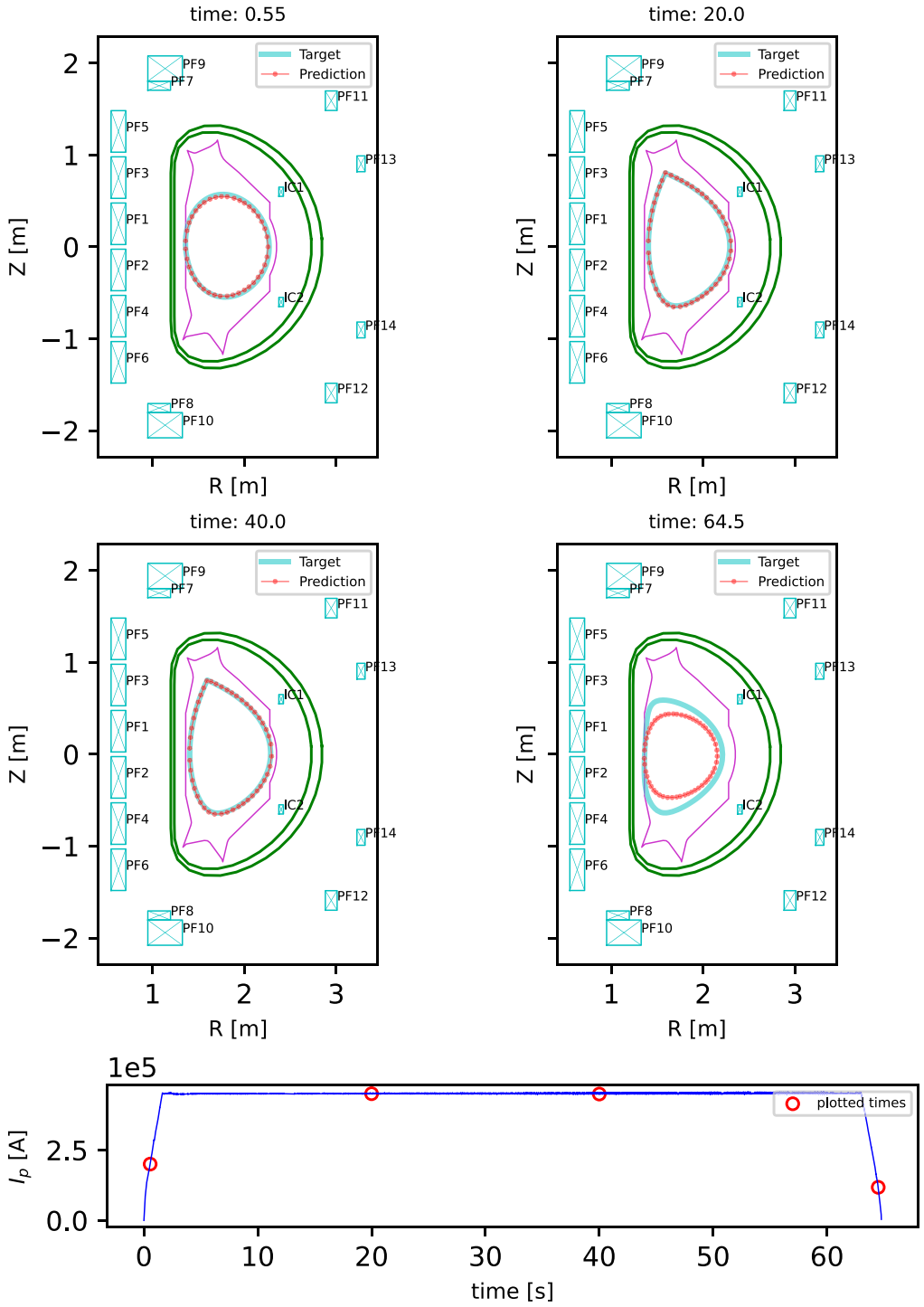

Figure 3. Discharge #73678 offline LCFS reconstruction. The LCFS was generated from EFIT using inputs from our model. The solid blue lines are the experimental target LCFS, the red 'star' markers are predicted LCFS.

the range #52804–88283 [59–61], whereas input and output signals can be found in section 2.2.

