Comparison of machine learning systems trained to detect Alfvén eigenmodes using the CO$_2$ interferometer on DIII-D

Alvin V. Garcia[1,*] , **Azarakhsh Jalalvand**[2,*] , **Peter Steiner**[3] , **Andy Rothstein**[2] , **Michael Van Zeeland**[4] , **William W. Heidbrink**[1] and **Egemen Kolemen**[2,5,*]

[1] Department of Physics & Astronomy, University of California, Irvine, CA 92617, United States of America
[2] Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08540, United States of America
[3] Department of Speech Technology and Cognitive Systems, Dresden University of Technology, Dresden, Germany
[4] General Atomics, San Diego, CA 92121, United States of America
[5] Princeton Plasma Physics Laboratory, Princeton, NJ 08540, United States of America

E-mail: alvin.garcia@uci.edu, Azarakhsh.Jalalvand@princeton.edu and ekolemen@princeton.edu

CrossMark

Abstract

A Machine-Learning (ML) based detection scheme that automatically detects Alfvén Eigenmodes (AE) in a labelled DIII-D database is presented here. Controlling AEs is important for the success of planned burning plasma devices such as ITER, since resonant fast ions can drive AEs unstable and degrade the performance of the plasma or damage the first walls of the machine vessel. Artificial Intelligence could be useful for real-time detection and control of AEs in steady-state plasma scenarios by implementing ML-based models into control algorithms that drive actuators for mitigation of AE impacts. Thus, the objective is to compare differences in performance between using two different recurrent neural network systems (Reservoir Computing Network and Long Short Term Memory Network) and two different representations of the CO$_2$ phase data (simple and crosspower spectrograms). All CO$_2$ interferometer chords are used to train both models, but only one is processed during each training step. The results from the model and data comparison show higher performance for the RCN model (True Positive Rate = 90% and False Positive Rate = 14%), and that using simple magnitude spectrograms is sufficient to detect AEs. Also, the vertical CO$_2$ interferometer chord passing near the center is better for ML-based detection of AEs.

1. Introduction

The successful operation of planned nuclear fusion devices such as ITER depends on confined populations of super-thermal particles that heat fuel ions for a self-sustaining plasma burn [1]. If confined, alpha particles born from fusion reactions can provide the heating required to reach ignition. If these alphas become unconfined, they can carry away fusion power to the inner walls of the vessel and damage the first walls [2, 3]. The heat loss can be replaced using auxiliary heating mechanisms such as Neutral Beam Injection (NBI) or Radio Frequency (RF) waves and both of these methods can create populations of fast ions that are useful for momentum transfer and current drive [4]. Fast ions born from fusion reactions or external heating can resonate with special types of plasma waves called Alfvén Eigenmodes (AEs) [5–7], transfer energy to the wave, drive the plasma unstable and degrade energy confinement [8, 9]. Also, particle redistribution can expel fast ions from the plasma [8–14] and damage the inner walls of the vessel [15, 16]. Therefore, studying fast ions and controlling AEs is imperative for the realization of controlled nuclear fusion.

Real-time control of AEs in high performance burning plasmas without damage to the inner walls is a high priority for the Plasma Control System (PCS) at ITER [17, 18]. It is currently an important goal to determine the best set of external actuators in order to control AEs and alpha losses [19]. Suitable techniques include Electron Cyclotron Resonance Heating (ECRH) and current drive, and NBI. Since AEs can appear for short time scales on the order of milliseconds, simple feed-forward physics models are used to detect and control AEs. There is a need in the community for models with quick response times that could accurately detect the presence of AEs in real-time experiments.

Machine Learning (ML) applications in magnetic confinement fusion energy are growing and exciting opportunities exist in the fast-ion physics research field. Currently, the largest application of ML is in the area of disruption mitigation, where models are trained to prevent the rapid loss of thermal and magnetic energy during a quench of the plasma [20–27]. Surrogate model generation and experimental planning also benefit from data-driven methods [28]. On the other hand, ML in fast-ion research is a relatively new field. For example, Alfvénic and magnetohydrodynamic modes were detected using deep learning-based models, manually-labeled targets and magnetics on TJ-II [29] and COMPASS [30]. More examples used supervised learning to detect AEs [29–31], and data mining techniques combined with clustering for extraction of plasma fluctuations [32, 33].

In recent years, significant advancements have been made in detecting and controlling AEs using Electron-Cyclotron Emission (ECE) data on DIII-D. Originally, in-shot variation of neutral beam energy showed promise for AE control [34], then the first active real-time control of AEs in a tokamak utilized modulated beams to tune the drive for AEs using feedback from high resolution ECE signals [15]. Shortly after, the Large 2009-2017 DIII-D AE Energetic Particle Database [35] was created to better understand low frequency AEs and was later used for ML analysis in two papers [36, 37]. Deep Neural Networks were trained using ECE data in both studies. Reservoir Computing Networks (RCN) and Multi-layer Perceptron (MLP) Networks were trained in the former and latter study, respectively, and both achieved high performance. Section 2 of this paper discusses the Large AE-EP Database in more detail.

In this work, we focus on training Recurrent Neural Networks and labels created from the Large AE-EP Database, but use CO_2 interferometer data instead of ECE since there are several advantages: (1) calculating crosspower spectrograms between two chords is common in the fast-ion physics community since AE patterns can be highly visible in this representation of the data, (2) the 1D phase signals are routinely processed by the PCS for nearly every discharge and can be used for real-time control in future DIII-D experiments, and (3) although ECE measurements are high resolution and can measure AE fluctuations with good signal-to-noise, issues associated with resonances and cutoff frequencies pose challenges for AE detection. Using the CO_2 interferometer for AE identification is useful for reliably detecting AEs since it does not have limitations with cutoffs. For reasons 2 and 3, more shots are available in the Large AE-EP Database to train data-driven models. The baseline technique was initially trained to detect AEs using CO_2 interferometer data in a conference paper [38], and we report significant advancements here.

Building from our prior work, the primary objective of this paper is to study the performance by comparing the following: (1) different feature sets (simple magnitude and crosspower spectrograms), (2) recurrent neural networks (RCN vs LSTM), and (3) stacking outputs vs individual crosspower. The State-Of-The-Art (SOTA) technique in [36] used 40 stacked time-domain signals of ECE data and created labels from the Large AE-EP Database to train an RCN. We instead use spectrograms of CO_2 interferometer data, and individually forward pass each chord through both Neural Networks one-at-a-time and compare the results. The aim is to match our prior performance by training with CO_2 interferometer data for the potential long-term goal of creating an ML detector that could be useful in real-time AE control.

