Machine Learning for Magnetic Fusion Science Part II

Cristina Rea

crea@mit.edu

MIT Plasma Science and Fusion Center, Cambridge, MA, USA

Joint ICTP-IAEA School on AI for Nuclear, Plasma and Fusion Science, Trieste, May 24th- 25th, 2023

PSFC

Outline

1. Brief Fusion primer
2. ~~The Universality theorem and brief ML taxonomy~~
3. Explainable deep learning vs design of interpretable models
   - disruptions!
4. Domain adaptation and transfer learning
   - disruptions!
5. Current challenges and opportunities for future research
   - disruptions!
6. Conclusions

Part I

Part II

Models interpretable by design vs black boxes that can be "explained"

existing dichotomy

Interpretable by design vs **Black boxes**

- Ensemble methods,
- Convolutional Neural Networks,
- Recurrent Neural Networks,
- Autoencoders,
- …

why did we get the output we got? → *post-hoc explainability*

explanation by

simplification → linear surrogates

feature relevance → saliency maps

additional models …

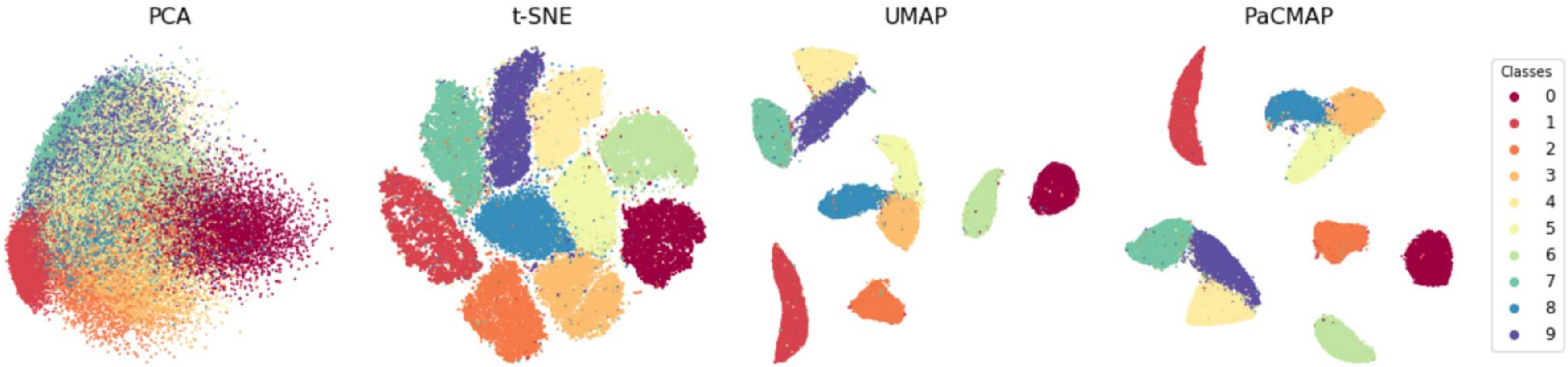
Examples of explainable models interpretable by design

Interpretable by design

- ❑ Dimensionality Reduction
- ❑ Physics Informed Neural Networks

Dimensionality reduction (DR) enables inspection of dataset structure

Dataset of handwritten digits **represented** through different **embeddings in latent space**.



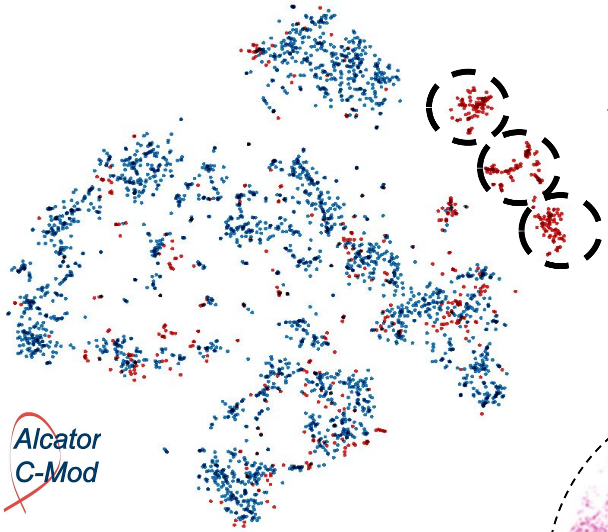
C. Rudin et al., 2022 *Stat. Surv.* **16** 1–85

- Latent space (no physical units) allows 2D visualization of **similar data points** in high-dimensional feature space.
  - All DR methods allow some form of data inspection and understanding.

Coloring done a-posteriori!

Clustering algorithms enable discovery of data patterns

Fusion

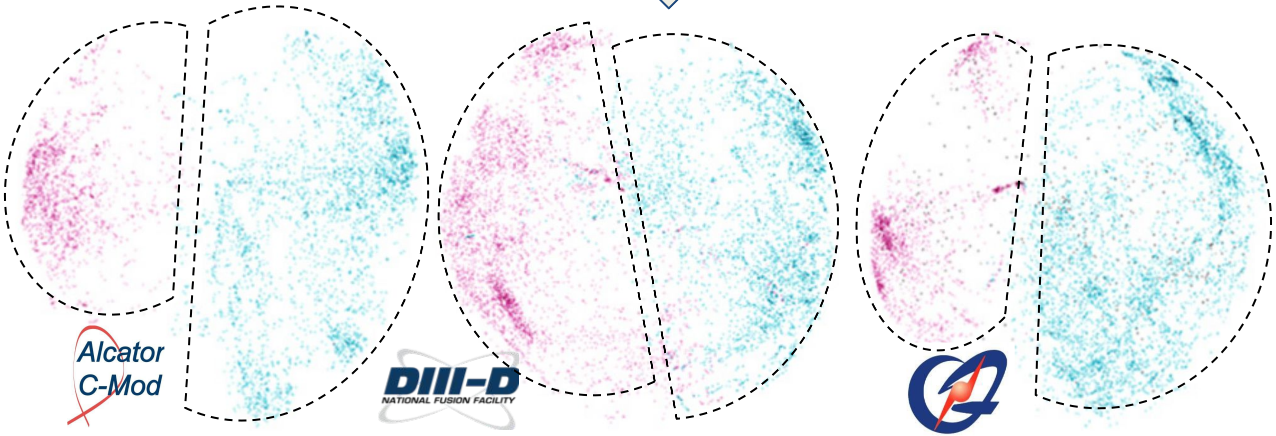


t-SNE clustering of C-Mod disruptive vs non-disruptive time sequences.

PCA clustering of two different performance regimes for three different tokamaks.

J.X. Zhu et al, 2021 Nucl. Fusion 61 026007

Coloring done a-posteriori!



J.X. Zhu et al, 2021 Nucl. Fusion 61 114005

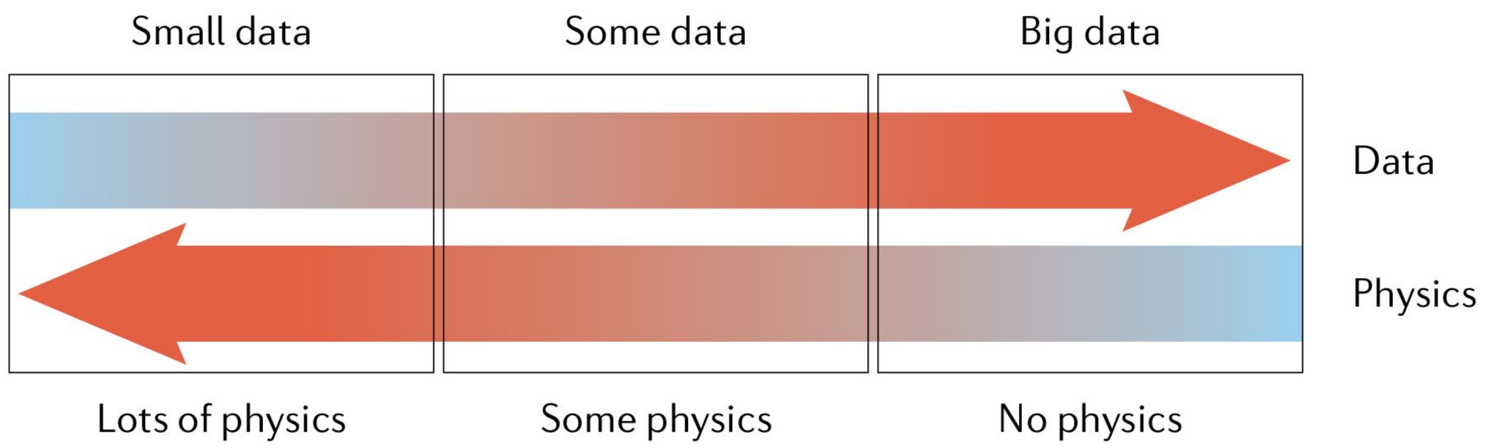
Examples of explainable models interpretable by design

Interpretable by design

- ❏ Dimensionality Reduction
- ❏ Physics Informed Neural Networks

Physics-informed machine learning seamlessly integrates data and governing physical laws

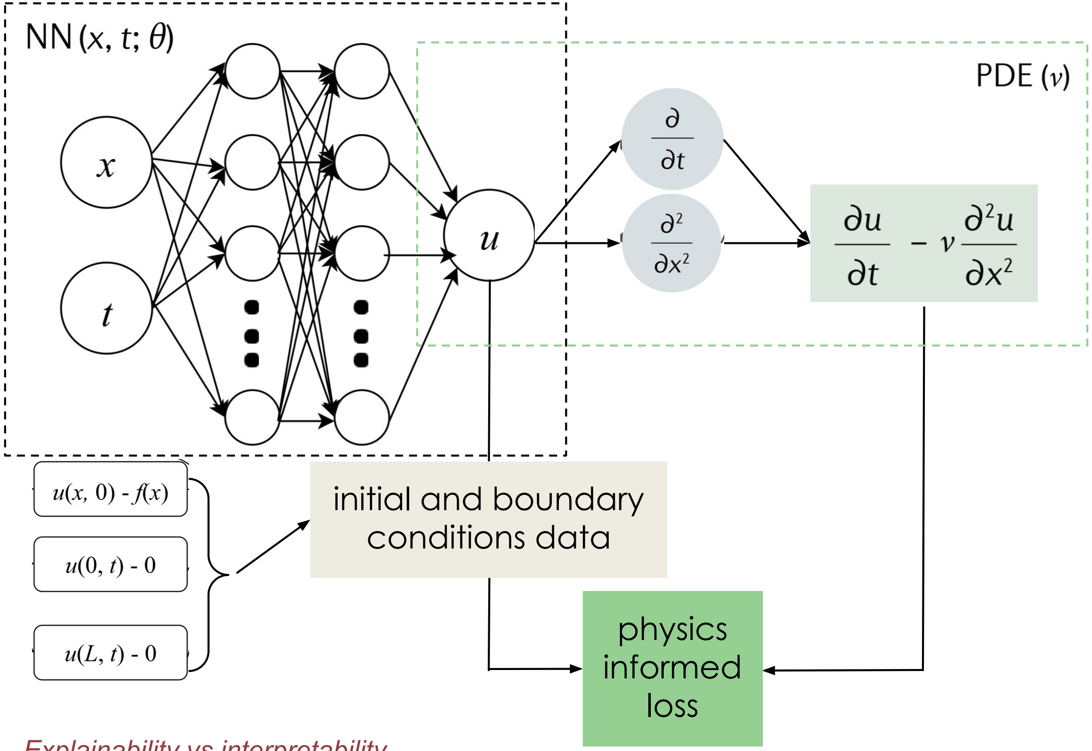
- NN and AutoDiff allow to design models with **partially missing physics (or data!)**
    - No need of domain adaptation or transfer learning.
    - Strong generalization, by enforcing/embedding physics constraints.
    - Can tackle high-dimensional problems.
    - Can address uncertainty due to physics, data, and learning models.