3.1. Offline model results

Figure 3 shows our offline model prediction for the LCFS in the EAST discharge #73678. The duration of this discharge is longer than 70 s, with the sequence length of $\sim 7 \times 10^4$, which is a typical long sequence modeling problem. The LCFS shown in this figure is generated through the equilibrium reconstruction code EFIT [50] by inputting the magnetic quantities predicted by the model into EFIT. The equilibrium reconstruction is a broad topic in tokamak research, extensively discussed in various papers and main plasma

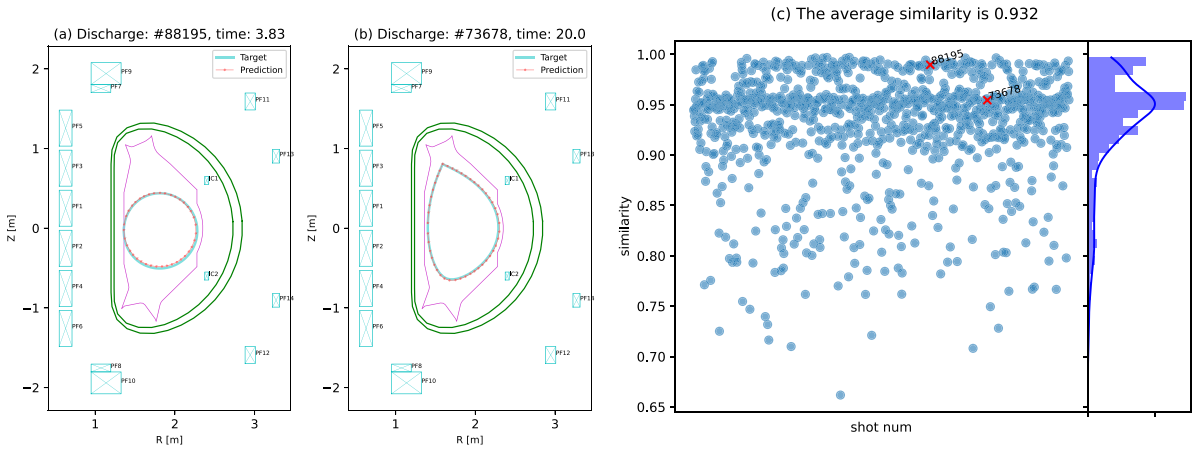

Figure 4. (*a*), (*b*) Discharges #88195 and #73678 offline LCFS reconstruction on the flat-top phase. (*c*) Similarity distribution of offline model predicted results on the test set. The test set (section 2.2) is in discharge range #82651–88283 and some long-time discharges for a total of 1677 discharges. Discharge #88195 is the limiter ('circle') configuration as shown in (*a*). Discharge #73678 is the diverter ('single null' or 'double null') configuration as shown in (*b*) (see details in figure 3).

physics books [62]; therefore, it will not be addressed in this paper. Figure 3 shows that the model can reconstruct the LCFS with high accuracynot only during the flat-top phase of the plasma current but also in the ramp-up and ramp-down phases, which are non-stationary phases. The model is able to reproduce the magnetic configuration during the various discharge phases, from the tokamak start-up 'circle,' going through the formation of a 'single null' shape, to the characteristic shapes in the shutdown 'circle.'

The performance of the model has been evaluated with the same similarity indicator discussed in [23]. The average similarity in the test set for the offline version of the model (figure 4) is 93.2%. Most of the discharges are concentrated around 95%, with the bulk of the distribution above 90%. The test set for this work comprises experiments in the discharge range #82651–88283 for a total of 1677 discharges, some of which have a very long duration (see details in section 2.2). Note that the similarity is computed on raw signal data instead of the reconstructed LCFS. As far as experiments with similarity less than 0.85 are concerned, there are 98 discharges, among which 89 are disruptions, whereas 9 are discharges with regular terminations. A disruption is an unexpected termination of the discharge where the plasma loses abruptly its thermal and magnetic confinement, involving huge electromagnetic forces and thermal loads, which can potentially damage the machine. Apart from experiments dedicated to the study of disruption physics and to the assessment of engineering limits during these violent transients, the design of the discharge itself together with robust real-time control strategies aim to avoid disruptions. Nevertheless, when operating close to stability limits, various sequences of events can potentially lead to disruption, strongly affecting the magnetic equilibrium and making it unavoidably deviate from offline modeling. The operational space characterizing disruptions is extremely complex and wide, making its coverage within the input domain unfeasible. The nine regular terminations with relatively high error are not well estimated probably because of inherent limitations in the model, or inaccuracies in the measurements, but they correspond to only the 0.5% of the test set. The similarity distribution in figure 4(*c*) has two peaks, one at ∼98% and the other at ∼95%. We have checked the related discharge numbers and found that peak 1 is the limiter 'circle' configuration (figure 4(*a*)), and peak 2 is the normal diverter ('single null' or 'double null') configuration (figure 4(*b*)). Based on these observations, we believe that the reason for the two peaks is that there are two modes in the test set, which are the limiter and diverter configurations. The limiter configuration is simple and easy to reconstruct offline; therefore, the similarity is higher. In contrast, the diverter configuration is complex and difficult to reconstruct offline; hence, the similarity is lower.