This paper is organized as follows: the CO_2 interferometer on DIII-D, labels from the Large AE-EP Database and important challenges are discussed in section 2. The results of model and feature comparison are shown in section 3.

Correlation analysis between predictions and metadata (equilibrium, beam, etc) are reported in section 4. Our conclusions appear in section 5.

2. Experimental data

DIII-D is a well diagnosed tokamak housing many diagnostic systems that measure the effects of AEs, with large amounts of available data from decades of experimental campaigns. Electron cyclotron emission [39], CO_2 interferometry [40], beam emission spectroscopy [41], and magnetic fluctuation diagnostic systems [42] can be used to study the effects of fast-ion driven instabilities. Diagnostic and plasma information can be relayed to actuators for real-time control of AEs in DIII-D experiments [19, 34, 43, 44].

The two-color vibration compensated CO_2 interferometer is a real-time system routinely used for feedback control of the plasma state at DIII-D. Additionally, it can provide useful information about the internal mode structure of AEs since it observes the AE induced density perturbations with a resolution in the $\Delta(nL)/nL \sim 10^{-5}$ range at the frequencies of interest. A layout of the CO_2 interferometer for an example equilibrium is displayed in figure 1. All four chords (three vertical and one horizontal) are digitized for 9 s per shot at a rate of $1.67\,\mathrm{MS\,s^{-1}}$, and the CO_2 phase data are studied in this work since AE frequencies are well above typical mechanical vibration frequencies. Also, the phase data are processed in real time by the PCS at DIII-D, making the AE detector in this work applicable for actuator driven mitigation of AE impacts.

In the past, identfication of AEs was usually done in a post-shot framework using crosspower spectrograms of CO_2 interferometer data and other AE fluctuation diagnostics (or plasma parameters). Doing spectral analysis is useful since generating spectrograms can remove low frequency noise and machine vibrations seen in the 1D signals. Although this method worked, it can be time consuming and requires extensive domain knowledge. In this work, we automate the identification process by training RCNs and LSTMs using simple and crosspower spectrograms of CO_2 interferometer data.

The original curated Large AE-EP Database was created to investigate the dependence of AE stability on plasma parameters in over 1139 shots [35]. It includes the occurrences of six plasma instabilities: Ellipticity (EAE), Toroidal (TAE), Reversed-Shear (RSAE), Beta-Induced (BAE), Low-Frequency Mode (LFM), and Energetic Particle-Induced Geodesic Acoustic Mode (EGAM) [35]. Table I of [36] shows a description of these modes. Times were selected when the various AEs were stable, marginal, or unstable. The number of time stamps per discharge was chosen to sample changes in plasma parameters and mode activity in a representative fashion. Time stamps usually appear in the middle of a type of activity, and many occur during the first 1.9 s since some AEs depend on the q profile and q steadily evolves during that phase of the discharge.

There are several challenges using the Large AE-EP Database and they are addressed here. The time stamps need to first be made binary and we adopt the one-hot encoding
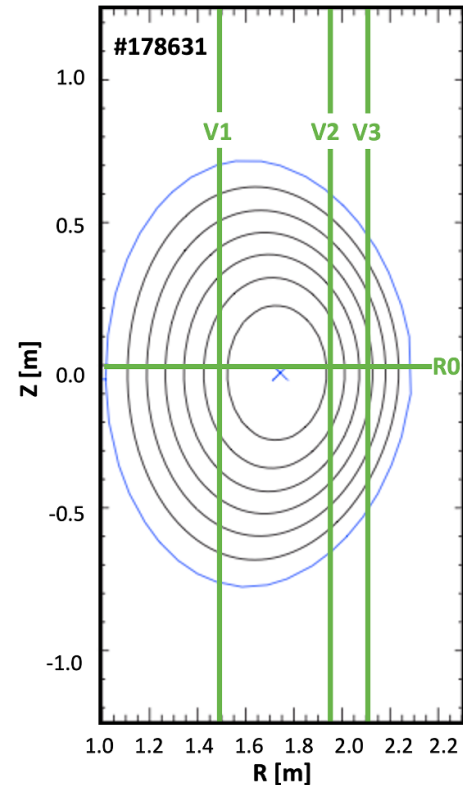

Figure 1. The elevation view of the CO_2 interferometer installed on DIII-D for shot #178631. Three vertical chords are located at R_m of 1.48 m, 1.94 m and 2.10 m, and the radial chord is horizontal on the midplane. The black curves are the magnetic flux surfaces (the last closed flux surface is in blue). The magnetic axis is denoted by the blue × symbol.
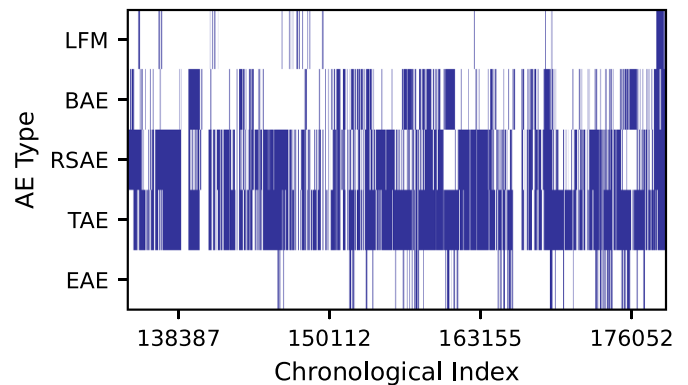
Figure 2. The CO_2 interferometer part of the AE-EP database begins and ends with DIII-D shot #132215 and #178880, respectively. The *presence* of AEs are plotted against these chronological shot numbers. In this figure, we show that TAE and RSAE are labelled frequently across many experimental campaigns. LFM, BAE and EAE have relatively sparse representation in the database.

method described in table II of [36]. We consider AEs originally marked unstable as being *present* in the discharge and mark them as 1, otherwise flags are reassigned to 0. Figure 2 shows the *presence* of AEs over the selected 1069 shots studied in this paper. Since predicting single time stamps is a challenge
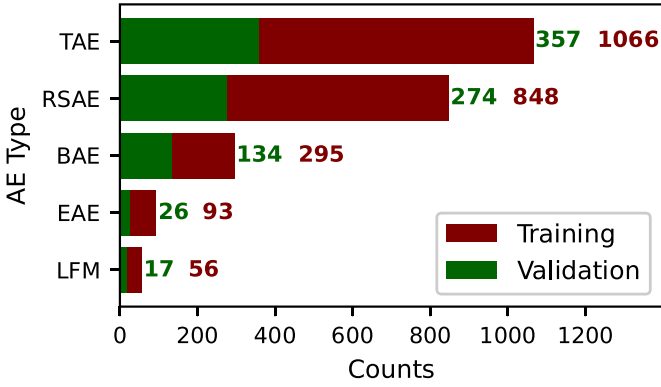
Figure 3. The occurrence of labels for the training set (801 shots) and validation set (268 shots) are skewed towards TAE and RSAE. The sets are randomly shuffled to preserve distribution shape. In comparision, there are barely any LFM or EAE instances throughout the database.

