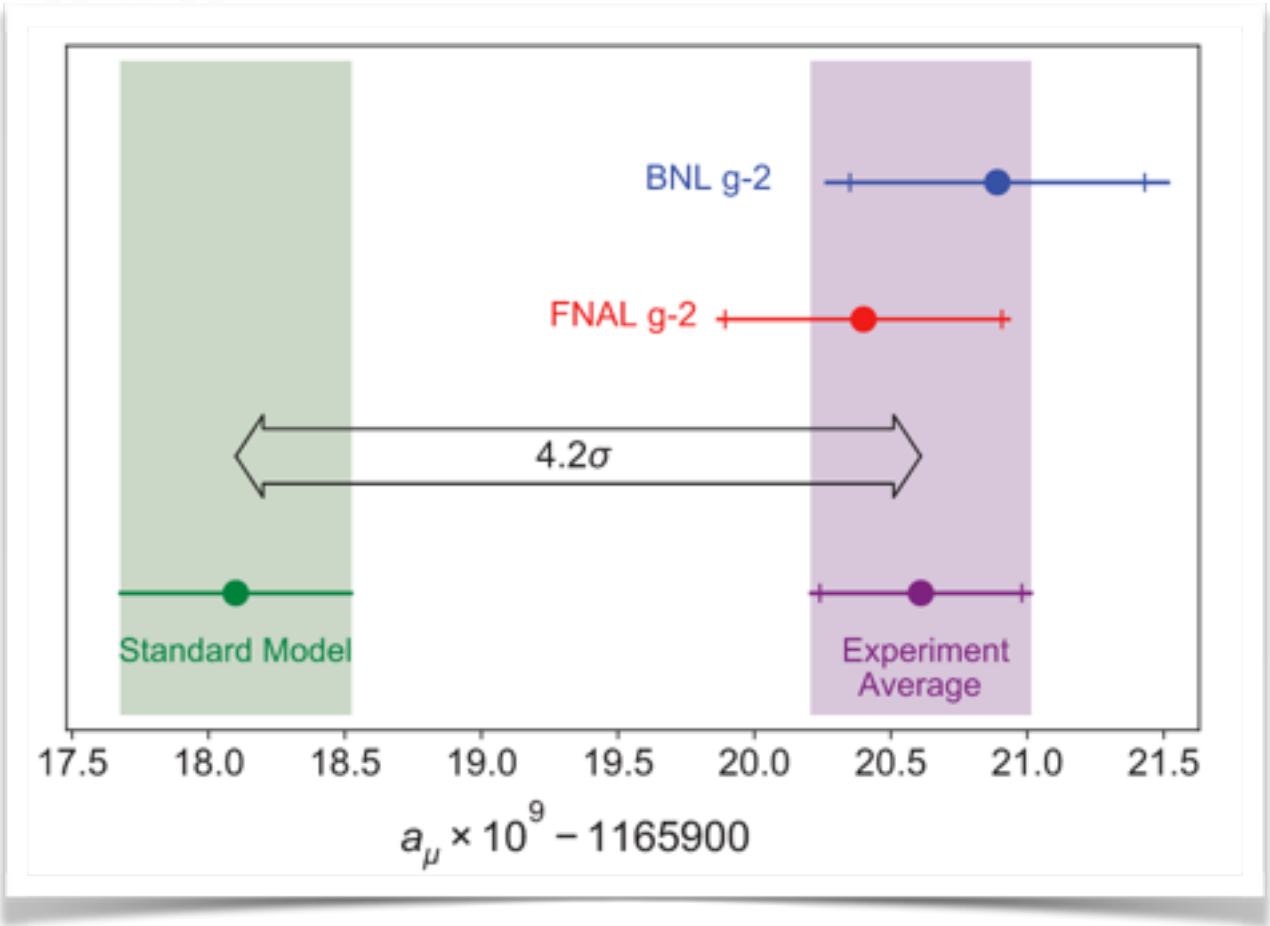
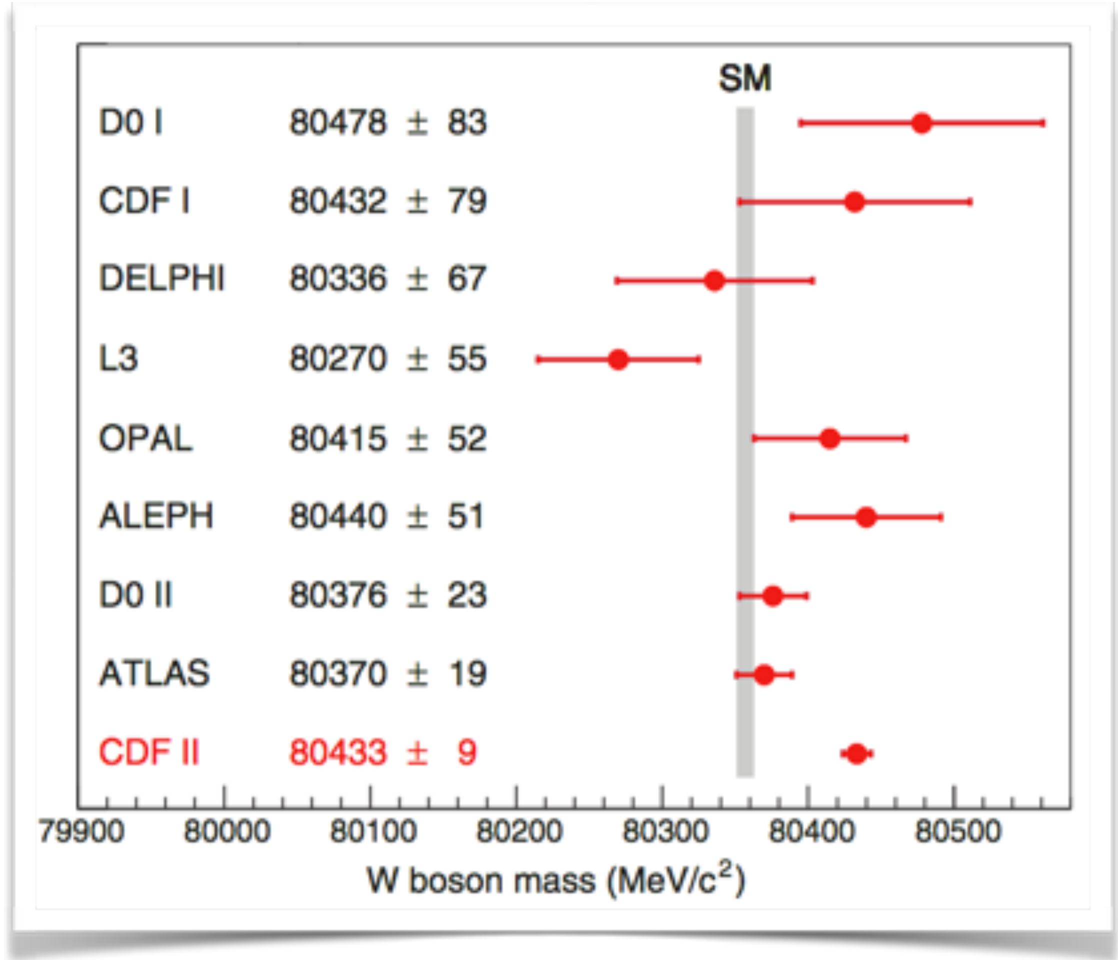
Joint Workshop of the CEPC Physics, Software and New Detector Concept

NOVELTY (ANOMALY) DETECTION AT COLLIDERS

Tao Liu
The Hong Kong University of Science and Technology

Hints for NP?