3.2. *Real-time model results*

The real-time model differs from the offline model both in terms of input quantities and inference time requirements (discussed in detail in section 2.1). Figure 5 shows the reconstruction results of the real-time model for the discharge #73678. In real-time settings, the real measurement of the magnetic field probe at the previous step is fed as an input to simulate the actual tokamak's control feedback process.

The similarity of the real-time model in the test set is shown in figure 6, which is the same test set as the offline model.

However, there is almost no difference between the modeling results of discharge #73678 in figures 3 and 5 on the flat-top phase. Figures 7 and 8 show that the main errors of the model prediction are concentrated in the ramp-up and ramp-down phases, especially in the ramp-down phases. The

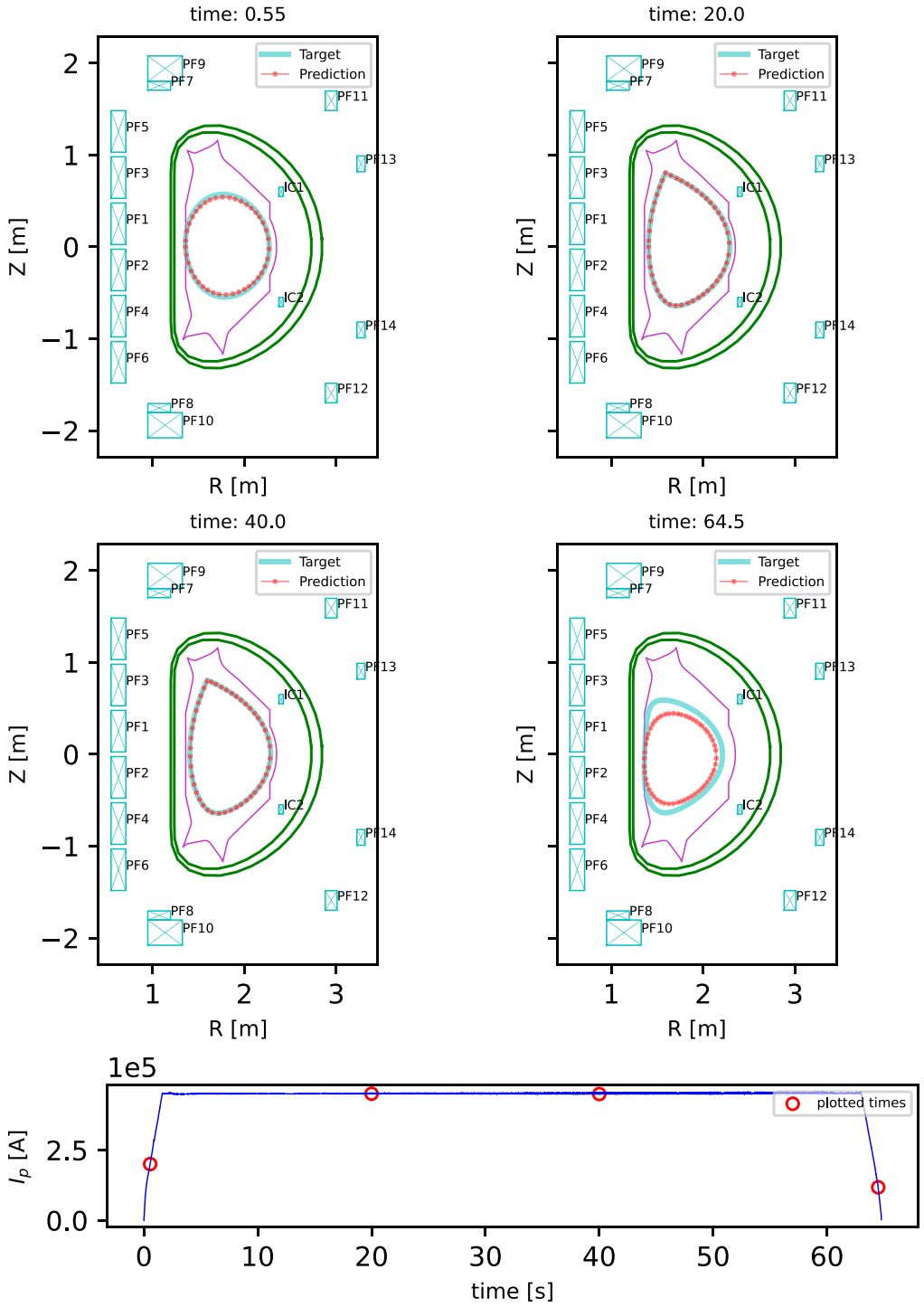

Figure 5. Discharge #73678 real-time LCFS reconstruction. The LCFS was generated by the same method as offline magnetic reconstruction, figure 3. The solid blue lines are the target LCFS, the red 'star' markers are predicted LCFS.

possible reasons for these errors are: 1. EFIT code result is unstable when the plasma current is small [49]. 2. The magnetic field changes rapidly in the ramp-up and ramp-down phases, and the sampling rate of 1 ms may not fully capture the changes.