for ML-based methods, the re-assigned flag for all AEs are widened over ± 125 ms. This completes the creation of the labels used to train the RCN and LSTM in this work. The third challenge is shown in figure 3, where the distribution of labels is imbalanced and heavily skewed towards TAE and RSAE. This imbalance motivates using True Positive Rate (TPR) and False Positive Rate (FPR) as the metrics of success since the accuracy metric would be 94% if a model always predicted 0. TPR and FPR are defined as follows:

$$
\begin{aligned}
\mathrm{TPR} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \\
\mathrm{FPR} &= \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}
\end{aligned}
\tag{1}
$$

where TP = true positive, FN = false negative, FP = false positive, and TN = true negative. Although the ML classifiers train using information over the entire discharge and the original curated label is only available at discrete random times, TP and FP are modified such that a given prediction is reassigned only if an AE label is nearby within a window of ± 140 ms. Lastly, CO_2 interferometer, ECE and magnetics were all used to originally classify AEs in the Large AE-EP Database, which creates a classification challenge since a certain mode might show up more clearly in a different diagnostic than the CO_2 interferometer.

Given these challenges, our prior work using ECE data accomplished TPR = 91% and FPR = 7%, (table III of [36]). Our aim here is to match or improve these results using different feature sets of a new diagnostic system (CO_2 interferometer) and recurrent neural networks (RCN vs. LSTM).

3. Comparisons and results

In an effort to discover the best performing ML-based model for this new CO_2 interferometer project, several methods are explored in the following order: (1) linear regression, (2) MLPs, (3) convolutional neural networks (convnet), and

(4) recurrent neural networks. A brief, qualitative summary for the regression, MLP and convnet classification appear in section of the appendix. Here, three major goals are addressed:

1. Compare the features of different inputs, i.e. simple magnitude and advanced crosspower spectrograms. The extraction of these different feature sets is discussed in section 3.1
2. Determine the best performing recurrent neural network (RCN or LSTM) for this study. The different models are introduced in section 3.2.
3. Compare the performance of stacking outputs vs. crosspower combinations (2 simple spectrogram chords and 1 crosspower calculation). Results are shown in section 3.3.

3.1. Inputs

The inputs for both recurrent neural networks are simple magnitude and advanced crosspower spectrograms. These are windowed Fourier transform calculations using a window length of 4.9 ms and overlap of 80%. The spectrograms are downsampled using a maxpool function and the final input shapes are (time, frequency) = (140, 508). Maxpooling is commonly used in computer vision tasks and produced good results in this work. For the LSTM model, spectrograms are 'cut' into 280 ms windows, concatenated and fed into the model. For the RCN case, windowing is not implemented and the model processes 1D vectors of frequencies per training step. More details about the input preparation for the LSTM model can be viewed in section III of [38].

3.2. Architectures

The Python toolbox PyRCN (Python Reservoir Computing Networks) [45] is used for optimizing and training the RCN in this classification project. We utilize the more common RCN architecture, Echo State Networks (ESNs) [46], to perform the classification of AEs. Also, the hyper-parameter optimization routine is handled within the PyRCN framework and is based on the search strategy introduced in [47].

A two-layer RCN is developed by sequentially stacking two RCNs [48] on top of each other. It has been shown in [47–49] that stacking RCNs increases the temporal model capacity and reduces errors learned in early layers by rectifying their outputs in the subsequent layers. Larger capacity (more temporal information) can improve the RCN's performance in detecting specific AEs such as LFM. Figure 4 shows a diagram of this architecture. The RCN processes a timestamp vector of frequency values ($N_{\mathrm{freq}} \times 1$) and uses the provided labels ($N_{\mathrm{modes}} \times 1$) to train the readout layer. The output of the first RCN are scores for each of the five AEs. This output vector is then fed into the second RCN as input and the second readout layer is trained using the same labels. The final outputs are rectified scores for each AE.

The hyperparameter optimization strategy closely follows the method described in section 3c of [47]. Table 1 shows the results from the hyperparameter optimization routine for both
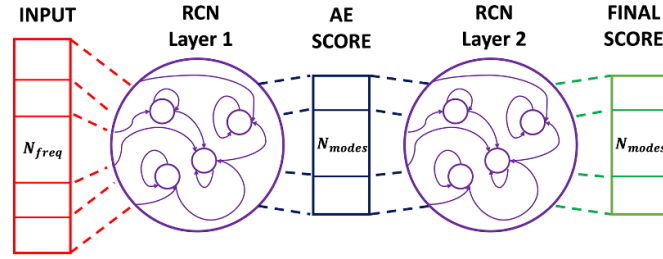
Figure 4. Schematic of the stacked two-layer RCN used to classify AEs trained with simple and crosspower spectrograms. The input layer of the first RCN is connected to a reservoir of nonlinear neurons and gets mapped to a higher dimensional space, where the data are more separable. The readout layer of the first RCN is trained using linear regression and processed as inputs for the second RCN. The second reservoir consists of less neurons since less model capacity is needed to rectify the mistakes of the first layer. The final outputs are AE scores.

Table 1. The results from the hyperparameter optimization routines used to train the RCN network. A sequential search hyperparameter optimization strategy is used to train the readout layers of the stacked two-layer RCN. Final values for each hyperparameter and each layer are reported in the final two columns.

Hyper- parameter	Range	Distribution	Layer 1	Layer 2
Input scaling	$1 \times 10^{-5} - 1$	uniform	0.03	0.09
Spectral radius	0 to 2	uniform	1.27	0.39
Leakage	$1 \times 10^{-5} - 1$	log uniform	0.71	0.71
Bias scaling	$1 \times 10^{-2} - 1$	uniform	0.27	0.13
Alpha	$1 \times 10^{-5} - 10$	loguniform	6.60	0.00

Table 2. Similar to table 1, only for the LSTM network. A simple sequential scan is implemented here. Final values are listed in the last column.