G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440

Physics Informed Neural Networks (PINNs) preserve interpretability through physics constraints

- PINN learns partial differential equations (PDEs) given initial and boundary conditions (I&BC): heat equation example.



- PINN training minimizes the PDEs residuals + I&BC, through **combined loss function** and **automatic differentiation**.
- **No need of labeled data, only generative constraints!**

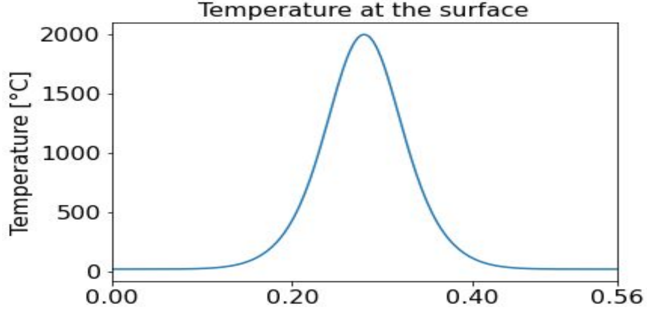
Adapted from:
C. Rudin et al., 2022 Stat. Surv. 16 1–85
G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440

PINN solves heat equation and computes heat flux on the top surface of W7-X divertor tiles

E. Aymerich et al., PSI-25 Poster

Fusion

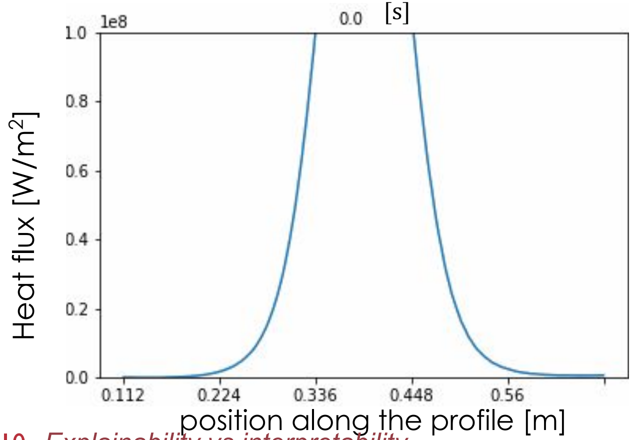
Gaussian top boundary condition constant in time $(t \in [0,0.1]\ \text{s})$



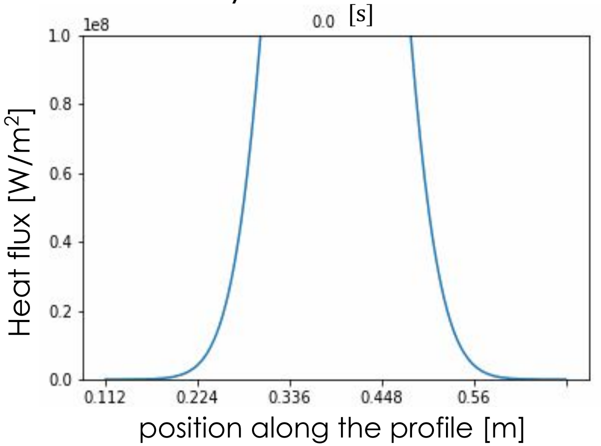
The PINN solves the equation and then computes the derivative on the top surface of the profile and estimate the normal heat flux thanks to automatic differentiation:

$$q = -\nabla_n u$$

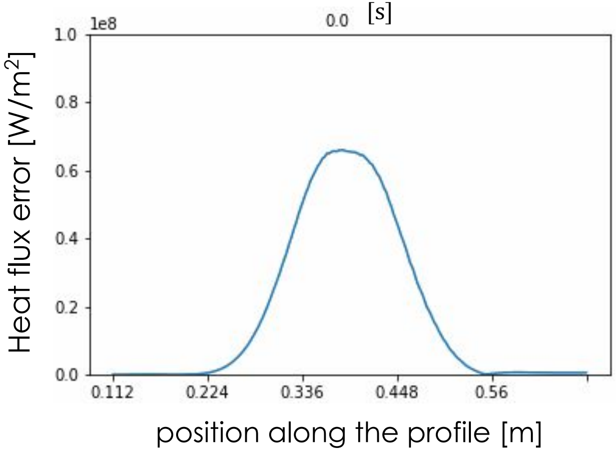
PINN



Physics simulation



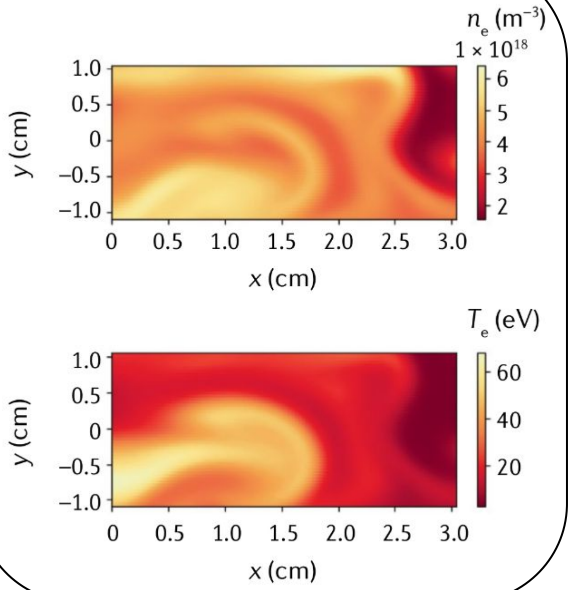
Error



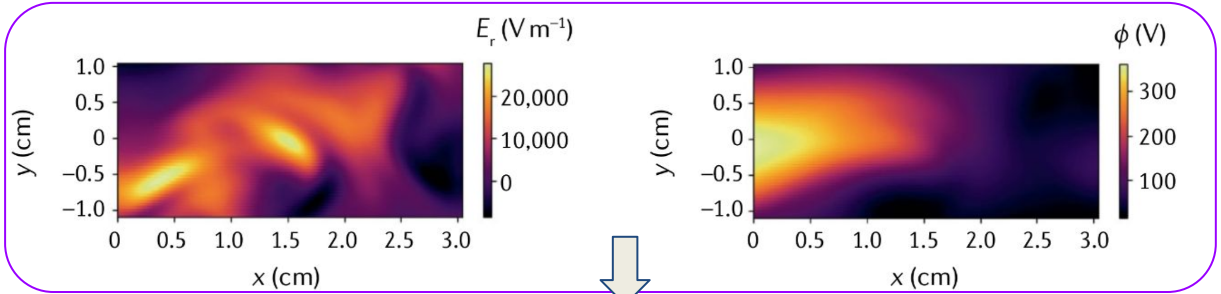
PINNs can accurately learn turbulent field dynamics consistent with theory, and from partial observations

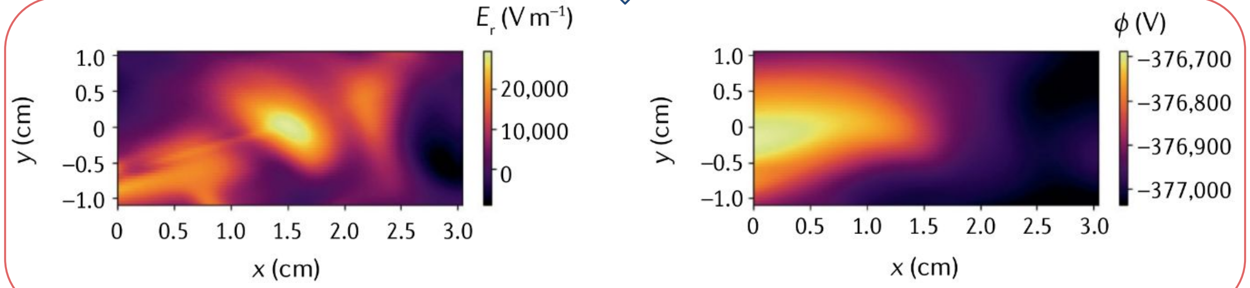
Fusion

Partial observations of Te, ne from one test discharge



Reference target solutions





A. Mathews, et al, Phys. Rev. E 104, 025205 (2021)

PINN reconstructions

Physics-informed machine learning: current limitations

- **Multiscale and multiphysics problems require further developments**.
  - High-frequency functions difficult to learn → F-principle or spectral bias.
- **PI ML involves highly non-convex optimization problems for complex loss functions**.
  - Need more robust algorithms and computational frameworks.
  - Meta-learning techniques to automate the design of best architectures?
- **Missing benchmarks** on openly available datasets from physics, chemistry, …
- **More research needed on the theoretical foundations of NN**.