The comparison of figures 4 and 6 reveal that the real-time model performs slightly better than the offline model. A possible reason is that the plasma magnetic field is not a rapidly time-varying process, and the system output at the current time step is a good 'guide' to forecast the evolution of the system

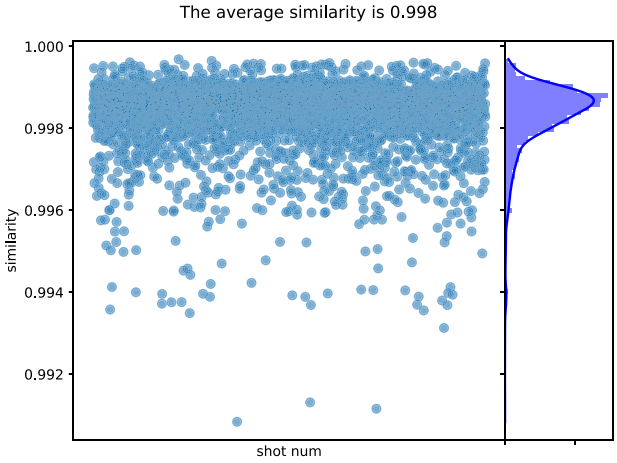

Figure 6. Similarity distribution of real-time model predicted results on the test set. The test set for the real-time and the offline models are the same.

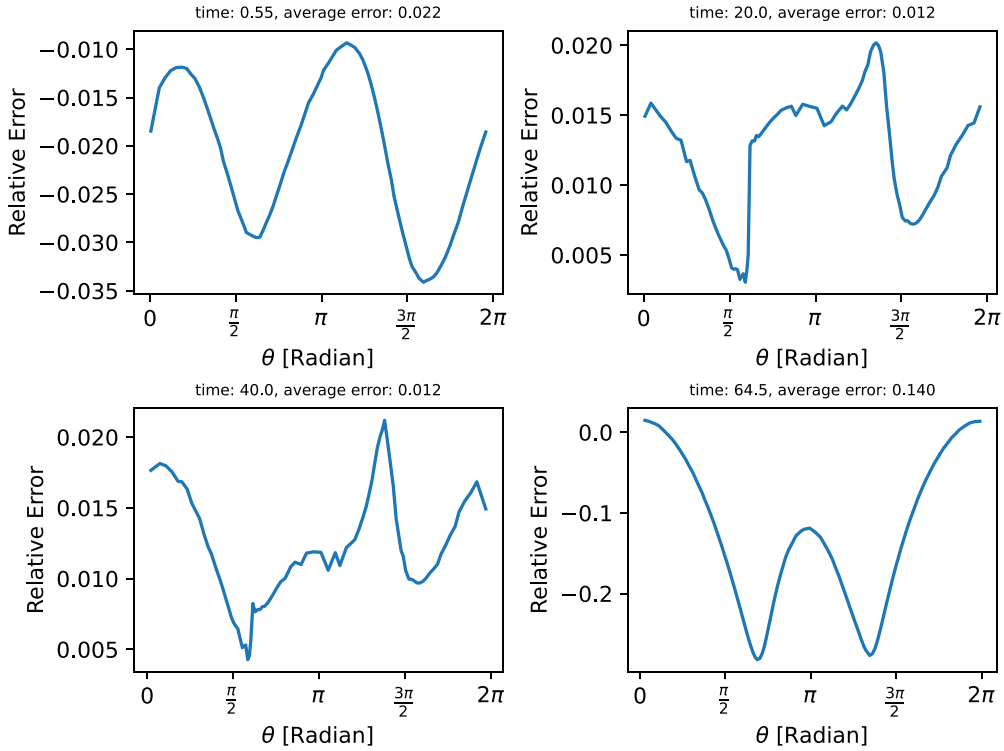

Figure 7. Discharge #73678 offline LCFS reconstruction relative error on the poloidal angle. The relative error is defined as $(a-\hat{a})/a$, a and $\hat{a}$ are the target and prediction minor radii.

in the subsequent time step. Figure 6 has only one peak for possibly the same reason. However, the offline model has no knowledge of the actual tokamak output, so even if bigger and more computationally demanding models are used for the offline task, the results are a bit less accurate compared to the real-time model.