Hyperparameter	Range	Step size	Final
Optimizer	Adam, Adamax, RMSprop	N/A	Adam
Learning rate	$10^{-2} - 10^{-6}$	$\times 10^{-1}$	10^{-4}
Layers (LSTM)	$2 - 9$	$+1$	3
Nodes (LSTM)	$32 - 256$	$\times 2$	64
Layers (MLP)	$1 - 3$	$+1$	3
Nodes (MLP)	$64 - 512$	$\times 2$	128
Dropout layers	$1 - 3$	$+1$	1
Dropout threshold	$0.25 - 0.75$	$+0.25$	0.5

layers. The process uses a three-step sequence of searches for the hyperparameters input scaling, spectral radius, bias scaling and leakage. The steps of the method are as follows:

1. Perform a random search for the input scaling and spectral radius while the bias scaling and leakage terms are held constant.
2. Fix the leakage to 1 and search for the bias scaling.
3. Search for the leakage term.

Step 1 determines the balance between forward and recurrent connections, step 2 can introduce more non-linearity into the system, and step 3 determines the attention the network gives to temporal information in the inputs. The hyperparameters at each step are selected based on the minimization of the Mean Squared Error (MSE) curve.

The LSTM model consists of three layers using 64 Long-Short Term Memory cells, one dropout layer with dropout probability of 0.5, and four layers using MLPs with 128 nodes per layer. All hidden layers utilize relu activation functions, and the weights are initialized using uniform variance scaling [35]. The prediction layer consists of 5 nodes with sigmoid activations and outputs scores for each AE. Binary Crossentropy is used evaluate the loss score, and the Adam optimizer with a learning rate of 0.0001 will tune the weights. The hyperparameters for this network are shown in table 2, and they are optimized by sequentially scanning values and analyzing predictions over three selected discharges.

3.3. Results

A data-driven convention is initially implemented to evaluate the performance of the model, and detailed analysis of the predicted errors follow. Thus, the TP and FP metrics are modified such that the time slice of each prediction is reassigned if any AE label information is available within a window of ± 140 ms of the predicted timeslice. Two examples at the end of this section show a few observed errors. Also, comparisons here evaluate performance over all four CO_2 interferometer chords.

Table 3. Comparison of results using simple and crosspower spectrograms for the RCN and LSTM models. RCN predictions are made binary (0 and 1) using AE scores = 0.05, 0.15, 0.11, 0.07 and 0.08 for the five AEs listed in the left column. Similarly, the LSTM threshold values are 0.06, 0.13, 0.13, 0.10 and 0.07. The RCN trained using simple spectrograms is the top performer.

| | SIMPLE | | | | CROSSPOWER | | | |
| | RCN | | LSTM | | RCN | | LSTM | |
AE	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
EAE	0.60	0.07	0.77	0.07	0.72	0.07	0.70	0.06
TAE	0.93	0.18	0.94	0.26	0.89	0.14	0.94	0.28
RSAE	0.94	0.19	0.91	0.29	0.89	0.15	0.92	0.28
BAE	0.80	0.23	0.79	0.23	0.69	0.13	0.79	0.27
LFM	0.81	0.05	0.80	0.10	0.64	0.02	0.78	0.07
TOTAL	0.90	0.14	0.90	0.18	0.85	0.10	0.90	0.18
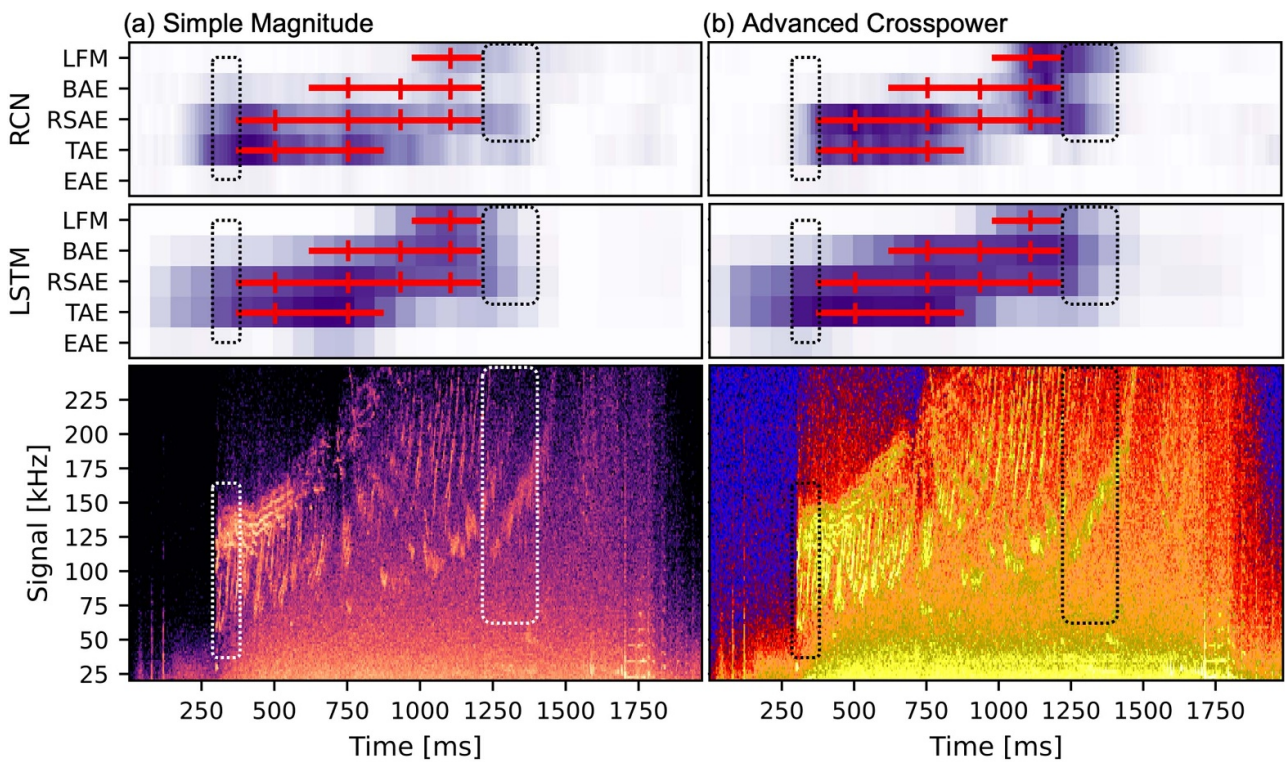
Figure 5. A comparison of the raw RCN and LSTM predictions using simple magnitude (panel (*a*)) and advanced crosspower (panel (*b*)) spectrograms for shot #178631. The simple spectrogram is calculated for chord V2 and the crosspower is between chords V2 and R0. The red vertical ticks and horizontal strikethroughs indicate the curated time stamp and label, respectively. The purple pixels are raw predictions for the RCN and LSTM models. Regions where the purple pixels overlap the red strikethroughs are considered good agreement. The dotted regions are times where the curated database does not indicate anything, yet the model is robust enough to capture the AE activity observed in the spectrograms.