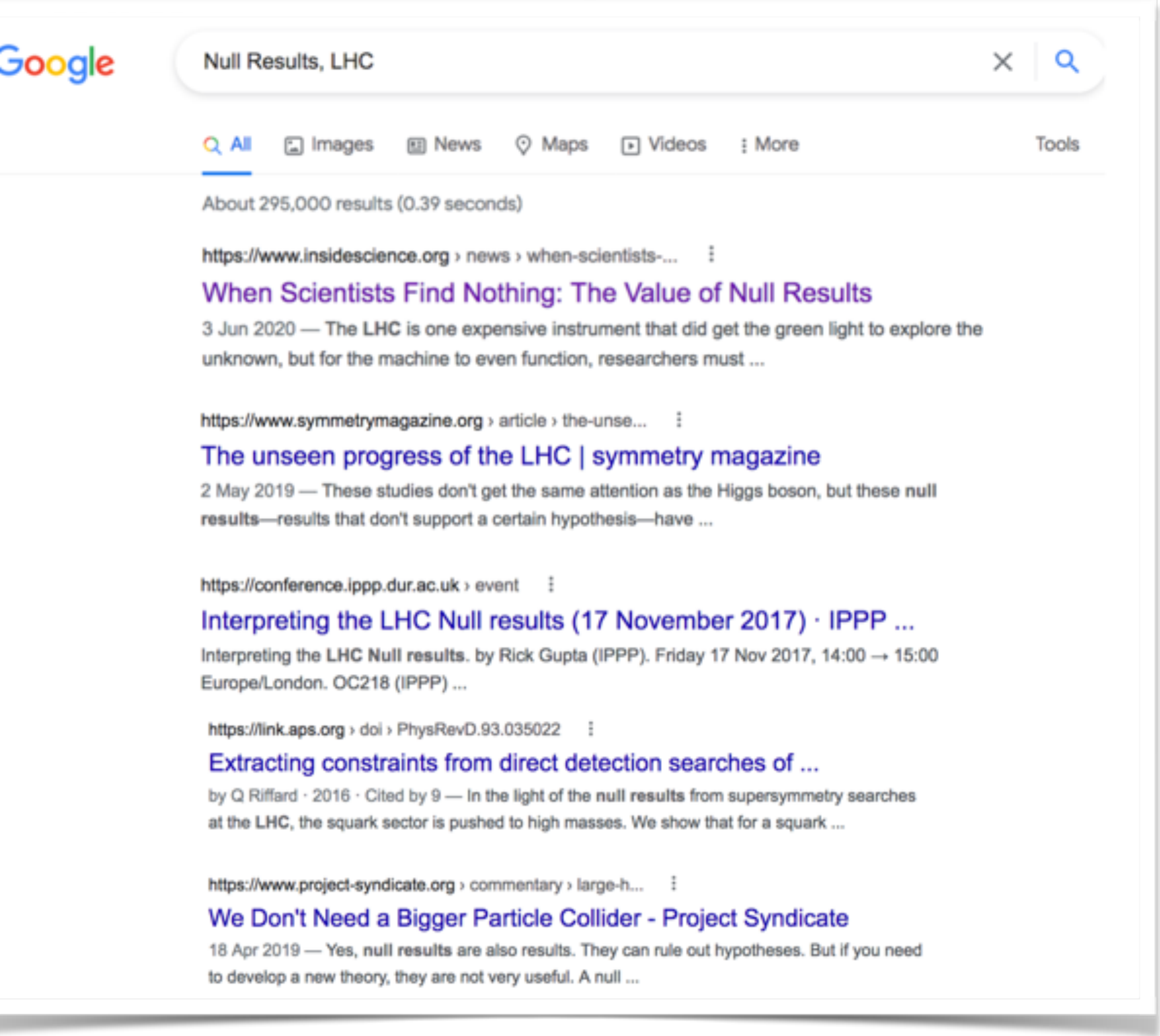
Muon g-2 Anomaly

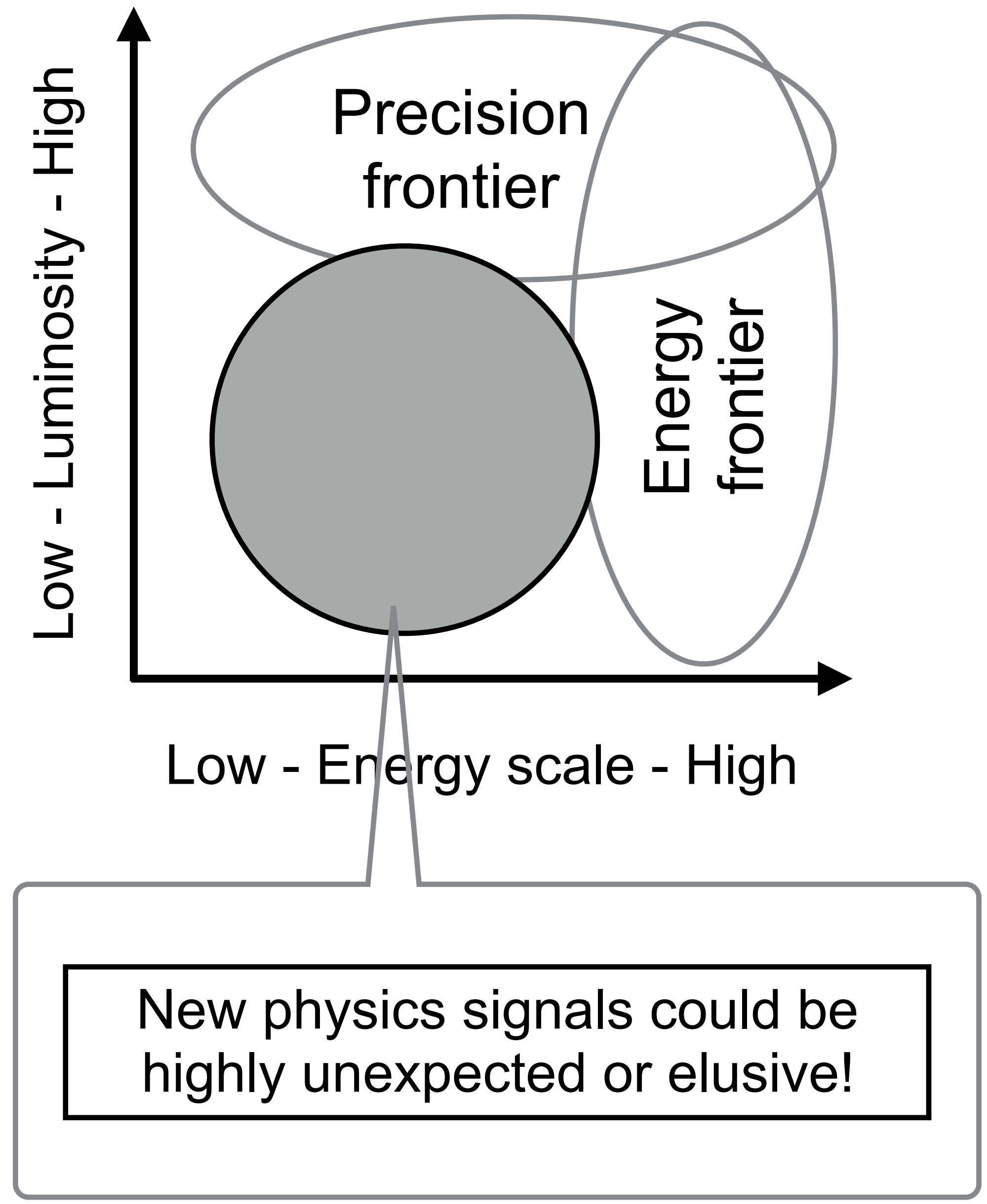
[PhysRevLett.126 (2021)]