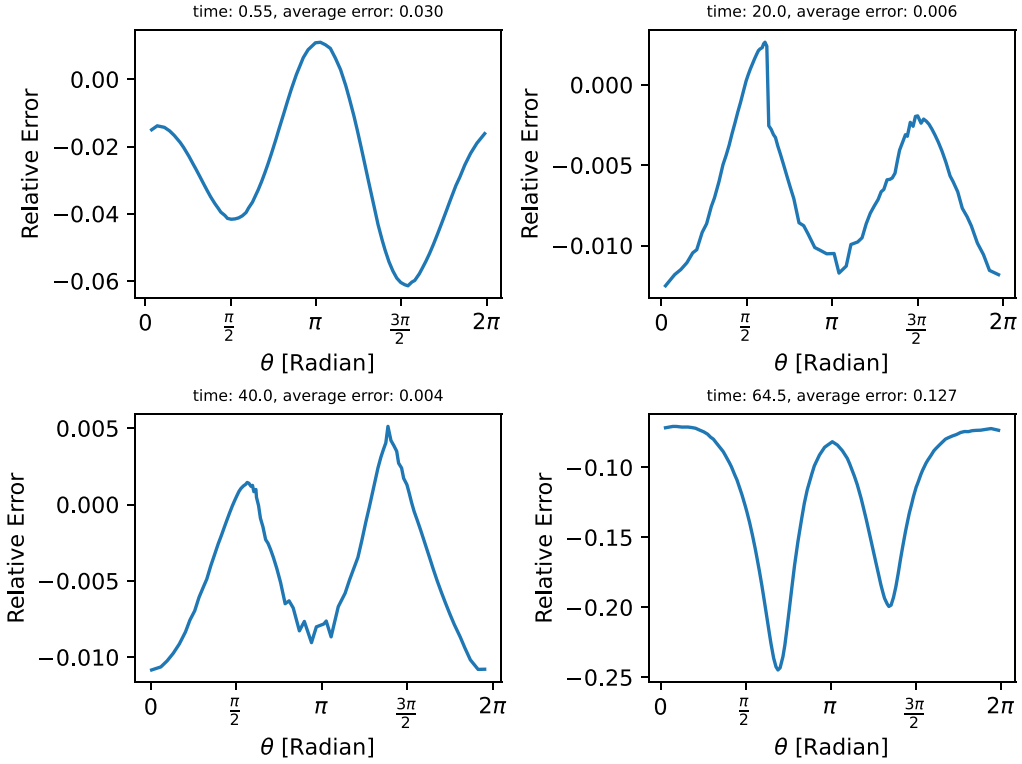

Figure 8. Discharge #73678 online LCFS reconstruction relative error on the poloidal angle. The definition of the relative error is the same as for offline magnetic reconstruction, figure 7.

4. Discussion and conclusion

In the current work, we propose a 1D shifted windows transformer model that can work with long sequences (up to a sequence length of 1×10^6 for LCFS reconstruction in this work), which reduces the computational complexity of the original model from a square to a linear dependence on the sequence length. The proposed model can form a general sequence processing backbone network for both real-time and offline sequence modeling. Thanks to the reduced computational complexity, the model can be efficiently used for very long sequences, exceeding a sequence length of 1×10^6 , as we demonstrate in this study. To the best of our knowledge, we have achieved the first data-driven modeling of the LCFS for the whole tokamak discharge, including the ramp-up and the ramp-down phases of the plasma current. Being dynamic phases, ramp-up and ramp-down are in general more difficult to model, and as such they are often not taken into account in data-driven applications. The inference time for the real-time task (one-step ahead forecasting) is $\sim$0.7 ms with an average similarity of $>$99%, whereas the average inference time for the offline modeling (entire discharge process) is 0.22 s with an average similarity of $>$93%.

From the ML point of view and to the best of our knowledge, this work is also the first proposing an attention-based mechanism for successfully modeling long time sequences. From the point of view of tokamak physics research, we have achieved high accuracy and fast tokamak magnetic field modeling, which can be used for critical applications, such as real-time control or offline validation of tokamak experimental proposals. When integrated with other existing discharge modeling data-driven frameworks, such as [24], the proposed approach can represent an extremely valuable tool to advance in the development of robust and high-performance tokamak scenarios. A first important milestone for the future will be the actual integration of real-time models within plasma control systems, which is of paramount importance to understand how reliable these systems are when operating routinely in real environments. Another exciting future perspective triggered by the achievements documented in this work is the validation of the full modeling of the plasma discharge, integrating magnetic reconstruction with the prediction of key 0-D physics quantities commonly describing the outcome of a plasma discharge. Finally, extending and testing the 1D shifted windows transformer to other general areas of ML, such as NLP, is also an exciting direction for prospective research.

Data availability statement

The data that supports the findings of this study belongs to the EAST team and is available from the corresponding author upon reasonable request.

Acknowledgment

The model training in this paper were performed on the ShenMa High Performance Computing Cluster in Institute of

Plasma Physics, Chinese Academy of Sciences. The authors would like to thank all the members of EAST Team for providing such a large quantity of tokamak experimental data. And especially thank Dr Peifeng Fan, Dr Ruirui Zhang, Dr Heru Guo, and other members of EAST Division of Control and Computer Application for explaining the experimental data.