The classification results that compare the RCN and LSTM model for simple and crosspower spectrograms are summarized by table 3. The results for the simple spectrograms are nearly equal or better than crosspower spectrograms for the RCN and LSTM model. Also, the RCN performs slightly better than the LSTM model when using simple spectrograms. The LSTM can trigger slightly stronger predictions than the RCN. This is visible in the slightly higher TPR for EAE and FPR for TAE and RSAE. Figure 5 is a specific example with a lot of AE activity that demonstrates
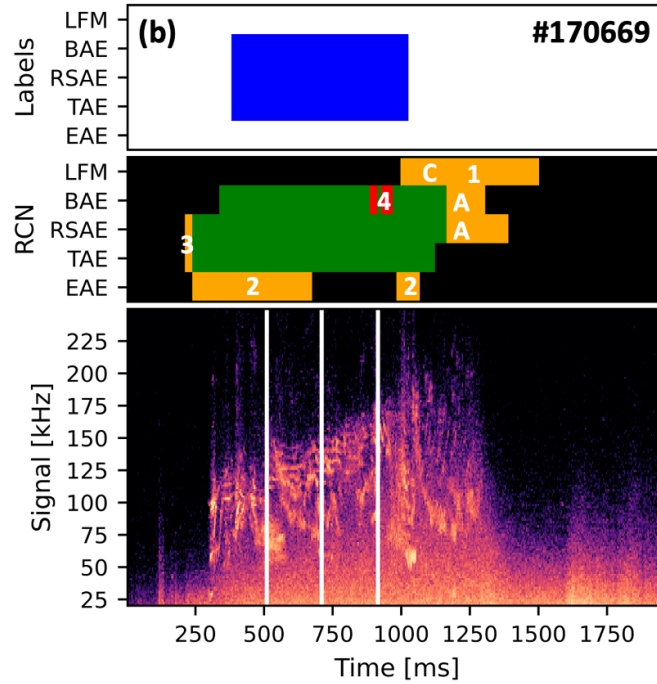
Figure 6. AE labels, thresholded predictions and simple magnitude spectrograms for shot #170669. The colored predictions are denoted as follows: TP = green, FP = orange, FN = red, and TP = black. White vertical lines in the spectrograms indicate the original timestamp. Error type 1 is due to effects from overfitting, since the model could be triggering scores for LFM due to the overall pattern of the discharge. Error type 2 occurs due to noise in the spectrograms. Error type 3 is attributed to time delays for predictions. Error type 4 is categorized as a general AI error, where the model failed to predict correctly. Letter A indicates an incorrectly assigned error since there is still activity but the Δt extension of the label is too short. Letter C also indicates an incorrectly assigned error due to ambiguity in the discharge.

the feature set comparison using the RCN and LSTM models.

Both models might be slightly overestimating detection of TAE and RSAE since the FPR is relatively higher for these AEs. Since the RCN triggers slightly lower, the overestimation effect is smaller. In regions where the AI failed, this is likely due to several reasons: (1) models are overfitting to training data, (2) noise in the CO_2 interferometer spectrograms, 3) latency associated with sparse time stamp, (4) general AI error. However, there are cases where the AI is working well, but an error is assigned. Possible reasons for this are the following: Incorrect value assigned to curated database through (A) Δt label extension and (B) calling no label stable, and C) some cases can be ambiguous. Figure 6 illustrates some of these points. Despite these issues, both models are capable of learning the patterns associated with AEs in this database and achieve high performance.

The RCN model demonstrates better results, and additional advantages include finer resolved predictions and the training speed for an RCN can be faster than for an LSTM. The RCN substantially improved the linear baseline technique using only linear regression since the memory of the model is higher with the addition of reservoirs containing random recurrent connections. Figure 7 shows the effect of adding a second layer to the RCN model. This effect is similarly observed when adding a third layer and increasing the number of nodes to 64 for the LSTM model.

In an effort to determine the set of chords with the highest AE detection, we check the performance for one chord,
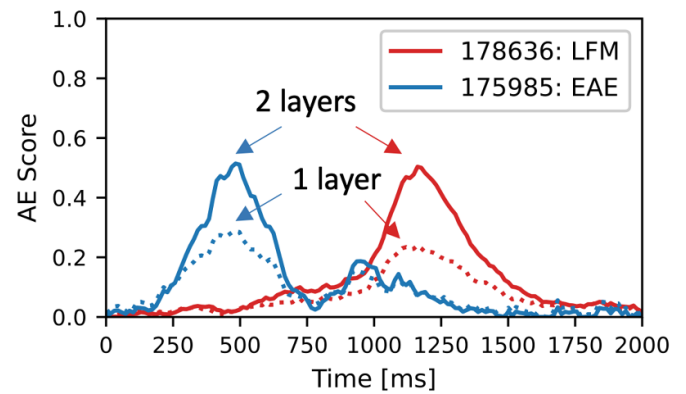

Figure 7. LFM and EAE predictions using the RCN model for shots 178 636 and 175 985 for CO_2 chord V2. A second reservoir rectifies the mistakes made by the first layer and produces better predictions for the least common modes in the database.

two-stacked chords or one crosspower combination using the F_2 score. This metric is a harmonic mean of the recall (TPR) and precision ($\frac{TP}{TP+FP}$) metrics, where $\beta = 2$ in the following equation:

$$F_\beta = \frac{1 + \beta^2}{\frac{\beta^2}{Recall} + \frac{1}{Precision}}. \tag{2}$$

The metrics TP and FP are further modified here by an additional $\Delta t = \pm 71$ ms in the calculation of the F_2 score to capture more information per chord from the discharges. These
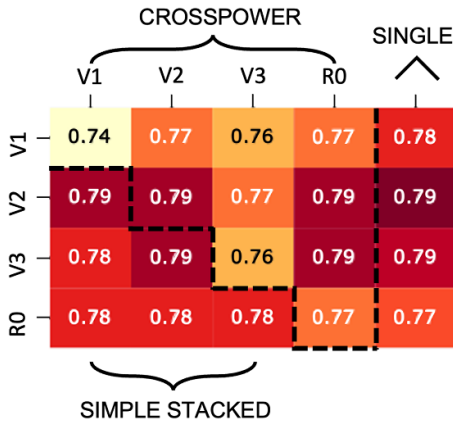
Figure 8. F2 scores for the crosspower (upper diagonal), stacked chords (lower diagonal) and single chord (right vertical bar) comparison using the RCN model. Stacking chords can perform better than crosspower, and chord V2 performs slightly better than the other three chords.

values are collected into a confusion matrix shown in figure 8. For the upper diagonal (crosspower), the best performing combinations are V2R0, V3R0 and V2V2 (autopower). The anecdotal favorite combination in the control room during experiments at DIII-D is V2R0 and the RCN model scores highest for this combination. For the lower diagonal (stacking outputs), adding predictions from V2 to V1 and any chord to R0 slightly improves the performance of V1 and R0, respectively. Although these differences are small, additional AE information from different chords might be needed when predicting using chords V1 or R0. Lastly, the darkest shaded region indicates that predictions for chord V2 achieve the best performance.