CDFII W Mass Anomaly

[Science 376 (2022)]

Null Results for NP Direct Searches at LHC





Machine Learning at Colliders

- The High Energy Physics (HEP) community has a long history of using ML for data analysis
  - Neural network for top quark search @D0 (1990)
  - BDT was first used by MiniBooNe for neutrino data (2004)
- Novelty (anomaly) detection represents a new task of ML at colliders

What Is Novelty (Anomaly) Detection?

Signal Processing 99 (2014) 215–249

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

ELSEVIER

SIGNAL PROCESSING

Review

A review of novelty detection

CrossMark

Marco A.F. Pimentel *, David A. Clifton, Lei Clifton, Lionel Tarassenko

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

A R T I C L E I N F O

Article history:
Received 17 October 2012
Received in revised form
16 December 2013
Accepted 23 December 2013
Available online 2 January 2014

Keywords:
Novelty detection
One-class classification
Machine learning

A B S T R A C T

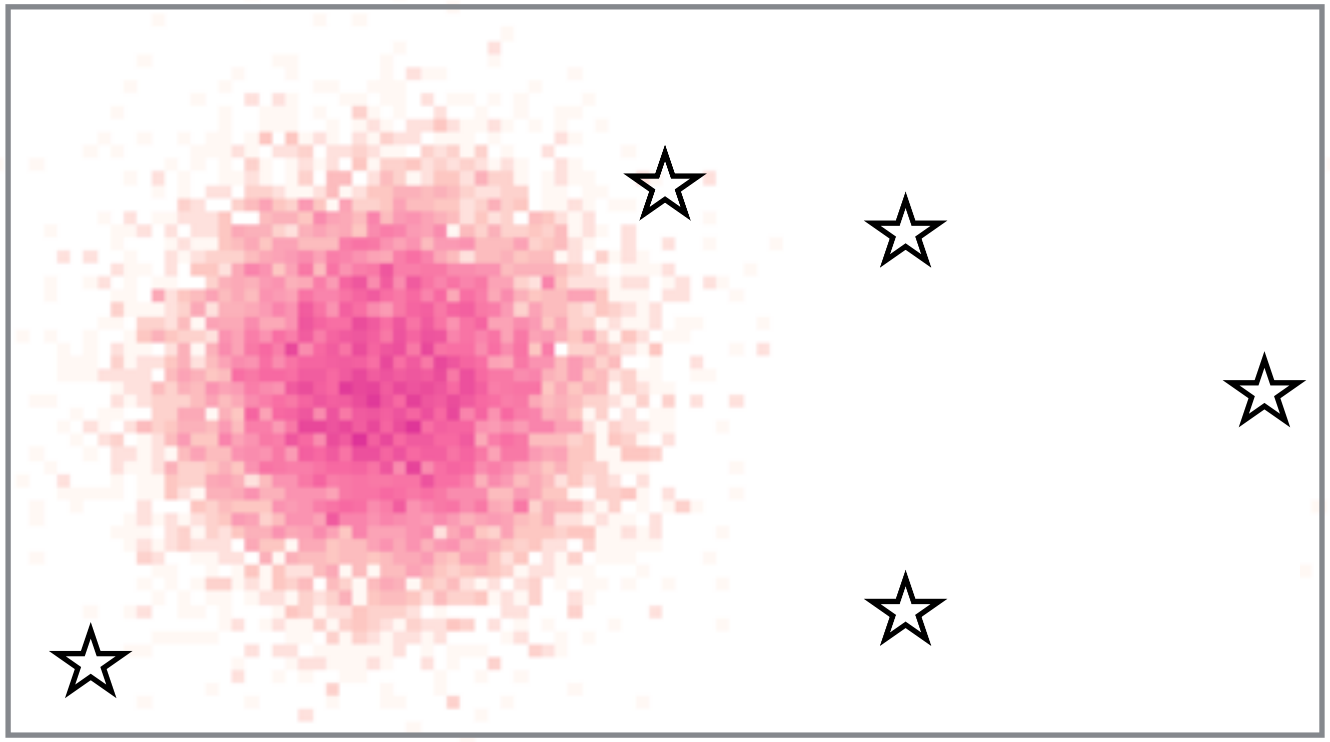
Novelty detection is the task of classifying test data that differ in some respect from the data that are available during training. This may be seen as "one-class classification", in which a model is constructed to describe "normal" training data. The novelty detection approach is typically used when the quantity of available "abnormal" data is insufficient to construct explicit models for non-normal classes. Application includes inference in datasets from critical systems, where the quantity of available normal data is very large, such that "normality" may be accurately modelled. In this review we aim to provide an updated and structured investigation of novelty detection research papers that have appeared in the machine learning literature during the last decade.