This work was supported by the National Key R&D project under Contract No. Y65GZ10593, the National MCF Energy R&D Program under Contract No. 2018YFE0304100, the Comprehensive Research Facility for Fusion Technology Program of China under Contract No. 2018-000052-73-01-001228, the National MCF Energy R&D Program of China under Grant No. 2018YFE0302100, and National Nature Science Foundation of China under Grant No. 12075285. This work was also supported in part by the Swiss National Science Foundation.

Code availability

The model code is open-source and can be found in github https://github.com/chgwan/1DSwin. The other codes for model training, data acquisition, and generate Figures belong to EAST team and are available from the corresponding author upon reasonable request.

ORCID iDs

Chenguang Wan https://orcid.org/0000-0002-6005-4460
Zhi Yu https://orcid.org/0000-0003-0000-8750
Alessandro Pau https://orcid.org/0000-0002-7122-3346
Olivier Sauter https://orcid.org/0000-0002-0099-6675
Xiaojuan Liu https://orcid.org/0000-0002-0331-8730
Qiping Yuan https://orcid.org/0000-0003-4292-1302

References

[1] Wesson J. and Campbell D.J. 2011 *Tokamaks* vol 149 (Oxford: Oxford University Press)
[2] De Tommasi G. 2019 Plasma magnetic control in tokamak devices *J. Fusion Energy* **38** 406–36
[3] Degrave J. *et al* 2022 Magnetic control of tokamak plasmas through deep reinforcement learning *Nature* **602** 414–9
[4] Moret J.-M., Duval B.P., Le H.B., Coda S., Felici F. and Reimerdes H. 2015 Tokamak equilibrium reconstruction code LIUQE and its real time implementation *Fusion Eng. Des.* **91** 1–15
[5] Walker M.L. and Humphreys D.A. 2006 Valid coordinate systems for linearized plasma shape response models in tokamaks *Fusion Sci. Technol.* **50** 473–89
[6] Blum J., Heumann H., Nardon E. and Song X. 2019 Automating the design of tokamak experiment scenarios *J. Comput. Phys.* **394** 594–614
[7] Ferron J.R., Walker M.L., Lao L.L., John H.E.S., Humphreys D.A. and Leuer J.A. 1998 Real time equilibrium reconstruction for tokamak discharge control *Nucl. Fusion* **38** 1055–66
[8] Carpanese F., Felici F., Galperti C., Merle A., Moret J.M. and Sauter O. 2020 First demonstration of real-time kinetic equilibrium reconstruction on TCV by coupling LIUQE and RAPTOR *Nucl. Fusion* **60** 066020
[9] Falchetto G.L. *et al* 2014 The European Integrated Tokamak Modelling (ITM) effort: achievements and first physics results *Nucl. Fusion* **54** 043018
[10] Kates-Harbeck J., Svyatkovskiy A. and Tang W. 2019 Predicting disruptive instabilities in controlled fusion plasmas through deep learning *Nature* **568** 526–31
[11] Hu W.H.H. *et al* 2021 Real-time prediction of high-density EAST disruptions using random forest *Nucl. Fusion* **61** 066034
[12] Guo B.H., Chen D.L., Shen B., Rea C., Granetz R.S., Zeng L., Hu W.H., Qian J.P., Sun Y.W. and Xiao B.J. 2021 Disruption prediction on EAST tokamak using a deep learning algorithm *Plasma Phys. Control. Fusion* **63** 115007
[13] Cannas B., Fanni A., Sonato P. and Zedda M.K. 2007 A prediction tool for real-time application in the disruption protection system at JET *Nucl. Fusion* **47** 1559–69
[14] Cannas B., Alessandra Fanni G.P., Sias G. and Sonato P. 2010 An adaptive real-time disruption predictor for ASDEX Upgrade *Nucl. Fusion* **50** 075004
[15] Yoshino R. 2003 Neural-net disruption predictor in JT-60U *Nucl. Fusion* **43** 1771–86
[16] Pau A., Fanni A., Carcangiu S., Cannas B., Sias G., Murari A. and Rimini F. 2019 Human immunodeficiency virus, associated neurocognitive disorders, consensus report, mind corresponding author and alternate corresponding author. A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET *Nucl. Fusion* **59** 106017
[17] Clayton D.J., Tritz K., Stutman D., Bell R.E., Diallo A., LeBlanc B.P. and Podestà M. 2013 Electron temperature profile reconstructions from multi-energy SXR measurements using neural networks *Plasma Phys. Control. Fusion* **55** 095015
[18] Honda M. and Narita E. 2019 Machine-learning assisted steady-state profile predictions using global optimization techniques *Phys. Plasmas* **26** 102307
[19] Meneghini O. *et al* 2017 Self-consistent core-pedestal transport simulations with neural network accelerated models *Nucl. Fusion* **57** 86034
[20] van de Plassche K.L., Citrin J., Bourdelle C., Camenen Y., Casson F.J., Dagnelie V.I., Felici F., Ho A. and Van Mulders S. 2020 Fast modeling of turbulent transport in fusion plasmas using neural networks *Phys. Plasmas* **27** 022310
[21] Ferreira D.R. and Carvalho P.J. 2020 Deep learning for plasma tomography in nuclear fusion pp 1–5
[22] Barana O., Murari A., Franz P., Ingesson L.C. and Manduchi G. 2002 Neural networks for real time determination of radiated power in JET *Rev. Sci. Instrum.* **73** 2038–43
[23] Wan C., Zhi Y., Wang F., Liu X. and Jiangang L. 2021 Experiment data-driven modeling of tokamak discharge in EAST *Nucl. Fusion* **61** 066015
[24] Wan C., Zhi Y., Pau A., Liu X. and Jiangang L. 2022 EAST discharge prediction without integrating simulation results *Nucl. Fusion* **62** 126060
[25] Murari A., Arena P., Buscarino A., Fortuna L. and Iachello M. 2013 On the identification of instabilities with neural networks on JET *Nucl. Instrum. Methods Phys. Res.* A **720** 2–6
[26] Boyer M.D., Kaye S. and Erickson K. 2019 Real-time capable modeling of neutral beam injection on NSTX-U using neural networks *Nucl. Fusion* **59** 056008
[27] Murari A., Mazon D., Martin N., Vagliasindi G. and Gelfusa M. 2012 Exploratory data analysis techniques to determine the dimensionality of complex nonlinear phenomena: the L-to-H transition at JET as a case study *IEEE Trans. Plasma Sci.* **40** 1386–94