4. Analysis of metadata

Additional information in the Large AE-EP Database can be used to study model interpretability, and correlations between misclassification and operating regime parameters. The following inferred and experimental data are available [35]: (1) EFIT equilibrium reconstructions [50] provide plasma shape, magnetic field, and beta information, (2) kinetic temperature, plasma rotation, and electron and impurity densities from between-shot profile fitting algorithms, and 3) information about neutral beams such as injected power, energy, voltage and orientation. The goal is to determine if there are any tendencies with misclassification by calculating Pearson correlation coefficients, r, between TPR and FPR with all 68 parameters in the database for the validation set.

Although many parameters have coefficient values near zero for both TPR and FPR, we report parameters with the highest values here. For the AE labels, BAE has the strongest Pearson correlation coefficient with values of -0.22 and -0.21 for TPR and FPR, respectively. For the plasma parameters, the strongest correlation for TPR is with
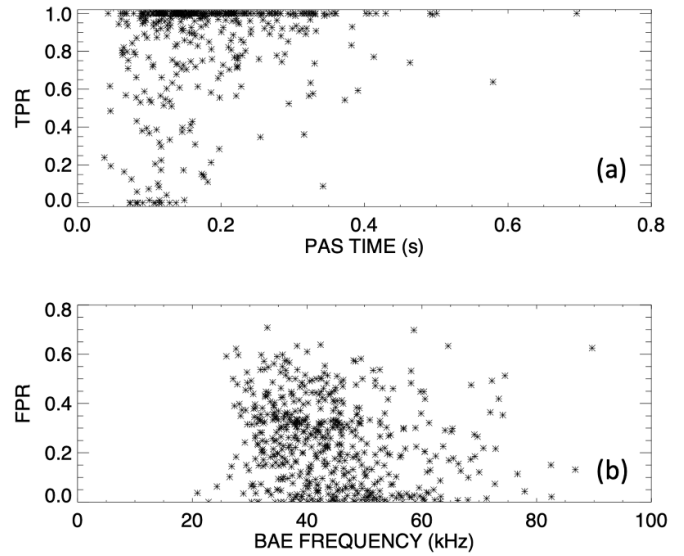
Figure 9. Points for the strongest Pearson correlation coefficient, r, in the comparison between AE metrics (TPR & FPR) and metadata are shown here. In panel a, the pitch-angle scattering (PAS) time is the 90-degree scattering time in the NRL Formulary [51]. The r between PAS and TPR is 0.20. In panel b, the BAE frequency is from equation (1) of [52], and the r with FPR is -0.17. Since most of the analysis shows low correlation values, concerns regarding the RCN model failing to predict AEs at the limits of the parameter range are alleviated.

Pitch-Angle-Scattering (PAS) time on axis, and for FPR is with an analytical calculation of the BAE frequency; see figure 9. The r for PAS with TPR and BAE frequency with FPR are 0.20 and -0.17, respectively. In both cases, there are about 500 points used in the comparison, and an $|r| \simeq 0.20$ indicates that the dependence is either non-existent or very weak. Thus, there is no evidence of any dependence on the operating regime-suggesting that we could safely use the identifier throughout this parameter range and likely somewhat beyond.

5. Conclusion

Recurrent neural networks are trained using CO_2 interferometer data and labels from the Large AE-EP Database on the DIII-D tokamak. Two models (RCN and LSTM) are trained separately using simple and crosspower spectrograms. The additional steps required to calculate crosspower are unnecessary since the predictions are similar for both types of inputs. Both models are trained using one CO_2 chord per training step and achieve high results. The RCN performance is slightly higher with TPR = 90% and FPR = 14%. Detection using any single chord is feasible (V2 is slightly better than the other three). Since the model is primarily trained using labels marked during the current ramp phase, more cases labelling the steady-state portion of the discharge would improve generalizability. Lastly, analysis of the metadata demonstrates that the RCN model still works at the limits of the experimental parameter ranges.

The CO_2 Interferometer is commonly used in fluctuation analysis, acquires data for nearly every DIII-D experiment, is available in the PCS in real-time, and does not have issues with cutoff frequencies. Given these results and advantages, it is strongly recommended to detect AEs using RCNs trained with simple magnitude spectrograms calculated using the vertical chord passing near center ($R_m = 1.94m$ at DIII-D).

Future work would consist of implementing the RCN reported in this paper into real-time control algorithms to detect AEs at DIII-D. The SOTA detector currently installed on the PCS is an RCN that trained using ECE data with 8000 and 500 nodes for layers 1 and 2, respectively. It processes time domain signals and makes predictions in approximately 400 microseconds for each time step. The RCN developed in this work is smaller for both layers (4000 and 50 nodes). Although there is an additional step of calculating spectrograms, the RCN trained using CO_2 Interferometer data could have a similar or faster response time during real-time experiments. Implementation of Fast Fourier Transforms into the PCS is currently under consideration and we plan to test it in the near future.

Acknowledgments

Disclaimer

Appendix. Linear regression, MLP and Convnets

The crosspower spectrograms were initially used to train Linear Regression, Multi-Layer Perceptron (MLP) and Convolutional Neural Networks. For regression, Tikhonov regularization is used as follows:

$$\mathbf{W}_{\text{out}} = \left(\mathbf{R}^T\mathbf{R} + \alpha\mathbf{I}\right)^{-1}\mathbf{R}^T\mathbf{D}, \tag{A1}$$

where $\mathbf{R}$ is the data, $\mathbf{D}$ are the labels and α is the regularization parameter. For the MLP network, the models are nearly the same as the LSTM model described in section 3.2, only without the LSTM block. Lastly, the convnet contained a sequence of 5 convolutional and max pooling layers followed by a small MLP block for the final classification.

All models were capable of detecting the most common modes (TAE and RSAE) and struggled to detect the other three modes (BAE, EAE and LFM). This is likely due to the lack of memory in the models. However, the convnet was capable of performing slightly better when trained over the entire discharge (similar to the popular cat-dog classification problem) instead of the windowing technique described in section 3.2. Although temporal information is lost when training over the entire discharge, the convnet is capable of predicting the *presence* of these modes within approximately 10% of the results reported in this paper.