No prior knowledge on the signals is available for model training

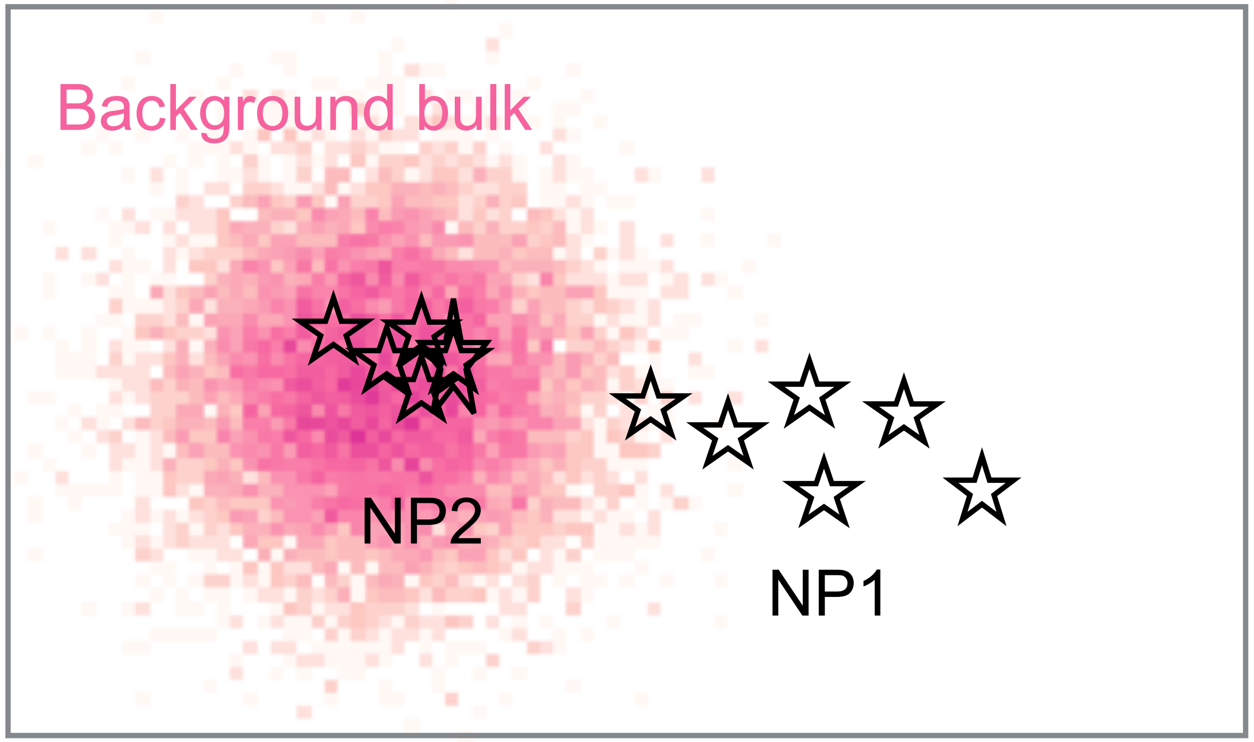
Applications in Our Daily Life

- Gate Control System (GCS): trained with personal data (face, fingerprint, iris, voice,…) of the residents/users of a building
- Is a visitor allowed to enter?
  - Recognized as a resident/user - Yes
  - Recognized as a stranger - No
- Many others: defect detection for materials, intrusion detection for cyber-security, event detection in sensor networks, health monitoring for people, etc.

Novel (Anomalous) Events at Colliders



``Novel (Anomalous) events''
evaluated by the GCS



``Novel (Anomalous) events''
evaluated at the collider

The ``usual'' novel (anomalous) events tend to be individual or unrelated, while the ones at colliders tend to be collective or clustered (QFT) and hence from a distribution. Individually, they may not be distinguishable from the backgrounds

=> To perform this task at colliders, dedicated designs for novelty evaluators are needed

First Efforts in This Direction… …

Weakly supervised learning

- [arXiv:1702.00414]: ``Weakly Supervised Classification in High Energy Physics'', L. Dery, B. Nachman, F. Rubbo and A. Schwartzman
- [arXiv:1706.09451]: ``(Machine) learning to do more with less'', T. Cohen, M. Freytsis and B. Ostdiek
- [arXiv:1708.02949]: ``Classification without labels: Learning from mixed samples in high energy physics'', E. M. Metodiev, B. Nachman, and J. Thaler
- [1805.02664]: ``Anomaly Detection for Resonant New Physics with Machine Learning'', J. Collins, K. Howe, B. Nachman
- [arXiv:1806.02350]: ``Learning New Physics from a Machine'', R. T. D'Agnolo and A. Wulzer
- [arXiv:1807.06038]: ``Guiding New Physics Searches with Unsupervised Learning'', De Simone and T. Jacques

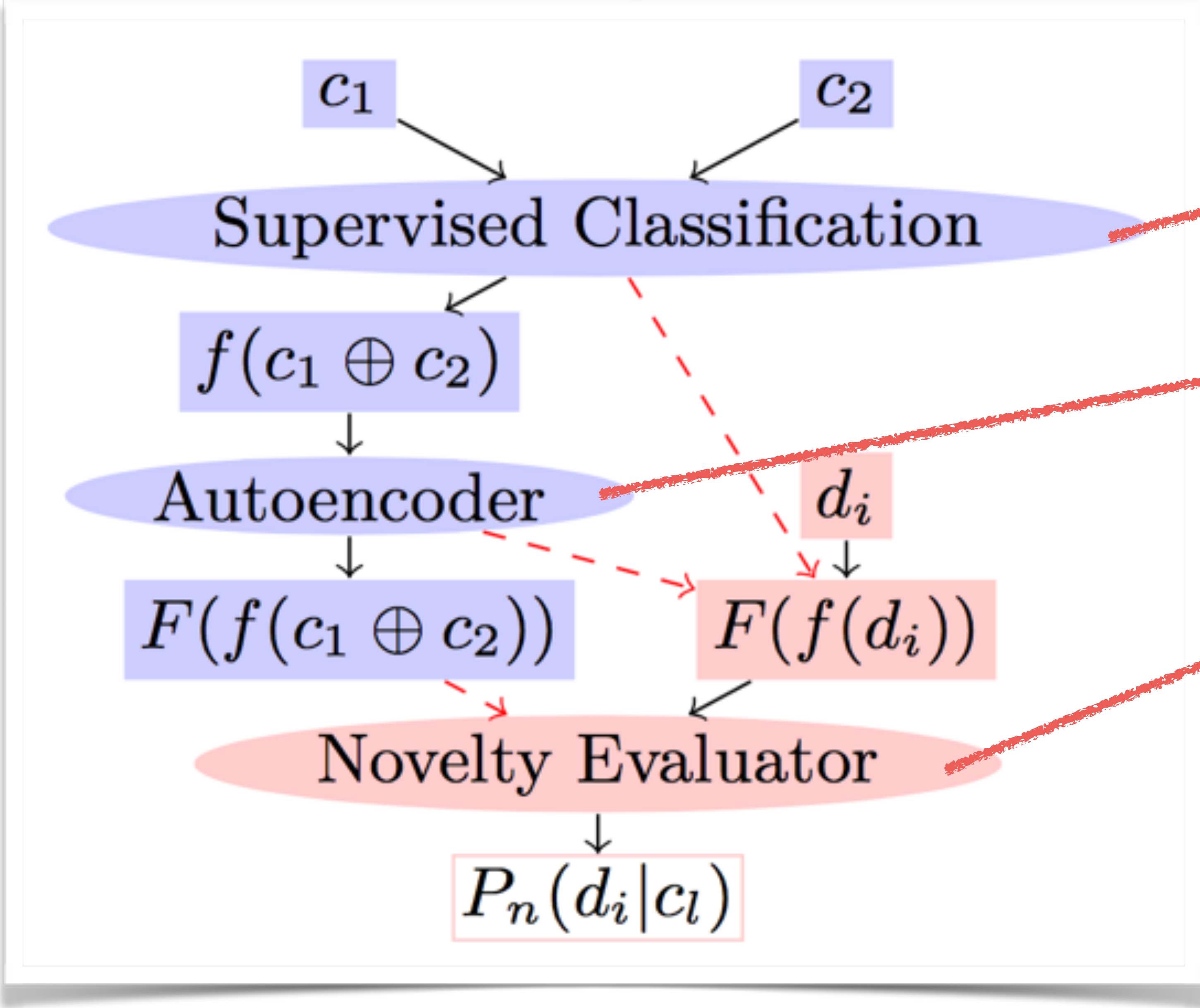
kNN method + evaluator complementarity

- [arXiv:1807.10261]: ``Novelty Detection Meets Collider Physics''; J. Hajer, Y.-Y. Li, TL, and H. Wang

AE reconstruction error

- [arXiv:1808.08992]: ``Searching for New Physics with Deep Autoencoders'', M. Farina, Y. Nakai, and D. Shih
- [arXiv:1808.08979]: ``QCD or What?'', T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson
- … …