G.E. Karniadakis et al., 2021 Nat Rev Phys 3, 422–440 and references therein

Explainable deep learning vs design of interpretable learning – recap slide

1. ML often used as black box, but high-stakes decisions imply need for inspections.

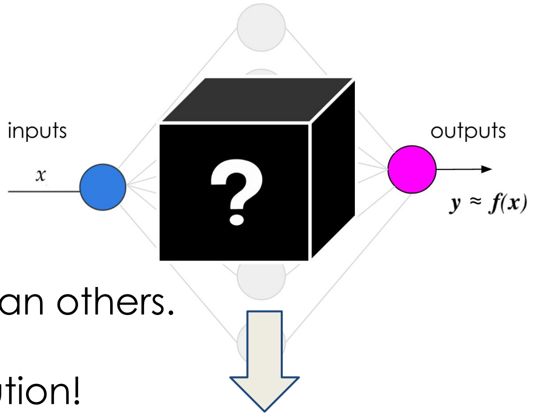
2. Fictional dichotomy between accuracy and interpretability.

3. Given a reasonably well described problem, multiple models of comparable accuracy may exist, some likely more interpretable than others.

4. Post-hoc explanation to black boxes viable approach, but use caution!

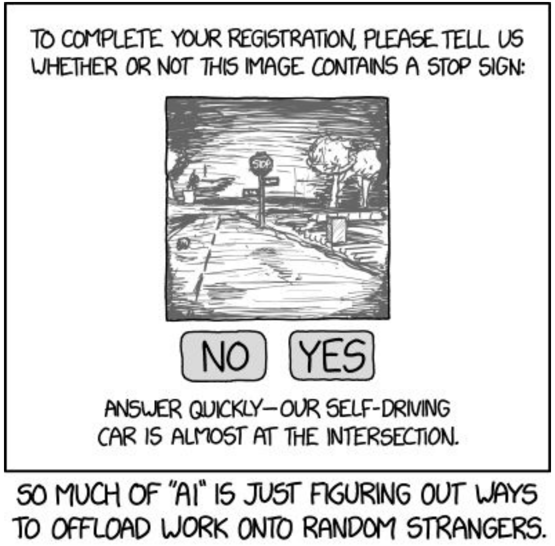
5. Design of interpretable models may be computationally hard → huge potential and should be preferred.

6. **Fusion examples already out there!**



Outline

1. Brief Fusion primer
2. The Universality theorem and brief ML taxonomy
3. Explainable deep learning vs design of interpretable models
4. Domain adaptation and transfer learning
5. Current challenges and opportunities for future research
6. Conclusions

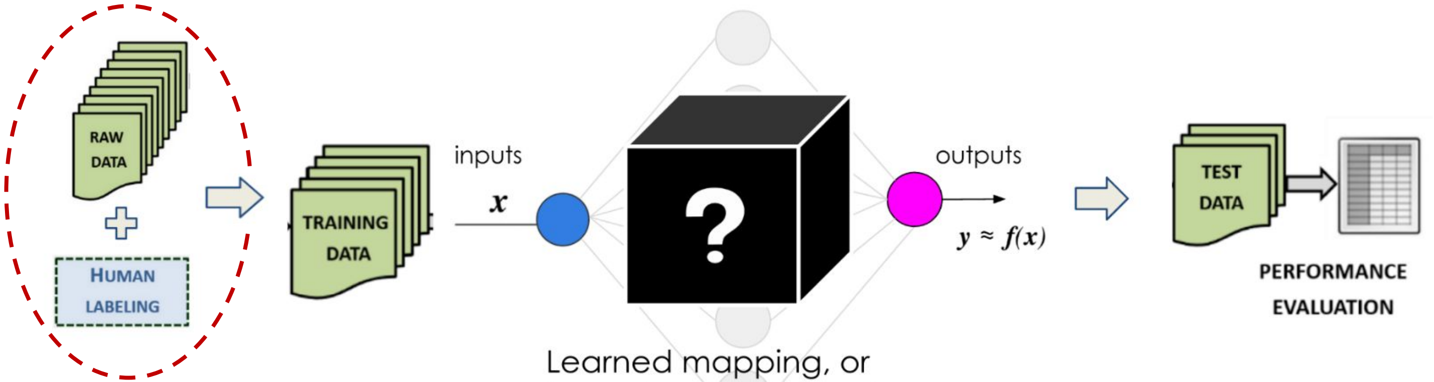


https://xkcd.com/1897/

ML mapping from inputs to outputs, or learning to perform a task

Simplified supervised ML classification workflow:

Adapted from A. Pau et al, Nuclear Fusion, 59(10):106017, 2019



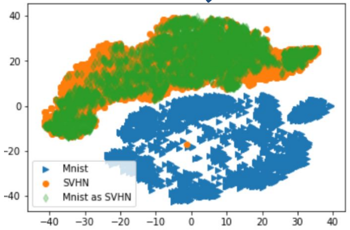
domain = data + labels

Learned mapping, or boundary between classes or **learning a task**

- Mapping from inputs to outputs through ML systems means to **learn to perform a task**.

Learning a task heavily depends on dataset composition

- What if the collected data is **not an accurate reflection of the population**? Too limited, not accurately labeled, …
  - Learning *a general data representation* → e.g., how can I predict *any* digits' labels? By finding common embeddings of source/target data!



(d) MNIST to SVHN

Source data →

Generated data → on target domain

Real target data ↓

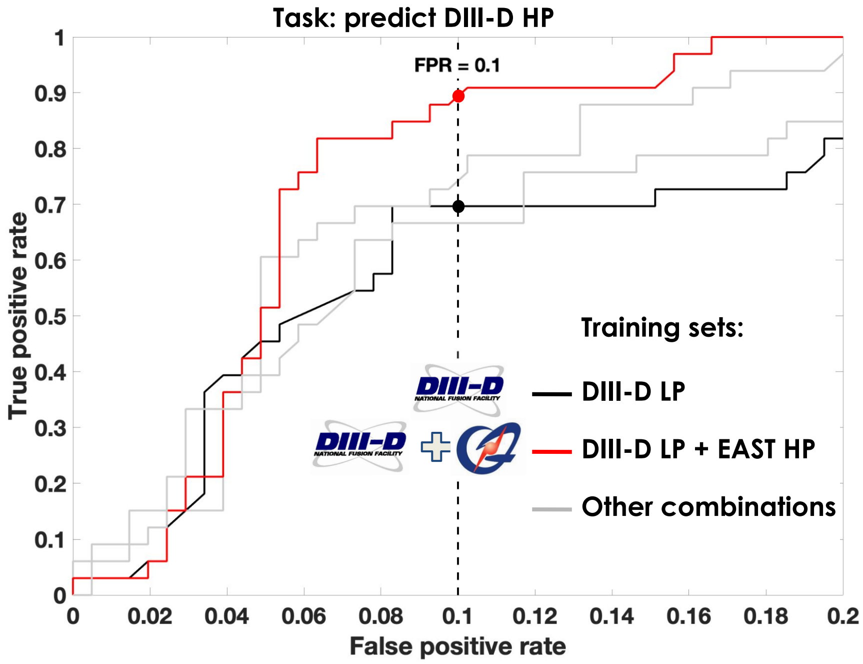
P. Russo et al, "From source to target and back: Symmetric Bi-Directional Adaptive GAN" https://arxiv.org/pdf/1705.08824.pdf

Domain adaptation and **transfer learning** designed to overcome **biased** dataset and/or **generalize** knowledge across **different tasks**.

Adaptive strategies designed to optimize predictions across different fusion devices

Fusion

- **Adapt** current state-of-the-art **ML predictors** to different operational regimes across devices (DIII-D/EAST).
- **Implications for next-gen, yet-to-be-built devices!**
- Adaptive strategies:
  - ad-hoc design of **training sets to match target domain** by fully exploiting existing data[1],
  - **retrain predictors** after performance degradation[2].



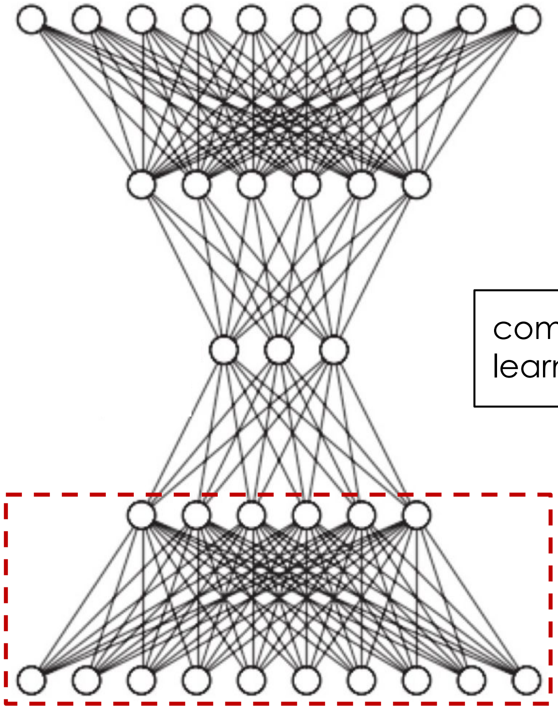
Adapted from [1]J.X. Zhu et al, NF (2021) 114005
[2]J. Vega et al., *Nat. Phys.* 18, 741–750 (2022)

Domain adaptation

Source domain (simulations) allows to learn how to reconstruct target data (experiments)

Fusion

task: map from sim back to sim



task: retrain with experimental observations

NN inputs: simulation outputs

↓

compress data, learn mapping

↓

reconstruction of original inputs

- **Large datasets** built through inexpensive but possibly inaccurate **simulations**.
- Networks (autoencoders) trained to learn **mapping** between **sim-to-sim** inputs to outputs.
- Mapping then transferred to new task → **learning corrective transformation mapping sim-to-exp** → **transfer learning**!

Adapted from Humbird et al., PoP 28, 042709 2021

Transfer learning

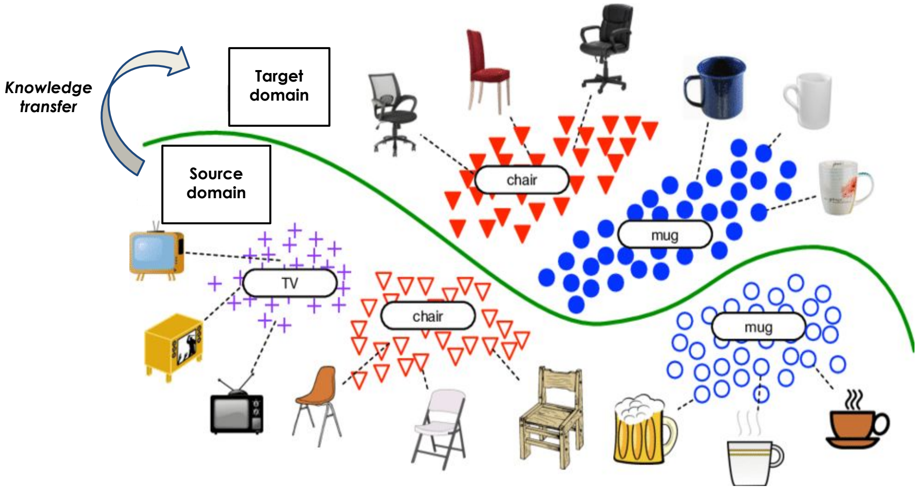
Domain adaptation and transfer learning – recap slide

1. Domain adaptation and transfer learning allow black box systems to overcome biased dataset and generalize knowledge across different tasks.

