Big Data Machine Learning for Disruption Predictions

William Tang[1], Matthew Parsons[1], Eliot Feibush[1], Jong Choi[2], Tahsin Kurc,[3] **(USA)**
(1) Princeton Plasma Physics Laboratory; (2) Oak Ridge National Laboratory; (3) Stony Brook University
A. Murari[4], J. Vega[4], A. Pereira[4], I. Nunes (IPFN, Portugal), and JET contributors[4]
[4]**(EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK)**

Email of Main Author: tang@pppl.gov

Building the scientific foundations needed to develop fusion power in a timely way can be facilitated not only by familiar “hypothesis-driven”/ first principles approaches but also by engaging modern big-data-driven statistical methods featuring machine learning (ML). An especially time-urgent and very challenging problem facing the development of a fusion energy reactor today is the need to reliably mitigate and avoid large-scale major disruptions in magnetically-confined tokamak systems such as the Joint European Torus (JET) -- today and the burning plasma ITER device in the near future. These major macroscopic events lead to rapid termination of plasma discharges including severe impulsive heat loads damaging material components. Avoiding or at least mitigating them is critical because ITER can sustain at most a very small number of full current disruptions. Since they can damage the surfaces of the machine, which in turn can cost hundreds of millions of dollars to remediate, it is critical that the international fusion mission engage in multiple avenues to accelerate progress toward achieving the capability to reliably avoid such events with better than 95% predictive capability [1].

In the current paper we will present results from the development and testing of ML-based-methodologies – an exciting R&D approach that is increasingly deployed in many scientific and industrial domains -- to help provide much-needed guidance for disruption avoidance in JET. Working on this repository of the most important and largest (nearly a half petabyte and growing) data base of fusion-grade plasmas, JET statistical scientists have successfully deployed ML software interfaced with the large JET data base over the course of the past 6 years [2,3]. This has produced encouraging results involving primarily the application of the support vector machine (SVM) approach. The goals for the present investigations are to: (i) achieve greater predictive reliability by improving the physics fidelity of the classifiers within the “supervised” ML workflow; and (ii) establishing the portability of the associated software beyond JET to other current tokamak systems and to ITER in the future. In order to so, it will be necessary to address the more realistic multi-dimensional, time-dependent, and much larger complex data instead of the simpler zero-dimensional, temporal data considered at present in all of the JET ML studies. Associated challenges in delivering higher physics-fidelity classifiers needed to enable establishing portability of the predictive software when applied to MFE systems different than JET will be presented. In addition, it is expected that deployment of such improved ML software will need to be accordingly upgraded from current modern clusters to much more powerful leadership class supercomputers. This presentation will include: (i) a description of a new workflow developed for our “supervised” ML SVM approach; (ii) results from associated applications – including profile information -- to the increasingly larger JET disruption-relevant data base; and (iii) highlights of current progress as well as key obstacles for this major big-data machine-learning problem. Scientific/technical progress achieved that will be discussed include:

• Systematic exploration (via MDS+ tree) of the JET disruption data base of associated signal and video data – enabled by formal approval of the EUROfusion JET leadership;

• Rewrite of SVM cross-validation routines now self-contained within Matlab, eliminating excessive file I/O and improving performance time by 100x;

• Description of PPPL’s SVM software “Machine Learning Disruption Classification Developer (ML-DCD)” that is interfaced with the JET data base, including successful benchmarking vs. results obtained using JET’s “Advanced Predictor of Disruptions (APODIS)” for zero-D time traces of both Carbon-wall and ITER-like-wall (ILW) cases.

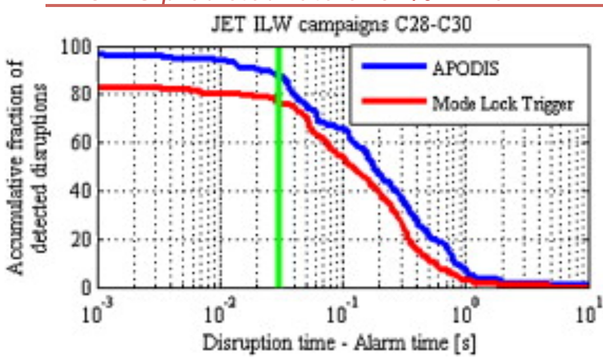
• Description of collaborative studies with the CS/Applied Math experts at ORNL and Stony Brook University in exploring alternative methods for improving the selection of classifiers indicating: (i) clustering based on deterministic annealing is a promising approach in that it does not require a pre-determined number of clusters; and (ii) systematic examination of multi-dimensional image data (e.g., from ECEI measurements) can provide additional information that improves prediction results since it contains spatial information that should be exploited.

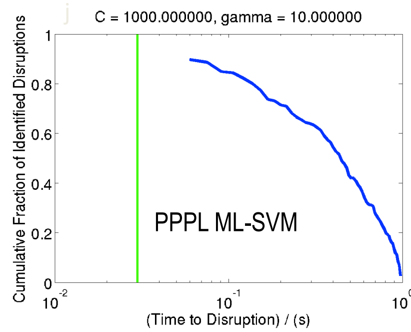
• Discussion of new results from initial ML studies that include electron temperature profile information indicating that since latent patterns emerge when additional signals (different from the set used to date), significant improvements to the algorithms can be achieved.

RECENT RESULTS for JET ILW Disruption Data: Comparison of Results from PPPL's "ML-DCD" Analysis with JET's "APODIS"

30 ms before disruption *(Ref.-- P. DeVries; M. Lehnen – ITER disruption prediction requirements → mitigation trigger time > 30 ms)*

APODIS predicted rate of 87% while PPPL ML-SVM gives 89%





• JET's "APODIS" (Advanced Predictor of Disruptions) trained on 738 disruptive and 2,035,000 non-disruptive samples

• PPPL's "ML-DCD" (Machine Learning Disruption Classifier Developer) trained to date on 975 disruptive and 975 non-disruptive samples

In moving toward the goal of a ML predictive disruption avoidance capability that would NOT require a significant number of ITER disruption events to "train" the predictive software, we will: (i) further examine how accounting for features from multi-D signals can improve prediction accuracy; and (ii) increase emphasis on "supervised" ML methodology via improving ML software portability through multi-machine studies that include integration of threshold conditions for key disruption-relevant physics events into the workflow to improve our classification software.

References

(1) "An Advanced Disruption Predictor for JET Tested in a Simulated Real-time Environment," G. A. Rattá, et al., Nuclear Fusion 50.2: 025005 (2010);

(2) "Development of an Efficient Real-time Disruption Predictor on JET & Implications for ITER," S. Dormido-Canto, J. Vega, et al., Nuclear Fusion 53: 113001 (March, 2013).

(3) "Survey of Disruption Causes at JET," P. C. DeVries, Nuclear Fusion 51: 053018 (April, 2011).