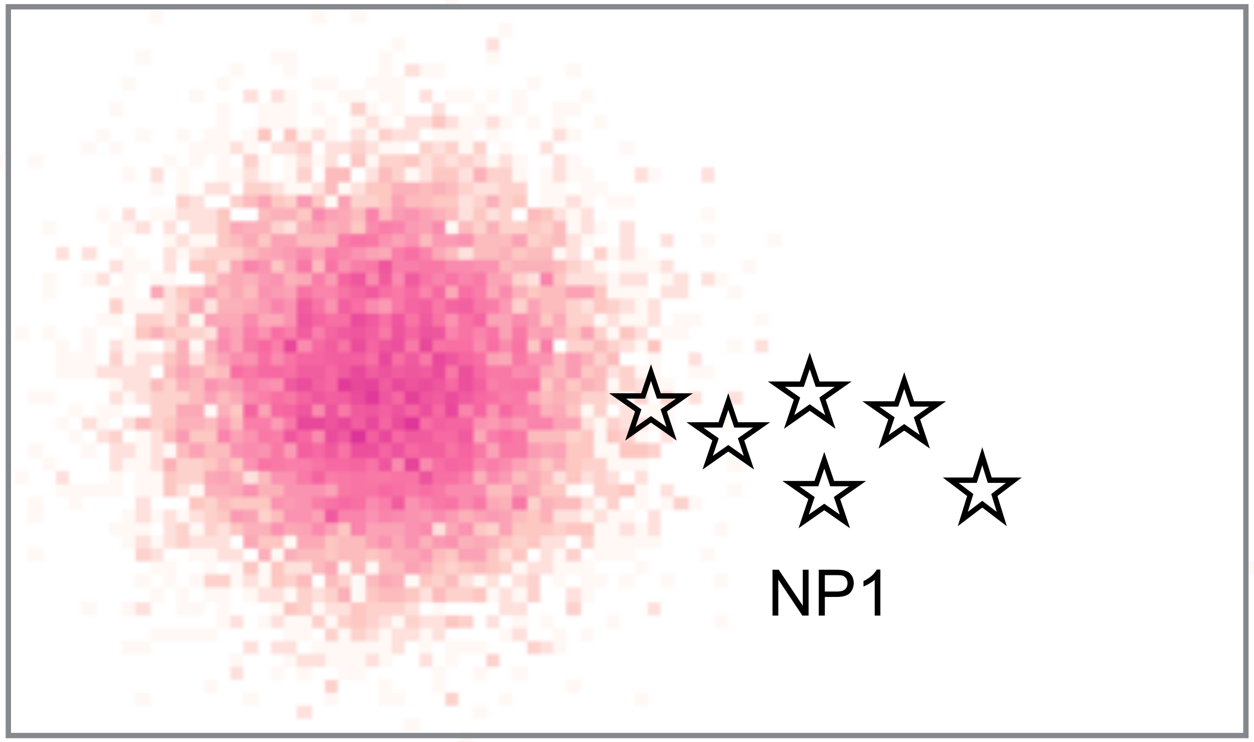
Workflow

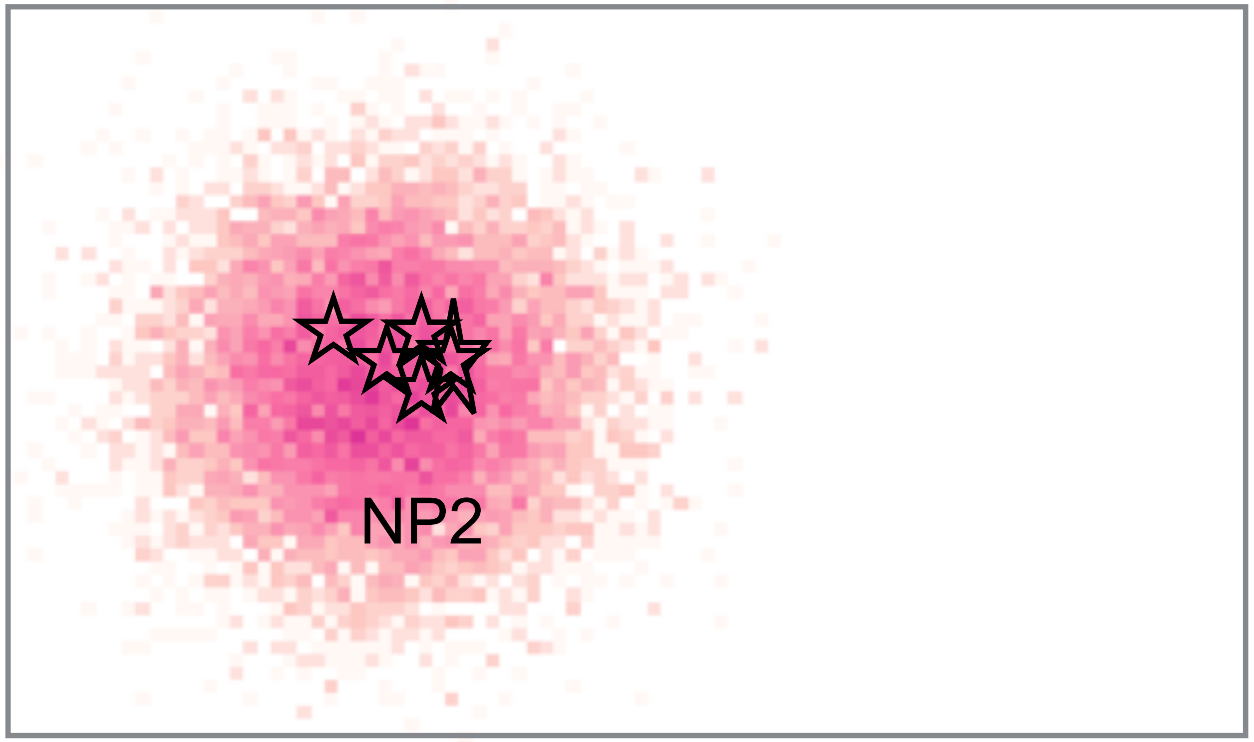


- Step 1: (SM/background) feature learning
- Step 2: dimensionality reduction of feature space (auto-encoder)
- Step 3: novelty evaluation of testing data
- Detection sensitivity can be analyzed based on novelty response of the testing sample

[arXiv:1807.10261; J. Hajer, Y.-Y. Li, TL, and H. Wang]

Two Classes of Evaluators





Isolation-based Evaluators

The novelty of a testing event is individually evaluated based on its distance to or isolation from the background bulk (known-pattern data) in the feature space.

Other testing events do not play a role in this process.

Clustering-based Evaluators

The novelty for a testing event is evaluated according to the event clustering/collective effect on top of the background bulk (known-pattern data) in the feature space.

Other testing events (especially the ones nearby) will contribute in this process.