2. Domain adaptation: learn the same task under differently distributed source and target domains.

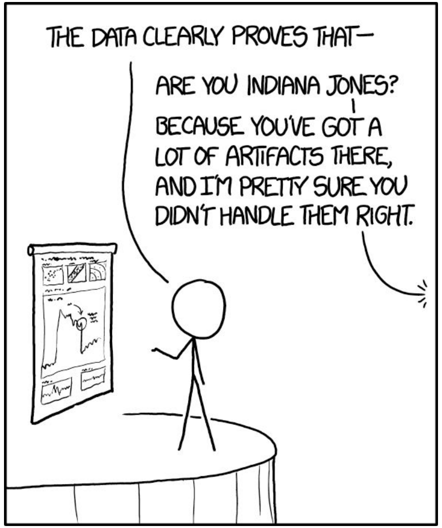
3. Transfer learning: learn to perform new task on limited target domain, given knowledge gained on source domain.

4. **Fusion examples already out there!**



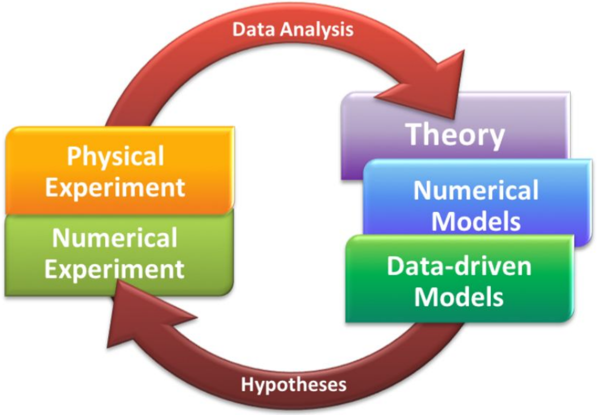
Outline

1. Brief Fusion primer
2. The Universality theorem and brief ML taxonomy
3. Explainable deep learning vs design of interpretable models
4. Domain adaptation and transfer learning
5. Current challenges and opportunities for future research
6. Conclusions



https://xkcd.com/1781/

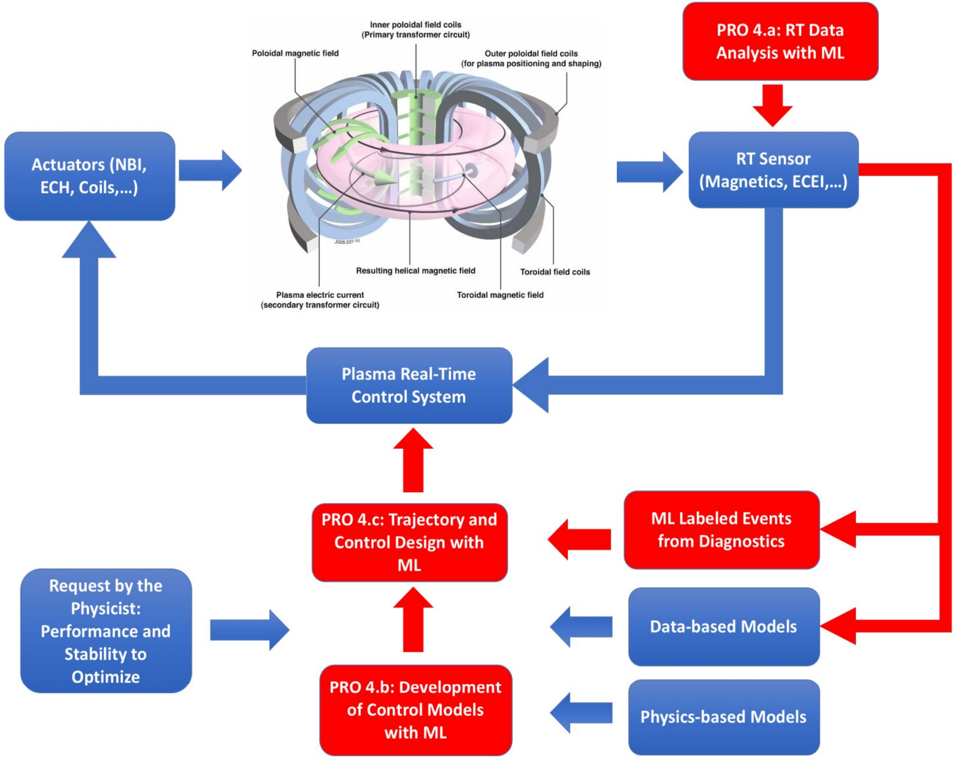
Data drives fusion experiments' design, simulation, analysis, and optimization …



D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)

- **ML enabling science discovery** bridging gaps in theoretical understanding

…and it will be an essential design and control tool for next-gen devices!



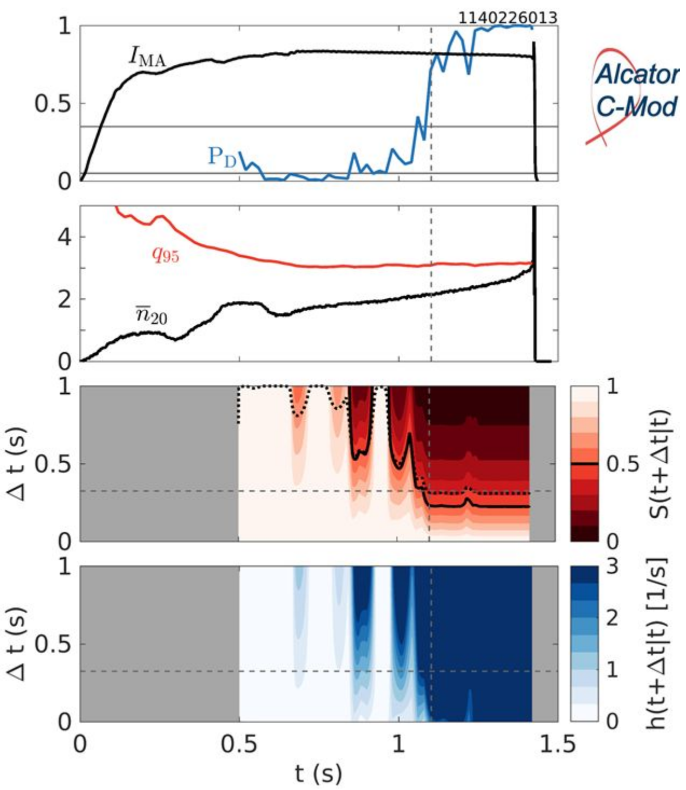
Adapted from D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)

Need:

- Robust model **interpretability/explainability**.
- Well-defined validity and extrapolability boundaries → **Uncertainty Quantification**.

Let's review some more examples…

Predict "time-to-disruption" risk using classification probability



Any classification probability (P_D) cast between [0,1] can be used to:

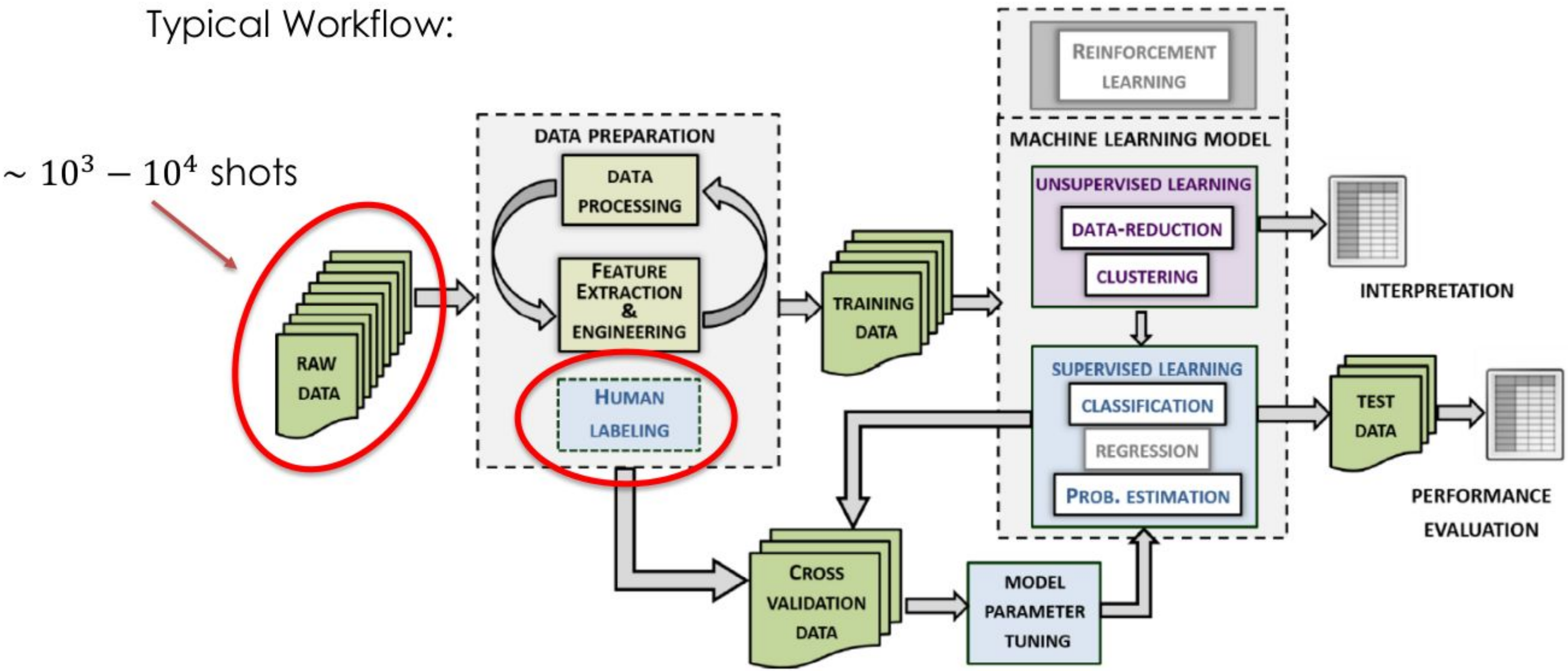
- Predict the **future probability** of **plasma survival** $S(t+\Delta t|t)$ [1]
  or
- Model the **instantaneous hazard** [2,3] h=d ln$S$/dt to be used as probability generator.