[28] Murari A., Vega J., Mazon D., Patané D., Vagliasindi G., Arena P., Martin N., Martin N.F., Rattá G. and Caloone V. 2010 Machine learning for the identification of scaling laws and dynamical systems directly from data in fusion *Nucl. Instrum. Methods Phys. Res.* A **623** 850–4

[29] Gaudio P., Murari A., Gelfusa M., Lupelli I. and Vega J. 2014 An alternative approach to the determination of scaling law expressions for the L–H transition in Tokamaks utilizing classification tools instead of regression *Plasma Phys. Control. Fusion* **56** 114002

[30] Cannas B., Carcangiu S., Fanni A., Farley T., Militello F., Montisci A., Pisano F., Sias G. and Walkden N. 2019 Towards an automatic filament detector with a faster R-CNN on MAST-U *Fusion Eng. Des.* **146** 374–7

[31] Böckenhoff D., Blatzheim M., Hölbe H., Niemann H., Pisano F., Labahn R. and Sunn Pedersen T. 2018 Reconstruction of magnetic configurations in W7-X using artificial neural networks *Nucl. Fusion* **58** 56009

[32] Coccorese E., Morabito C. and Martone R. 1994 Identification of noncircular plasma equilibria using a neural network approach *Nucl. Fusion* **34** 1349–63

[33] Bishop C.M., Haynes P.S., Mike E.U.S., Todd T.N. and Trotman D.L. 1994 Fast feedback control of a high temperature fusion plasma *Neural Comput. Appl.* **2** 148–59

[34] Jeon Y.-M., Yong-Su N., Kim M.-R. and Hwang Y.S. 2001 Newly developed double neural network concept for reliable fast plasma position control *Rev. Sci. Instrum.* **72** 513–6

[35] Wang S.Y., Chen Z.Y., Huang D.W., Tong R.H., Yan W., Wei Y.N., Ma T.K., Zhang M. and Zhuang G. 2016 Prediction of density limit disruptions on the J-TEXT tokamak *Plasma Phys. Control. Fusion* **58** 055014

[36] Joung S., Kim J., Sehyun Kwak J.G.B., Lee S.G., Han H.S., Kim H.S., Lee G., Kwon D. and Ghim Y.-C.C. 2020 Deep neural network Grad-Shafranov solver constrained with measured magnetic signals *Nucl. Fusion* **60** 16034

[37] Ph B. van Milligen V.T. and Jiménez J.A. 1995 Neural network differential equation and plasma equilibrium Solver *Phys. Rev. Lett.* **75** 3594–7

[38] Bishop C.M., Haynes P.S., Mike E.U.S., Todd T.N. and Trotman D.L. 1995 Real-time control of a tokamak plasma using neural networks *Neural Comput.* **7** 206–17

[39] Yang B., Liu Z., Song X. and Xiangwen L. 2020 Design of HL-2A plasma position predictive model based on deep learning *Plasma Phys. Control. Fusion* **62** 125022