Below we demonstrate using the k-Nearest-Neighbors (kNN) method (simple, intuitive and representative)

kNN-based O_iso

$$\Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'_{\text{train}} \rangle^{1/2}}$$

$$\mathcal{O} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$

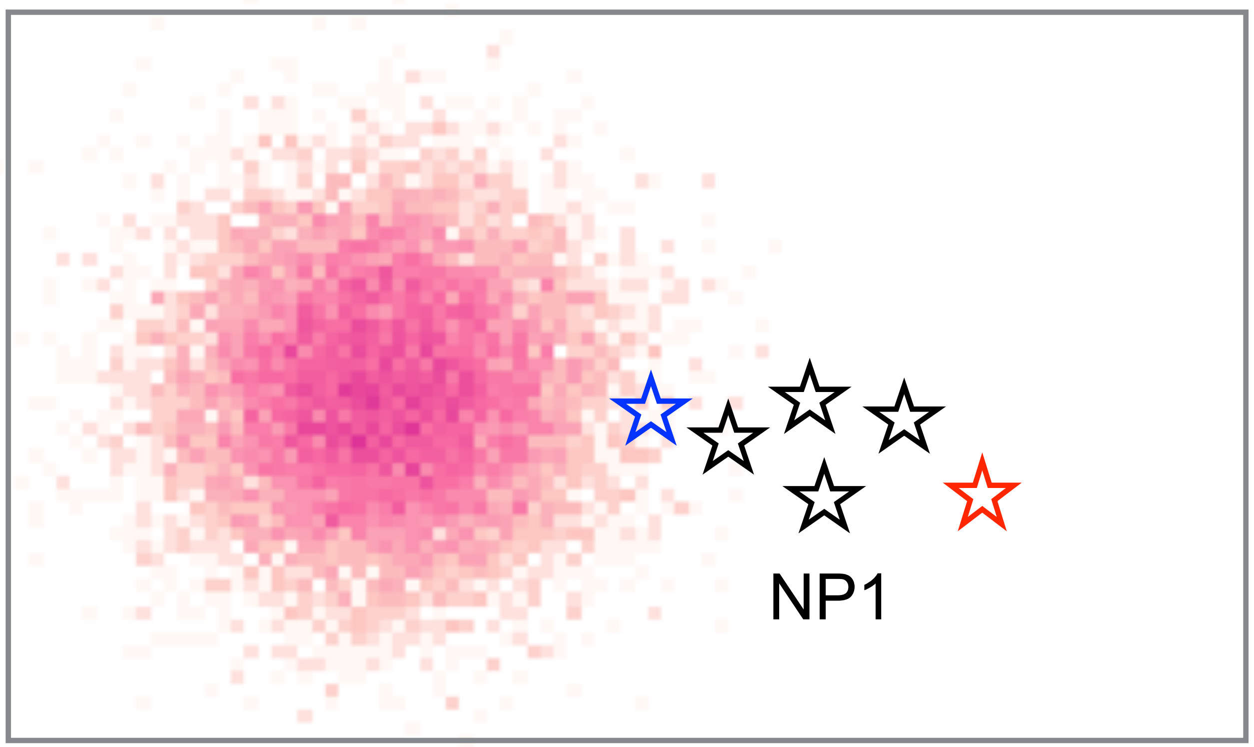
Novelty measure: not normalized

Novelty evaluator: $0 \leq \mathcal{O} \leq 1$

- d_{train} : mean distance of a testing data point to its k nearest neighbors in the training sample (physical distance)
- $\langle d'_{\text{train}} \rangle$: average of the mean distances of its k nearest neighbors to their respective k nearest neighbors in the training sample (reference distance) :
- Novelty response is evaluated by comparing the physical distance of the testing event and the reference distance.

kNN-based O_iso

$$\Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'_{\text{train}} \rangle^{1/2}} \qquad \mathcal{O} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$



Tends to be scored low

Tends to be scored high

kNN-based O_clu

$$\Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'_{\text{train}} \rangle^{1/2}} \qquad \Delta_{\text{clu}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

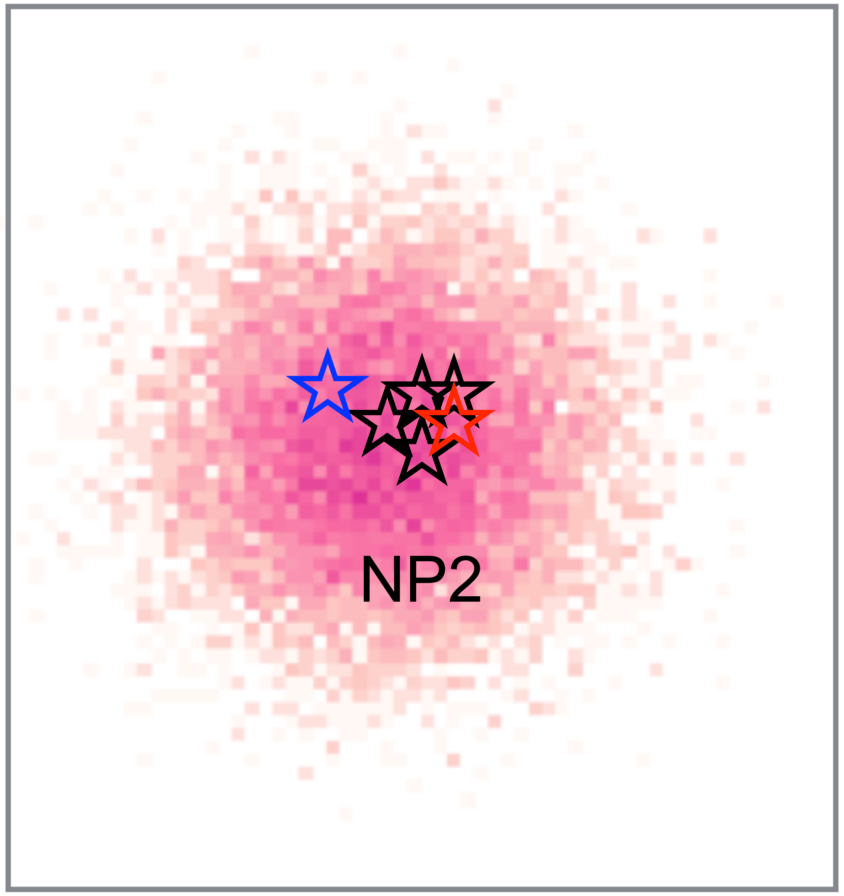
- d_{train} : mean distance of a testing event to its k nearest neighbors in the training dataset
- d_{test} : mean distance of a testing event to its k nearest neighbors in the testing dataset
- m: dimension of the feature space => d^-m: a measure of local density
- Novelty response is evaluated by comparing local densities of the testing event defined by the training and testing datasets

kNN-based O_clu

$$\Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'_{\text{train}} \rangle^{1/2}} \qquad \Delta_{\text{clu}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