Hazard function modeling connects dynamical systems and risk-aware control design by probability generation.

C-Mod data used as proof of concept to combine DPRF (or <u>any classifier</u>) disruptivity with survival analysis.

[1] RA Tinguely et al 2019 PPCF 61
[2] KEJ Olofsson et al 2018 PPCF 60
[3] KEJ Olofsson et al 2018 FED 146

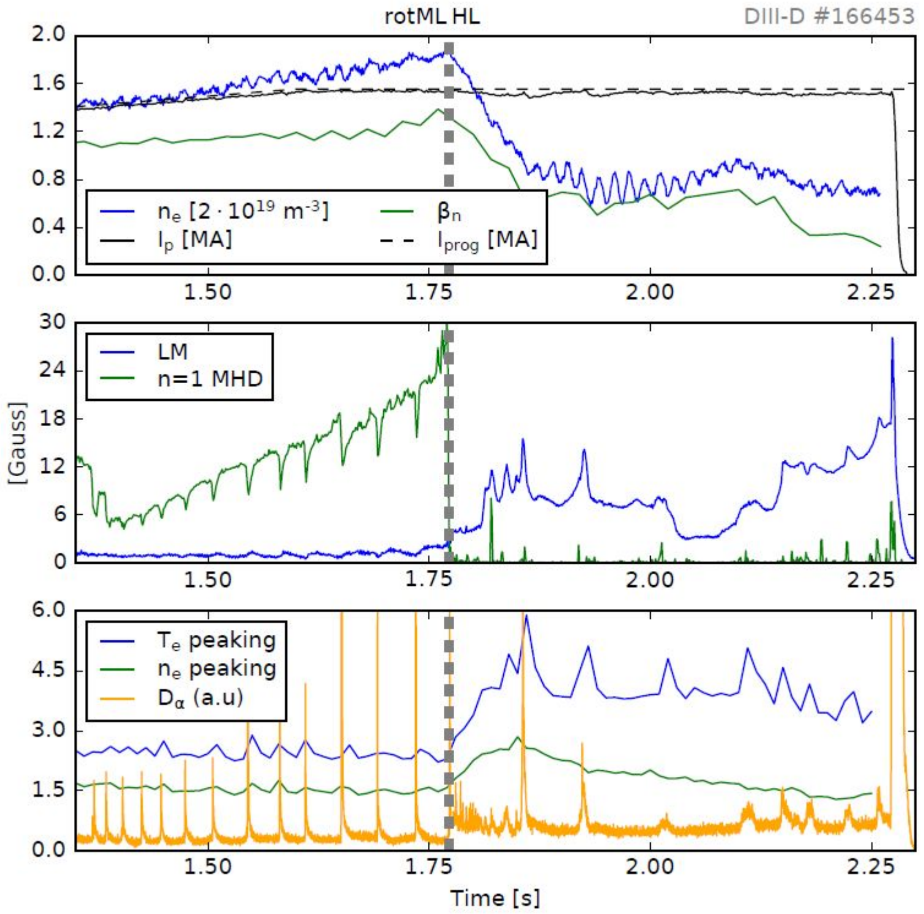
ML requires large datasets with target events labeled



A. Pau et al, Nuclear Fusion, 59(10):106017, 2019

Labeling events often requires manual inspection of multiple signals

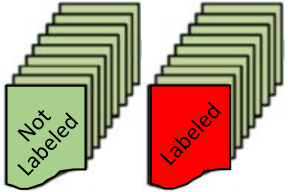
- **Mode locking (ML) event**
  - Rotating and non-rotating mode signals
- **H-L back transition (HL) event**
  - n_e and $\mathrm{T_e}$ profile pedestals
  - Pressure/stored energy
  - ELMs



Semi-supervised ML used to accelerate event labeling

Most shots go unexamined

Partially Labeled Event Data



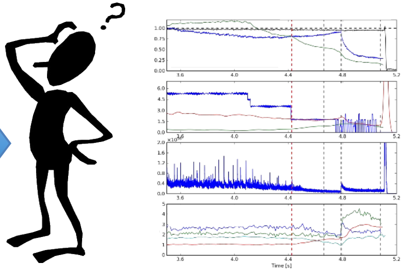
- Common signals from many shots
- A fraction (at least 1 shot) has labeled event occurrence

Diagnostic Data

- From MDSplus, etc.
- Many ($\sim 10^3$) shots

Some shots analyzed

Human Expert



- Use diagnostic signals to identify events
- Record times at which event occurs
- Validate 'learned' event detections

Predictions & performance results

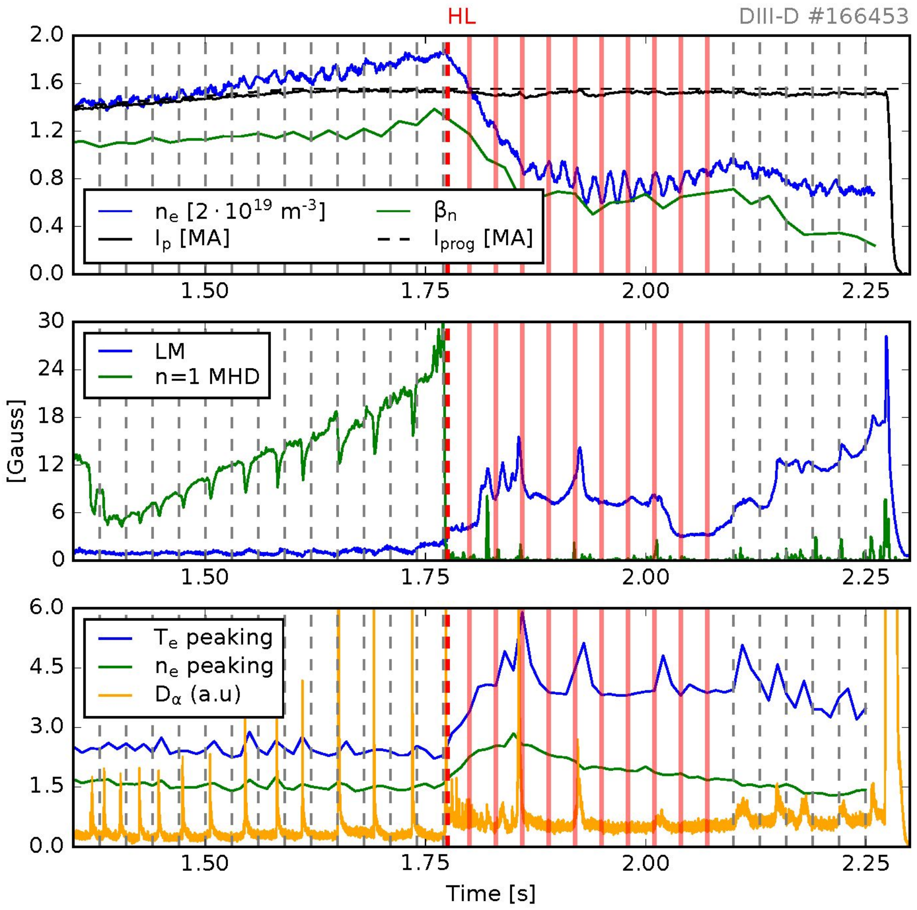
Semi-Supervised Learner

- Use unlabeled data distribution to find other events
- Output times when event likely occurred

Problem setup requires user-specified time scales and shots with example events

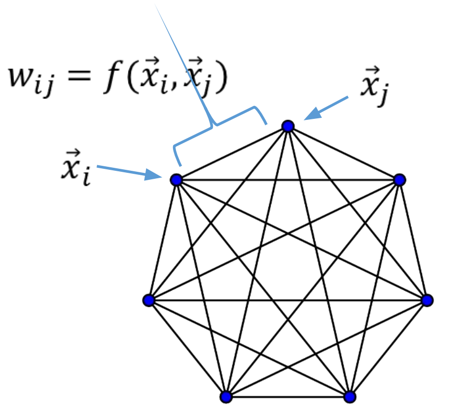
- Sample time sequences from each shot (endpoints are shown) with…
  - Duration > event timescale
  - # of steps > event resolution
- Choose N signals for event detection
  - Time sequences $\vec{x}_i \in \mathbb{R}^{N \cdot (\# \; of \; steps)}$
- Class assigned to each observed sequence
  - Positive (overlaps with event time)
  - Negative (otherwise)

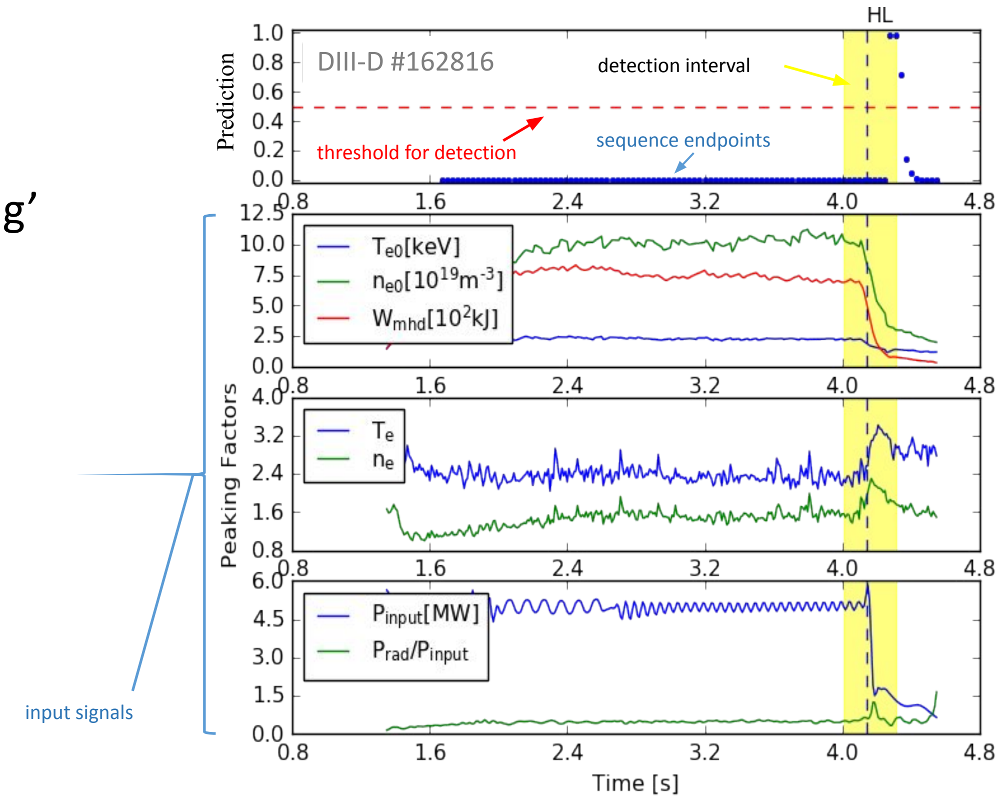
Semi-supervised learning:
Infer classes of unobserved sequences