ORCID iDs

Alvin V. Garcia ⬤ https://orcid.org/0000-0001-8647-2727
Azarakhsh Jalalvand ⬤ https://orcid.org/0000-0001-8739-1793
Peter Steiner ⬤ https://orcid.org/0000-0002-8149-2275
Andy Rothstein ⬤ https://orcid.org/0000-0003-2183-0997
Michael Van Zeeland ⬤ https://orcid.org/0000-0002-7911-2739
William W. Heidbrink ⬤ https://orcid.org/0000-0002-6942-8043
Egemen Kolemen ⬤ https://orcid.org/0000-0003-4212-3247

References

[1] Pinches S.D. *et al* (JET-EFDA Contributors) 2004 The role of energetic particles in fusion plasmas *Plasma Phys. Control. Fusion* **46** B187–200

[2] Duong H., Heidbrink W., Strait E., Petrie T., Lee R., Moyer R. and Watkins J. 1993 Loss of energetic beam ions during TAE instabilities *Nucl. Fusion* **33** 749

[3] White R., Fredrickson E., Darrow D., Zarnstorff M., Wilson R., Zweben S., Hill K., Chen Y. and Fu G. 1995 Toroidal Alfvén eigenmode-induced ripple trapping *Phys. Plasmas* **2** 2871–3

[4] Gorelenkov N.N., Pinches S. and Toi K. 2014 Energetic particle physics in fusion research in preparation for burning plasma experiments *Nucl. Fusion* **54** 125001

[5] Heidbrink W.W. 2008 Basic physics of Alfvén instabilities driven by energetic particles in toroidally confined plasmas *Phys. Plasmas* **15** 055501

[6] Chen L. and Zonca F. 2016 Physics of Alfvén waves and energetic particles in burning plasmas *Rev. Mod. Phys.* **88** 015008

[7] Todo Y 2019 Introduction to the interaction between energetic particles and Alfvén eigenmodes in toroidal plasmas *Rev. Mod. Plasma Phys.* **3** 1–33

[8] Heidbrink W.W. *et al* 2014 Confinement degradation by Alfvén-eigenmode induced fast-ion transport in steady-state scenario discharges *Plasma Phys. Control. Fusion* **56** 095030

[9] Holcomb C.T., Heidbrink W.W., Ferron J.R., Van Garofalo M.A., Solomon W.M., Garofalo A., Gong X., Mueller D., Grierson B. and Collins C. 2015 Fast-ion transport in $q_{m}in > 2$, high-β steady-state scenarios on DIII-D *Phys. Plasmas* **22** 055904

[10] Berk H., Breizman B. and Ye H. 1992 Scenarios for the nonlinear evolution of alpha-particle-induced Alfvén wave instability *Phys. Rev. Lett.* **68** 3563

[11] White R., Gorelenkov N., Heidbrink W. and Van Zeeland M. 2010 Particle distribution modification by low amplitude modes *Plasma Phys. Control. Fusion* **52** 045012

[12] García-Muñoz M. *et al* (the ASDEX Upgrade Team) 2010 Convective and diffusive energetic particle losses induced by Shear Alfvén waves in the ASDEX Upgrade tokamak *Phys. Rev. Lett.* **104** 185002

[13] Sigmar D., Hsu C., White R. and Cheng C. 1992 Alpha-particle losses from toroidicity-induced Alfvén eigenmodes. Part II: Monte Carlo simulations and anomalous alpha-loss processes *Phys. Fluids* B **4** 1506–16

[14] Appel L., Berk H., Borba D., Breizman B., Hender T., Huysmans G., Kerner W., Pekker M., Pinches S. and Sharapov S. 1995 Alfvén eigenmode induced energetic particle transport in JET *Nucl. Fusion* **35** 1697

[15] Hu W. *et al* 2018 Active real-time control of Alfvén eigenmodes by neutral beam and electron cyclotron heating in the DIII-D tokamak *Nucl. Fusion* **58** 124001

[16] Kurki-Suonio T. *et al* 2009 ASCOT simulations of fast ion power loads to the plasma-facing components in ITER *Nucl. Fusion* **49** 095001

[17] Snipes J., Campbell D., Casper T., Gribov Y., Loarte A., Sugihara M., Winter A. and Zabeo L. 2011 MHD and plasma control in ITER *Fusion Sci. Technol.* **59** 427–39

[18] Snipes J. *et al* 2014 Physics of the conceptual design of the ITER plasma control system *Fusion Eng. Des.* **89** 507–11

[19] Garcia-Munoz M. *et al* 2019 Active control of Alfvén eigenmodes in magnetically confined toroidal plasmas *Plasma Phys. Control. Fusion* **61** 054007

[20] Rea C. and Granetz R.S. 2018 Exploratory machine learning studies for disruption prediction using large databases on DIII-D *Fusion Sci. Technol.* **74** 89–100

[21] Fu Y., Eldon D., Erickson K., Kleijwegt K., Lupin-Jimenez L., Boyer M.D., Eidietis N., Barbour N., Izacard O. and Kolemen E. 2020 Machine learning control for disruption and tearing mode avoidance *Phys. Plasmas* **27** 022501

[22] Cannas B., de Vries P.C., Fanni A., Murari A., Pau A. and Sias G. 2015 Automatic disruption classification in JET with the ITER-like wall *Plasma Phys. Control. Fusion* **57** 125003

[23] Rea C., Granetz R.S., Montes K., Tinguely R.A., Eidietis N., Hanson J.M. and Sammuli B. 2018 Disruption prediction investigations using machine learning tools on DIII-D and Alcator C-Mod *Plasma Phys. Control. Fusion* **60** 084004

[24] Murari A., Lungaroni M., Peluso E., Gaudio P., Vega J., Dormido-Canto S., Baruzzo M. and Gelfusa M. (JET Contributors) 2018 Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET *Nucl. Fusion* **58** 056002

[25] Kates-Harbeck J. 2019 Svyatkovskiy A and Tang W 2019 Predicting disruptive instabilities in controlled fusion plasmas through deep learning *Nature* **568** 526–31

[26] Aymerich E., Sias G., Pisano F., Cannas B., Carcangiu S., Sozzi C., Stuart C., Carvalho P. and Fanni A. (JET Contributors) 2022 Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles *Nucl. Fusion* **62** 066005

[27] Montes K.J. *et al* 2019 Machine learning for disruption warnings on Alcator C-Mod, DIII-D and EAST *Nucl. Fusion* **59** 096015