Training dataset

VS

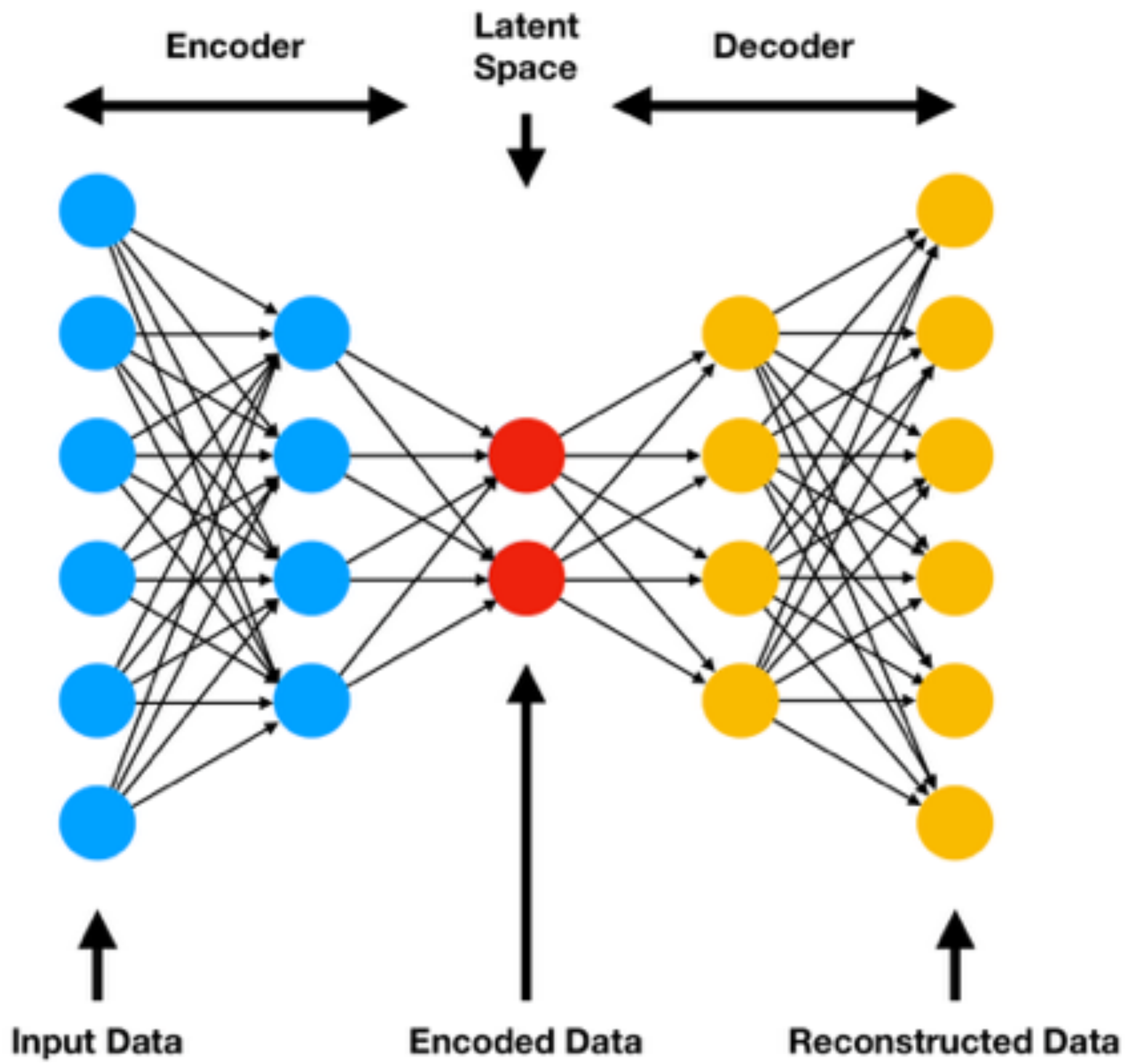


Testing dataset

☆ Tends to be scored low

☆ Tends to be scored high

Reconstruction Error of AE: Isolation-Based

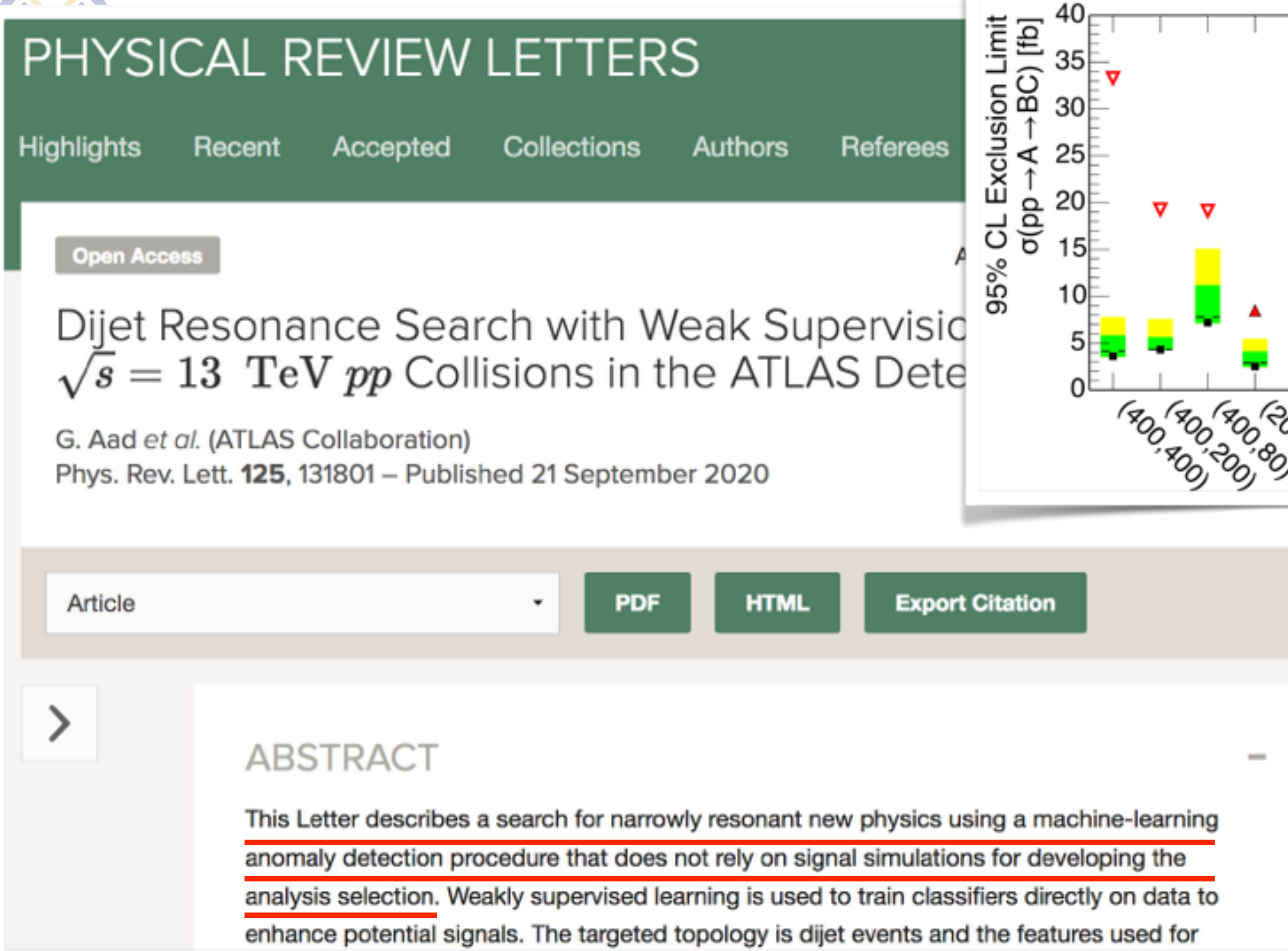


$$L = |x - x'|^2$$

[T. Heimel et. al.; M. Farina et. al; 2018]

- Autoencoder: a special NN with its latent space being defined with a demand of minimizing event reconstruction error
- Reconstruction error as a measure of distance to/difference from the known-pattern data
- Novelty is evaluated individually for the testing events => Isolation-based!