Applied label spreading algorithm to automate detection of physics events[1]

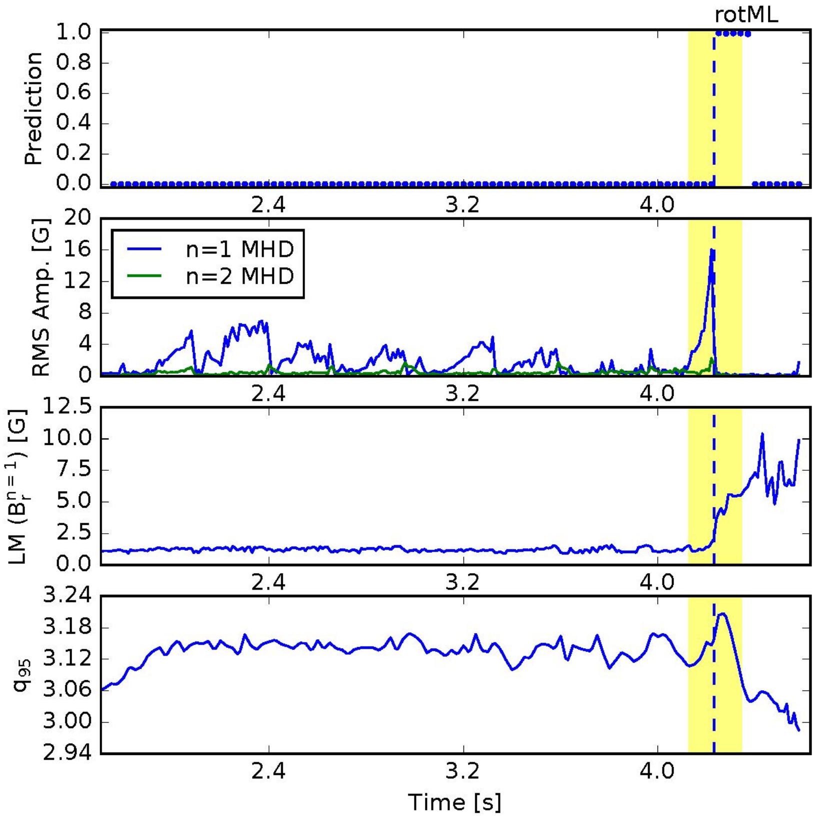
- Graph-based algorithm
  - Nodes are time sequences
  - Edges are weighted by proximity
  - Classifications made by 'spreading' information from labeled to unlabeled nodes



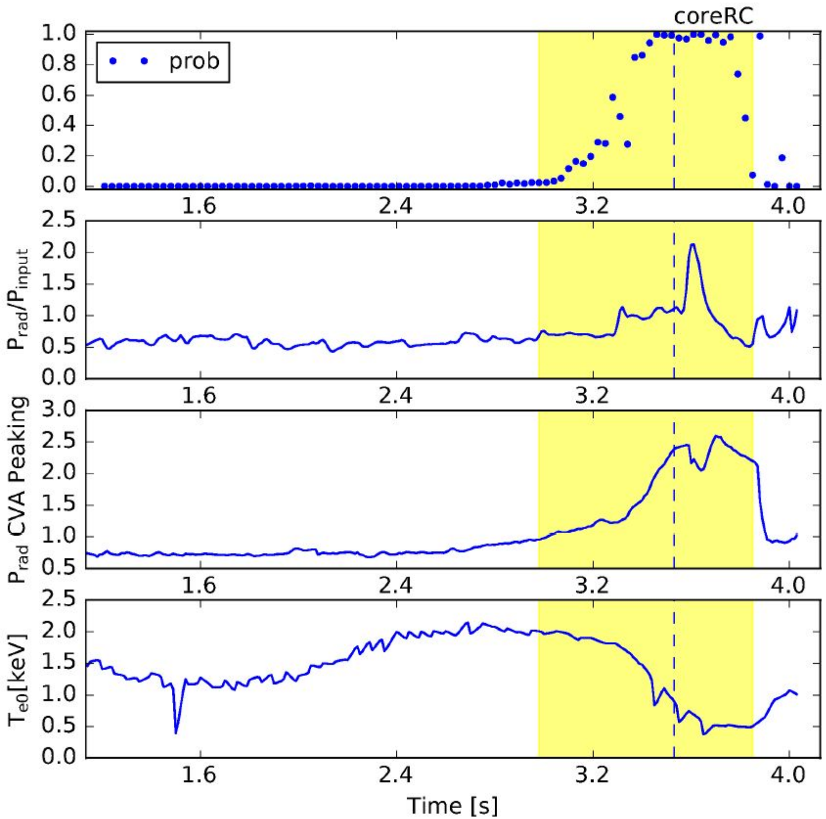


Applied label spreading algorithm to automate detection of physics events[1]

Rotating mode locks (**rotML**)



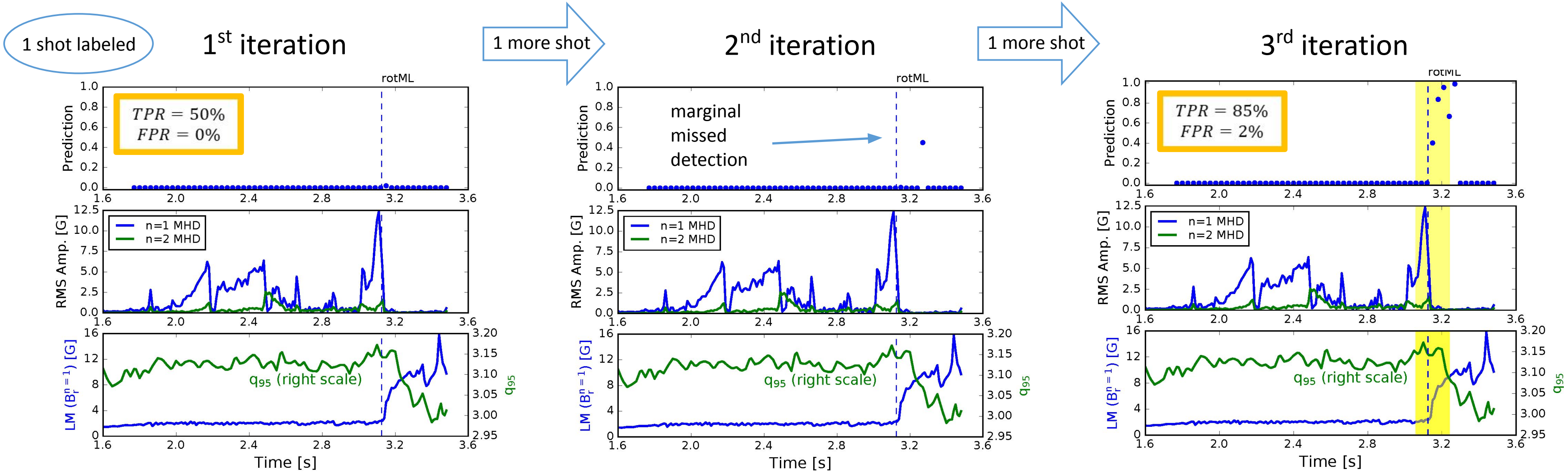
Core radiative collapse (**coreRC**)



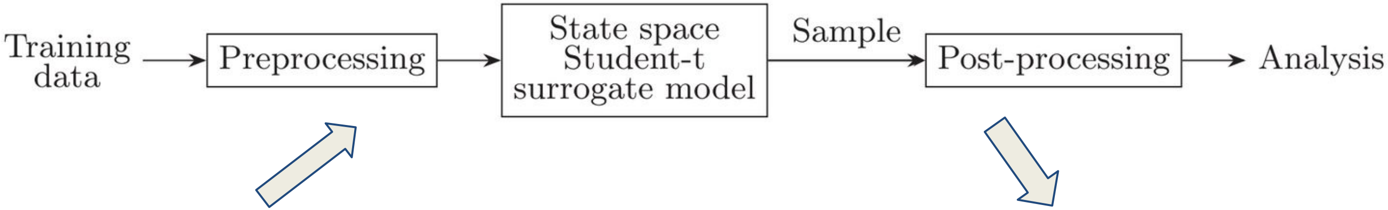
[1] Montes et al 2021 Nuclear Fusion **61** 026022

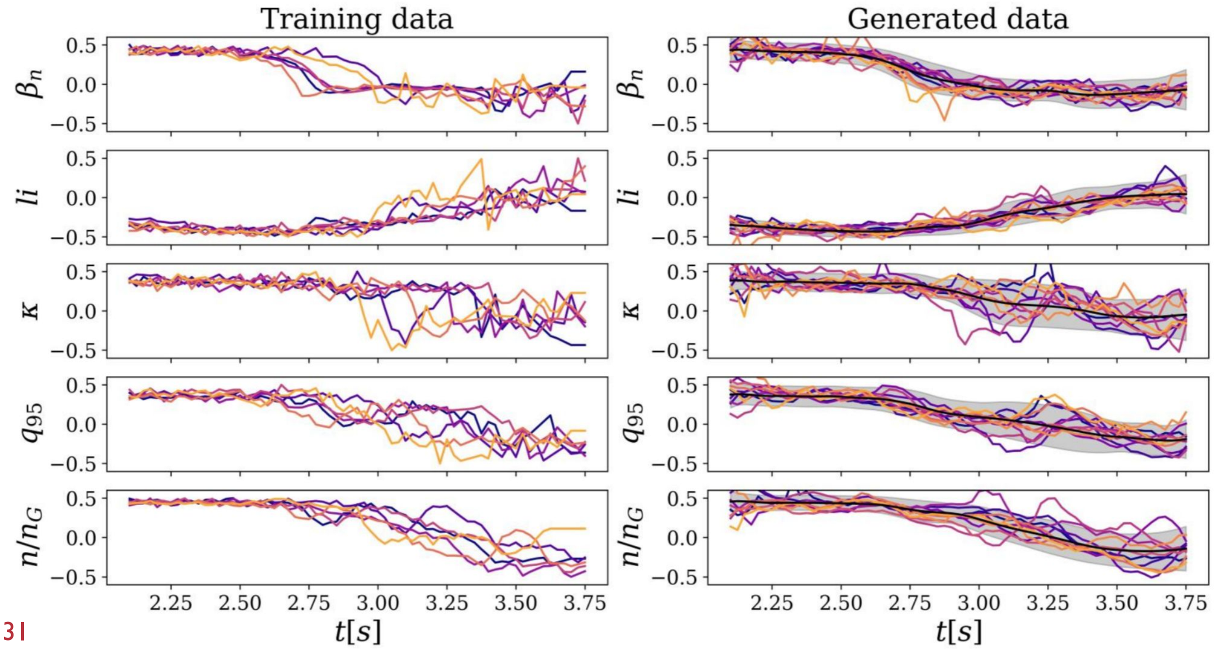
Can be applied to accelerate the construction of events databases

- Iteratively choose initially labeled shot from set of marginal detections
- Prediction quality on unlabeled shots improves consistently
  - Detected events in ~ 85% of hundreds of shots after manually analyzing just ~ 1%



Data augmentation to learn disruptive dynamics





- **DL models are data-greedy**: need comprehensive training database to achieve satisfying and reliable results.
- Robust augmentation of the training database using **state space Student-T surrogate models**.