[40] Wakatsuki T., Suzuki T., Hayashi N., Oyama N. and Ide S. 2019 Safety factor profile control with reduced central solenoid flux consumption during plasma current ramp-up phase using a reinforcement learning technique *Nucl. Fusion* **59** 066022

[41] Rasouli H., Rasouli C. and Koohi A. 2013 Identification and control of plasma vertical position using neural network in Damavand tokamak *Rev. Sci. Instrum.* **84** 023504

[42] Yang B., Liu Z., Song X., Xiangwen L.I. and Yan L. 2020 Modeling of the HL-2A plasma vertical displacement control system based on deep learning and its controller design *Plasma Phys. Control. Fusion* **62** 75004

[43] Jaemin Seo Y.-S.N., Kim B., Lee C.Y., Park M.S., Park S.J. and Lee Y.H. 2021 Feedforward beta control in the KSTAR tokamak by deep reinforcement learning *Nucl. Fusion* **61** 106010

[44] Mathews A., Francisquez M., Hughes J.W., Hatch D.R., Zhu B. and Rogers B.N. 2021 Uncovering turbulent plasma dynamics via deep learning from partial observations *Phys. Rev.* E **104** 025205

[45] Institute of Plasma Physics Chinese Academy of Science 1,056 seconds, another world record for EAST

[46] D.E. Rumelhart, G.E. Hinton and R.J. Williams 1986 *Learning Internal Representations by Error Propagation* Rumelhart edn (Cambridge, MA: MIT Press) pp 318–62

[47] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł. and Polosukhin I. 2017 Attention is all you need *Advances in Neural Information Processing Systems (Nips)* vol 2017 pp 5999–6009

[48] Lao L.L., St H. John R.D.S., Kellman A.G. and Pfeiffer W. 1985 Reconstruction of current profile parameters and plasma shapes in tokamaks *Nucl. Fusion* **25** 1611–22

[49] Lao L.L., Ferron J.R., Groebner R.J., Howl W., St H. John E.J.S. and Taylor T.S. 1990 Equilibrium analysis of current profiles in tokamaks *Nucl. Fusion* **30** 1035–49

[50] Lao L.L., St H.E. John Q.P., Ferron J.R., Strait E.J., Taylor T.S., Meyer W.H., Zhang C. and You K.I. 2005 MHD equilibrium reconstruction in the DIII-D tokamak *Fusion Sci. Technol.* **48** 968–77

[51] Hofmann F. and Tonetti G. 1988 Tokamak equilibrium reconstruction using Faraday rotation measurements *Nucl. Fusion* **28** 1871–8

[52] Felici F., Sauter O., Coda S., Duval B.P., Goodman T.P., Moret J.-M. and Paley J.I. 2011 Real-time physics-model-based simulation of the current density profile in tokamak plasmas *Nucl. Fusion* **51** 083052

[53] Liu Z., Lin Y., Cao Y., Han H., Wei Y., Zhang Z., Lin S. and Guo B. 2021 SWIN transformer: hierarchical vision transformer using shifted windows

[54] Kolen J.F. Kremer S.C., Kolen J.F. and Kremer S.C. 2009 Gradient flow in recurrent nets: the difficulty of learning longterm dependencies *A Field Guide to Dynamical Recurrent Networks* (IEEE)

[55] Devlin J., Chang M.-W., Lee K., and Toutanova K. 2018 BERT: pre-training of deep bidirectional transformers for language understanding

[56] Huang Z., Zweig G., Levit M., Dumoulin B., Oguz B. and Chang S. 2013 Accelerating recurrent neural network training via two stage classes and parallelization *2013 IEEE Workshop on Automatic Speech Recognition and Understanding* (IEEE) pp 326–31

[57] Bergstra J., Bardenet R., Bengio Y. and Kégl B. 2011 Algorithms for hyper-parameter optimization *Advances in Neural Information Processing Systems* vol 24, ed J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K.Q. Weinberger (Curran Associates, Inc)

[58] Smith L.N. and Topin N. 2017 Super-convergence: very fast training of neural networks using large learning rates

[59] Wan B., Jiangang L.i., Guo H., Liang Y., Guosheng X., Wang L. and Gong X. 2015 Advances in H-mode physics for long-pulse operation on EAST *Nucl. Fusion* **55** 104015

[60] Wan B., Jiangang L.i., Guo H., Liang Y. and Guosheng X. 2013 Progress of long pulse and H-mode experiments in EAST *Nucl. Fusion* **53** 104006

[61] Jiangang L.i. and Wan B. 2011 Recent progress in RF heating and long-pulse experiments on EAST *Nucl. Fusion* **51** 094007

[62] Freidberg J.P. 2007 *Plasma Physics and Fusion Energy* vol 9780521851 (Cambridge: Cambridge University Press)