[28] Humphreys D. *et al* 2020 Advancing fusion with machine learning research needs workshop report *J. Fusion Energy* **39** 123–55

[29] Bustos A., Ascasíbar E., Cappa A. and Mayo-García R. 2021 Automatic identification of MHD modes in magnetic fluctuation spectrograms using deep learning techniques *Plasma Phys. Control. Fusion* **63** 095001

[30] Škvára V., Šmídl V., Pevný T., Seidl J., Havránek A. and Tskhakaya D. 2020 Detection of Alfvén eigenmodes on COMPASS with generative neural networks *Fusion Sci. Technol.* **76** 962–71

[31] Woods B.J.Q., Duarte V.N., Fredrickson E.D., Gorelenkov N.N., Podesta M. and Vann R.G.L. 2020 Machine learning characterization of alfvénic and sub-alfvénic chirping and correlation with fast-ion loss at NSTX *IEEE Trans. Plasma Sci.* **48** 71–81

[32] Haskey S., Blackwell B. and Pretty D. 2014 Clustering of periodic multichannel timeseries data with application to plasma fluctuations *Comput. Phys. Commun.* **185** 1669–80

[33] Haskey S., Blackwell B., Nührenberg C., Könies A., Bertram J., Michael C., Hole M. and Howard J. 2015 Experiment-theory comparison for low frequency BAE modes in the strongly shaped H-1NF stellarator *Plasma Phys. Control. Fusion* **57** 095011

[34] Pace D.C. *et al* 2016 Control of power, torque and instability drive using in-shot variable neutral beam energy in tokamaks *Nucl. Fusion* **57** 014001

[35] Heidbrink W.W., Van Zeeland M.A., Austin M.E., Bierwage A., Chen L., Choi G., Lauber P., Lin Z., McKee G.R. and Spong D.A. 2020 BAAE' instabilities observed without fast ion drive *Nucl. Fusion* **61** 016029

[36] Jalalvand A., Kaptanoglu A.A., Garcia A.V., Nelson A.O., Abbate J., Austin M.E., Verdoolaege G., Brunton S.L., Heidbrink W.W. and Kolemen E. 2022 Alfvén eigenmode classification based on ECE diagnostics at DIII-d using deep recurrent neural networks *Nucl. Fusion* **62** 026007

[37] Kaptanoglu A.A., Jalalvand A., Garcia A., Austin M.E., Verdoolaege G., Schneider J., Hansen C.J., Brunton S., Heidbrink W.W. and Kolemen E. 2022 Exploring data-driven models for spatiotemporally local classification of Alfvén eigenmodes *Nucl. Fusion* **62** 106014

[38] Garcia A.V., Jalalvand A., Steiner P., Rothstein A., Van Zeeland M., Heidbrink W.W. and Kolemen E. 2023 Alfvén eigenmode detection using long-short term memory networks and CO_2 interferometer data on the DIII-D national fusion facility *2023 Int. Joint Conf. on Neural Networks (IJCNN)* (*Gold Coast, Australia*, *June 18*) (IEEE) pp 1–8 (available at: https://ieeexplore.ieee.org/document/10191760)

[39] Austin M. and Lohr J. 2003 Electron cyclotron emission radiometer upgrade on the DIII-D tokamak *Rev. Sci. Instrum.* **74** 1457–9

[40] Van Zeeland M., Kramer G., Nazikian R., Berk H., Carlstrom T. and Solomon W. 2005 Alfvén eigenmode

observations on DIII-D via two-colour CO2 interferometry *Plasma Phys. Control. Fusion* **47** L31

[41] McKee G., Ashley R., Durst R., Fonck R., Jakubowski M., Tritz K., Burrell K., Greenfield C. and Robinson J. 1999 The beam emission spectroscopy diagnostic on the DIII-D tokamak *Rev. Sci. Instrum.* **70** 913–6

[42] King J.D. *et al* 2014 An upgrade of the magnetic diagnostic system of the DIII-D tokamak for non-axisymmetric measurements *Rev. Sci. Instrum.* **85** 083503

[43] Li P. *et al* 2021 Dynamics between toroidal Alfvén eigenmode evolution and turbulence suppression under RMP on EAST *Nucl. Fusion* **61** 086020

[44] Tang S., Carter T., Crocker N., Heidbrink W., Lestz J., Pinsker R., Thome K., Van Zeeland M. and Belova E. 2021 Stabilization of Alfvén eigenmodes in DIII-D via controlled energetic ion density ramp and validation of theory and simulations *Phys. Rev. Lett.* **126** 155001

[45] Steiner P., Jalalvand A., Stone S. and Birkholz P. 2022 PyRCN: a toolbox for exploration and application of reservoir computing networks *Eng. Appl. Artif. Intell.* **113** 104964

[46] Jaeger H. 2001 The "echo state" approach to analysing and training recurrent neural networks *Technical Report GMD Report* 148 (German National Research Center for Information Technology)

[47] Steiner P., Stone S., Birkholz P. and Jalalvand A. 2021 Multipitch tracking in music signals using echo state networks *2020 28th European Signal Processing Conf. (EUSIPCO)* (*Virtual, 18–21 January 2021*) (IEEE) pp 126–30 (available at: https://ieeexplore.ieee.org/abstract/document/9287638)

[48] Triefenbach F., Jalalvand A., Schrauwen B. and Martens J.P. 2010 Phoneme recognition with large hierarchical reservoirs *Advances in Neural Information Processing Systems* (*Vancouver, Canada, 6–11 December 2010*) vol 23 (available at: https://proceedings.neurips.cc/paper/2010/hash/2ca65f58e35d9ad45bf7f3ae5cfd08f1-Abstract.html)

[49] Jalalvand A., Vandersmissen B., De Neve W. and Mannens E. 2019 Radar signal processing for human identification by means of reservoir computing networks *IEEE Radar Conf. (RadarConf)* (*Boston, Massachusetts, USA, 22–26 April 2019*) *IEEE Radar Conf. (RadarConf)* pp 1–6 (available at: https://ieeexplore.ieee.org/abstract/document/8835753)

[50] Lao L., John H.S., Stambaugh R., Kellman A. and Pfeiffer W. 1985 Reconstruction of current profile parameters and plasma shapes in tokamaks *Nucl. Fusion* **25** 1611

[51] Richardson A.S. 2019 2019 NRL plasma formulary *Technical Report* (US Naval Research Laboratory)

[52] Heidbrink W. *et al* 2021 Isotope dependence of beta-induced alfvén eigenmode (BAE) and low frequency mode (LFM) stability in DIII-D *Nucl. Fusion* **61** 106021