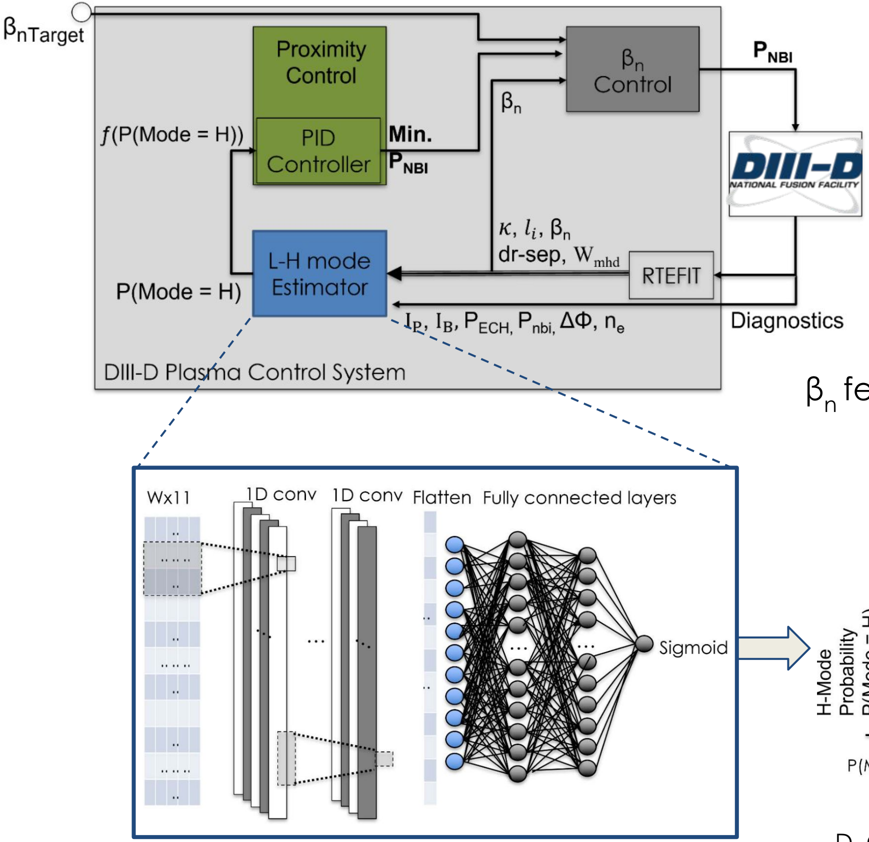
K. Rath,…, C Rea et al, "Data augmentation for disruption prediction via robust surrogate models" J. Plasma Phys. (2022), vol. 88, 895880502

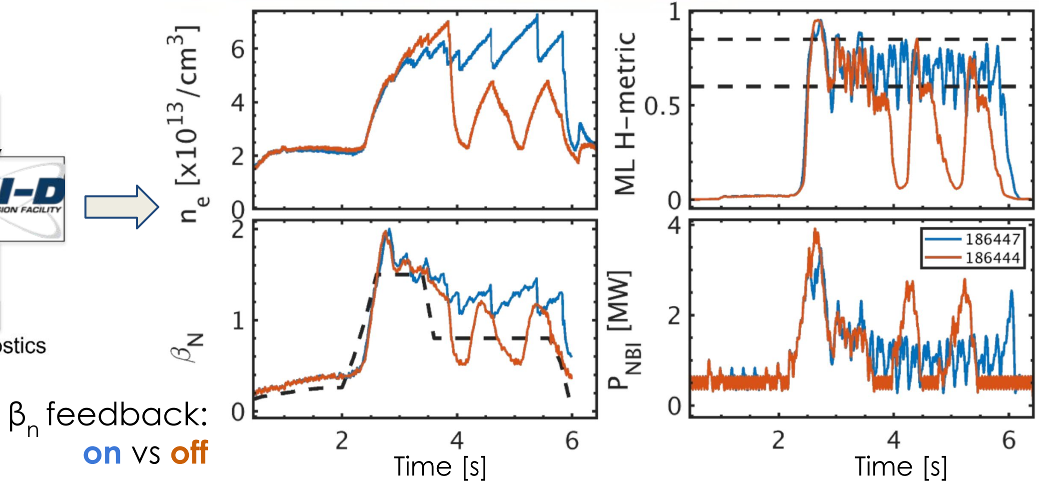
Hybrid Deep Learning predictor – if we have time

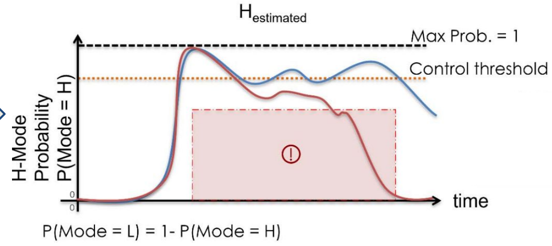
If not: J.X. Zhu, C. Rea et al, 2023 Nucl. Fusion 63 046009

More (non-disruptive) examples:

Temporal Convolutional Neural Network predicts confinement probability 1ms in the future







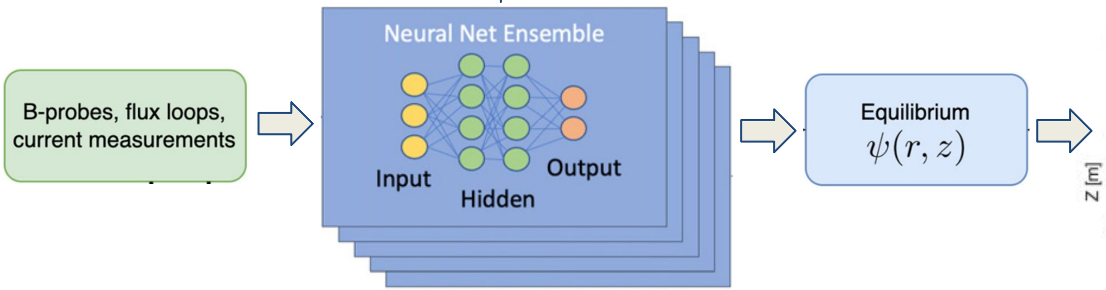
caution: limited expert-labeled dataset constraints model applicability

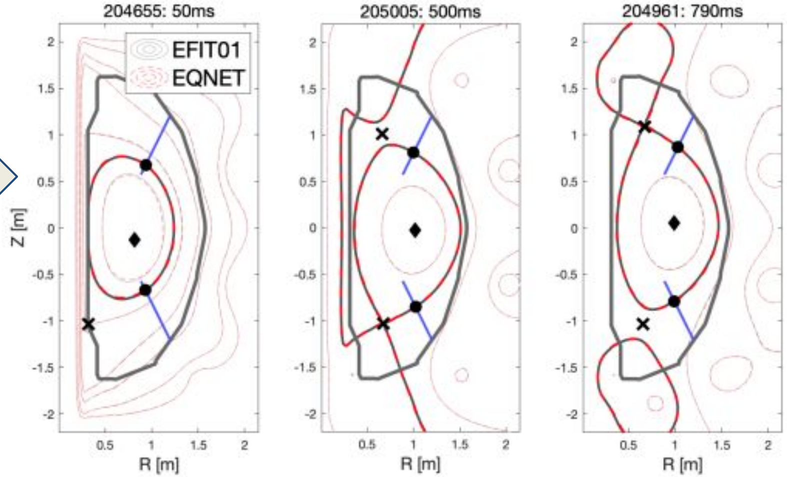
D. Orozco et al, IEEE TPS 2022

Neural networks accelerate equilibria reconstructions and profile evolution for shot planning and real-time control

caution: extrapolation to never seen equilibria or to other machines

$$\frac{\partial^2 \psi}{\partial r^2} - \frac{1}{r}\frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} = -\mu_0 r^2 \frac{dp}{d\psi} - \frac{1}{2}\frac{dF^2}{d\psi}$$





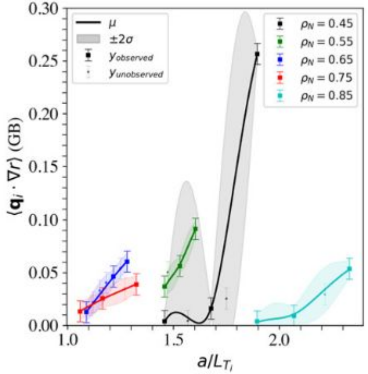
- NN learns free-boundary Grad-Shafranov solutions and reconstructs tokamak equilibria.

Adapted from J. Wai, M. Boyer, E. Kolemen, *Nucl. Fusion 62 (2022) 086042*

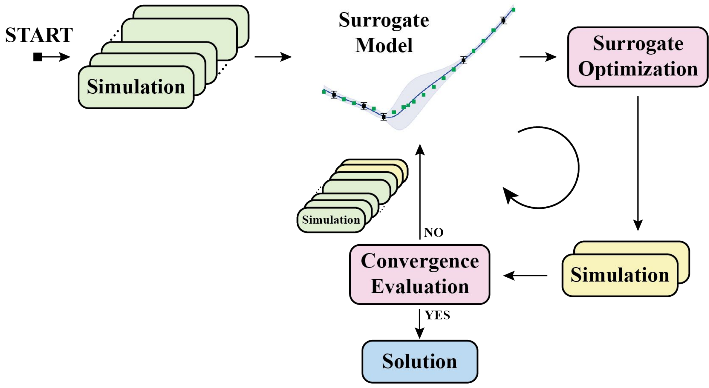
Gaussian Processes (GP) enable nonlinear simulations for performance prediction and gyrokinetic validation

- Few (10-20) simulations required to reach convergence, thanks to Bayesian Optimization (BO) workflow + GP surrogate modeling.
- Enabling profile predictions of unprecedented accuracy for:
  - ✔ Prediction of burning plasma performance (e.g. SPARC)
  - ✔ Validation of gyrokinetic codes (e.g. DIII-D)

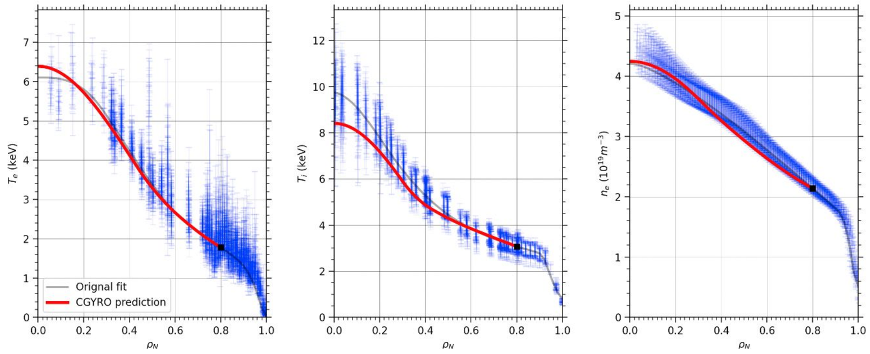
P. Rodriguez-Fernandez et al, Nucl. Fusion 62 (2022) 076036

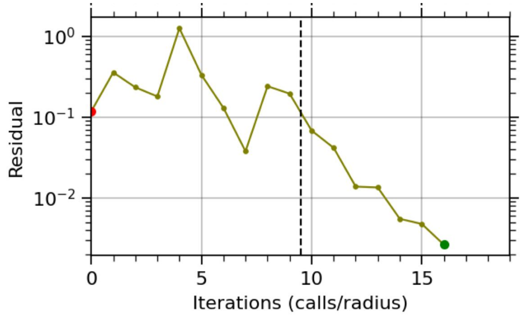


Standard BO workflow to find steady-state profile solutions

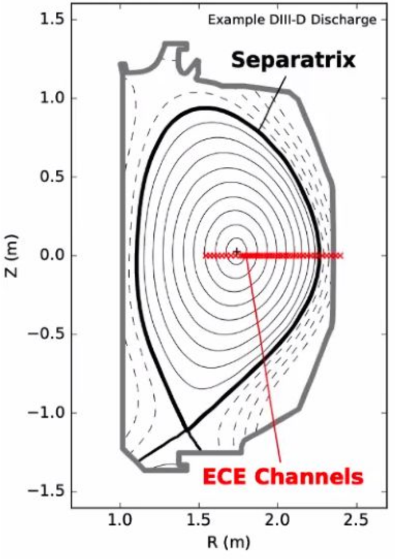


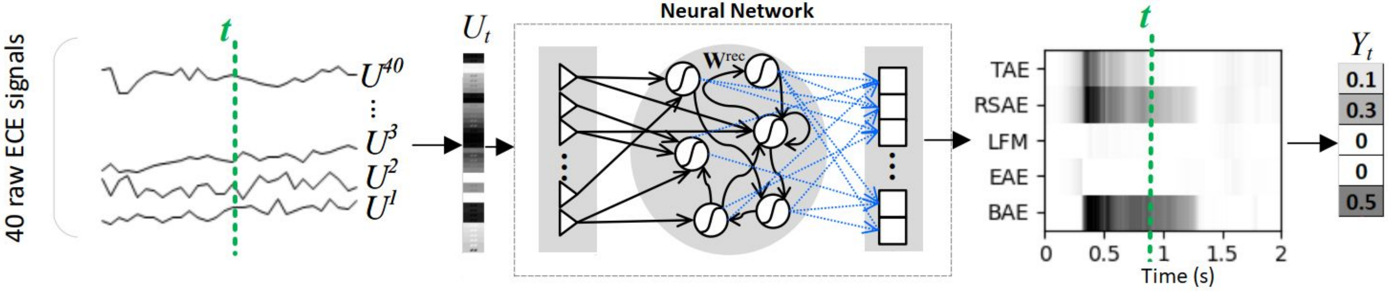
Open question: detangle local minima from unique physical solutions



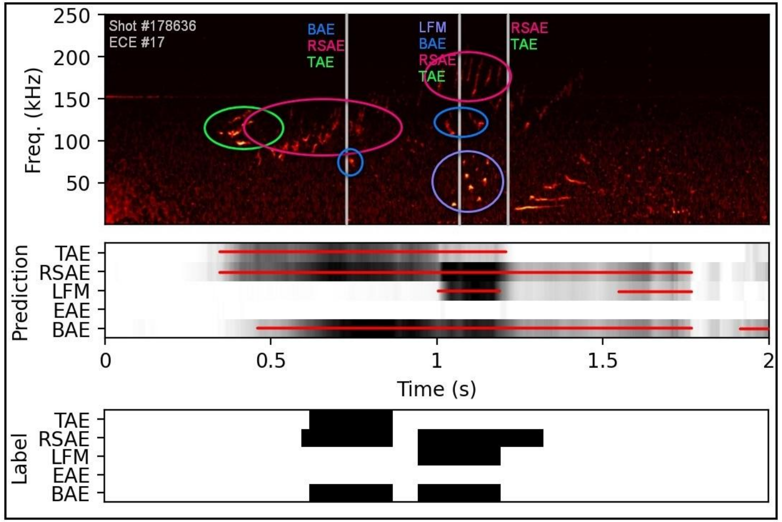


Raw ECE time series input data to Reservoir Computing Network to compute Alfvén Eigenmode score





Jalalvand et al 2022 Nucl. Fusion 62 026007

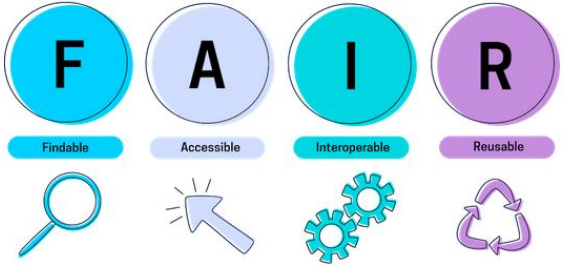


True Positive Rate: %91
False Positive Rate: %7

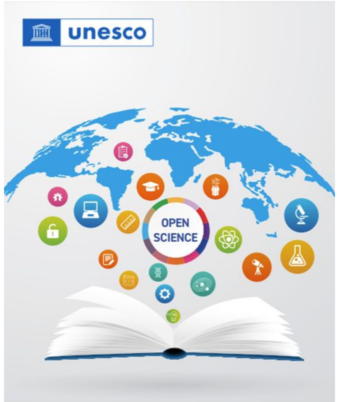
caution: how sensor failures affect ML workflow accuracy

Existing current challenges 😟 but also (!) opportunities for future research 🧑‍🎓 :

- trust in performance metrics → **missing benchmarks**
- trust in predictive output and learning → **model interpretation and explanation accuracy**
- prediction of out of distribution samples → **domain shifts, data shifts**
- integration with legacy architectures → **real-time vs offline implementations**
- **lack of labeled data** or of reliable (and automated) metadata extraction
- **uncertainty quantification**
- **open** and **FAIR** (!) **access** to data and models



M. Wilkinson, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016)



UNESCO Recommendation on Open Science

DOE and International Agencies strongly support ML research to accelerate Fusion progress

- **DOE-sponsored workshop in 2019** critical PROs identified
- **DOE Public Reusable Research Data** (PuRE) initiative

https://science.osti.gov/Initiatives/PuRe-Data

- **IAEA Coordinated Research Project 2022** addressing cross-cutting issues

Workshop on Advancing Fusion with Machine Learning Priority Research Opportunities (PROs)	
Accelerating Science	**Enabling Fusion Energy**
PRO 1: Science Discovery with ML *Hypothesis Generation and Experimental Guidance*	**PRO 4: Control Augmentation with ML** *Diagnostics to Data, Dynamic Models for Control, Fusion Trajectory Design*
PRO 2: ML Boosted Diagnostics *ML Boosted Diagnostics, Physics Enhanced Data*	**PRO 5: Extreme data algorithms** *Extreme-scale Processing, In-situ Data Analysis*
PRO 3: Model Extraction and Reduction *Data-driven Models, Reduction of Complex Code Algorithms*	**PRO 6: Data-enhanced Prediction** *Prediction of Disruption Events and Effects, Plasma Phenomena and State Prediction*
PRO 7: Fusion Data ML Platform	

D. Humphreys et al. "Advancing Fusion With Machine Learning" DOE Workshop (2020)

https://nucleus.iaea.org/sites/ai4atoms/ai4fusion/

Summary and conclusions

1. Fusion science and technology advancements also accelerated by ML

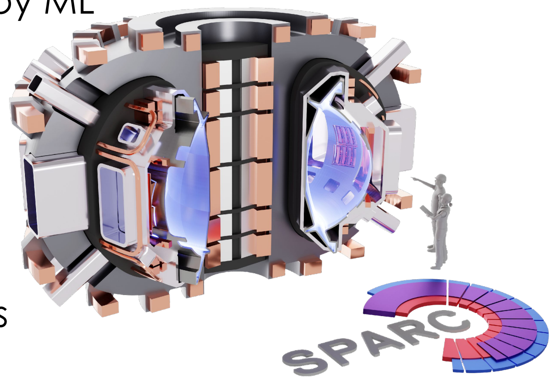
2. Neural networks as universal approximators
   - and accuracy does not prevent interpretability!

3. Interpretable by design models are really powerful

4. Adaptive learning addresses changing domains and learning tasks

5. Fusion examples already out there employing
   a. Interpretable algorithms
   b. Explainable predictions
   c. Transfer learning and statistical optimization
   d. Surrogate modeling for fast reconstructions

Long list of **open questions and cross-cutting challenges**,
but also **opportunities for future research, enabling change in the field!**



Credits: S. Mordijck 💖