Weakly-supervised Learning



PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees

Open Access

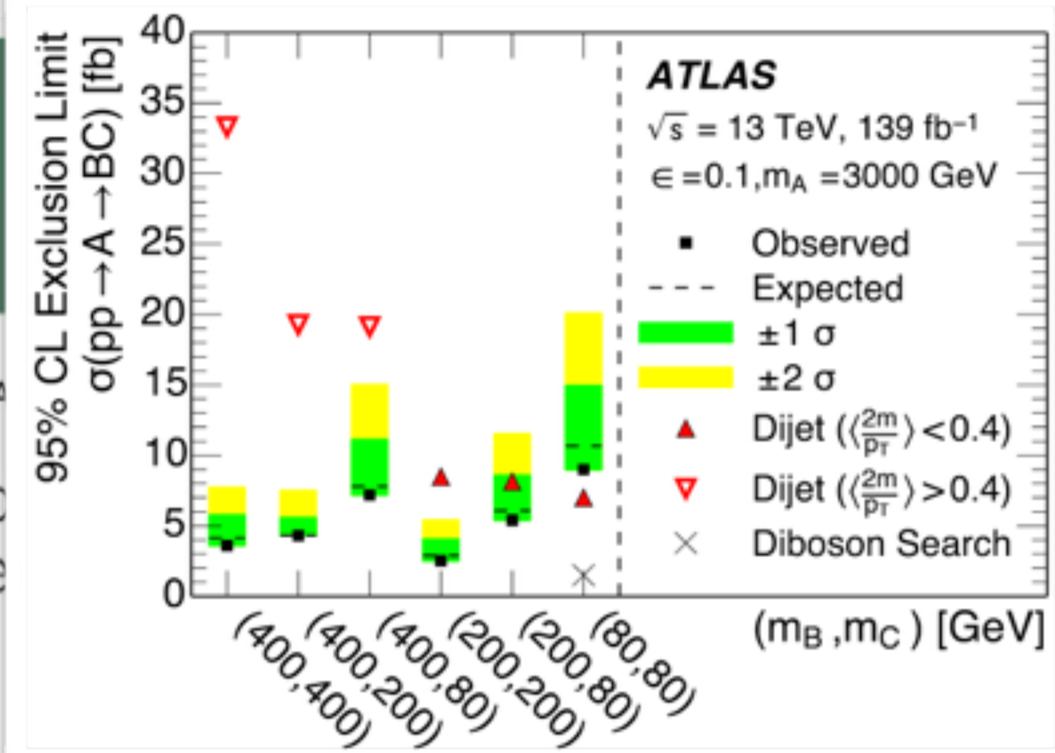
Dijet Resonance Search with Weak Supervisio
$\sqrt{s} = 13$ TeV pp Collisions in the ATLAS Dete

G. Aad *et al.* (ATLAS Collaboration)
Phys. Rev. Lett. **125**, 131801 – Published 21 September 2020

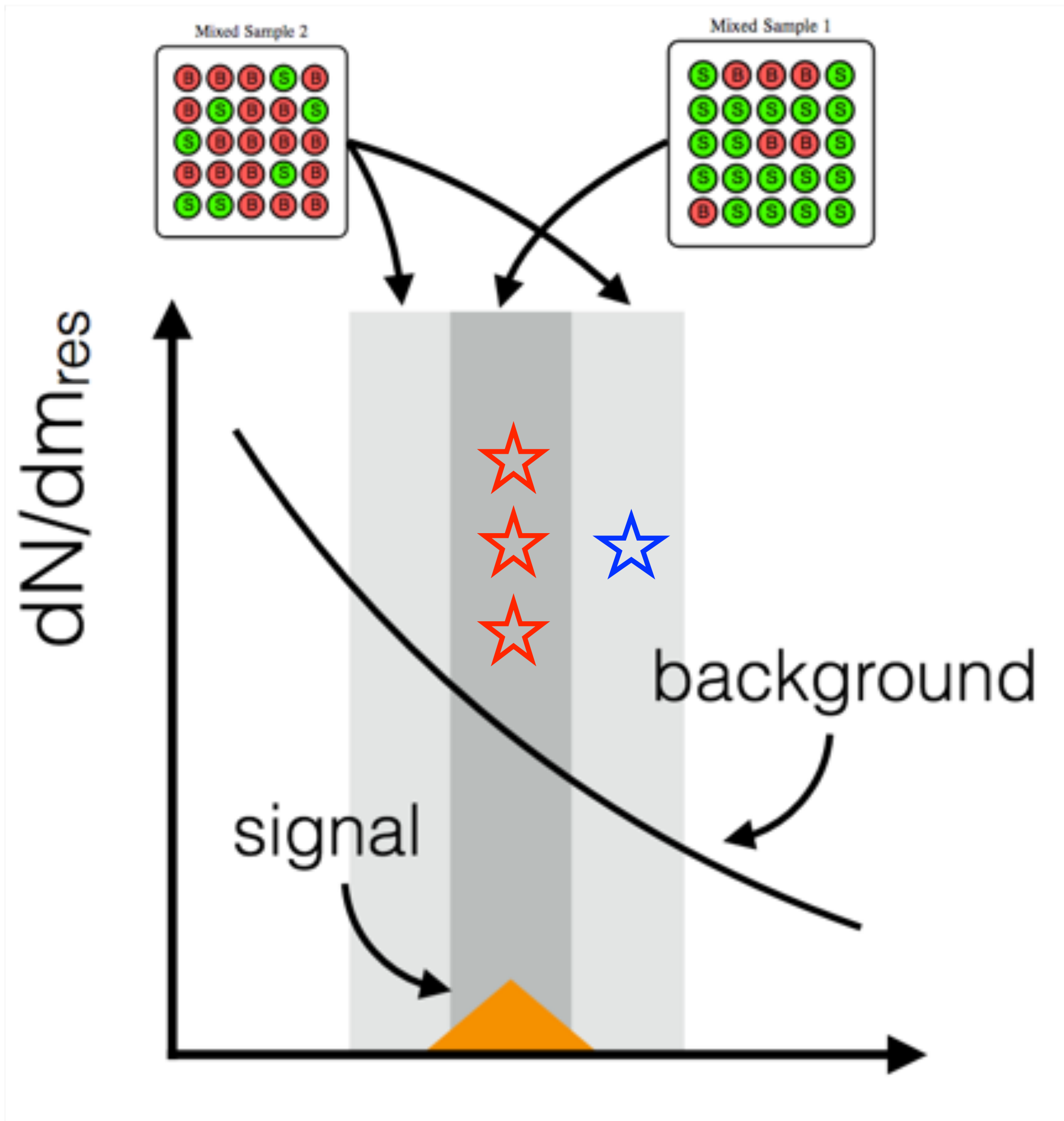
Article PDF HTML Export Citation

ABSTRACT

This Letter describes a search for narrowly resonant new physics using a machine-learning anomaly detection procedure that does not rely on signal simulations for developing the analysis selection. Weakly supervised learning is used to train classifiers directly on data to enhance potential signals. The targeted topology is dijet events and the features used for



Weakly-supervised Learning: Clustering-Based



[Collins, Howe, Nachman, 1805.02664]

- Assign labels to each event in the mixed samples (0 => M0; 1 => M1) and then performs supervised learning
- S0/B0 =/ S1/B1: optimize S0/B0 vs S1/B1 <=> optimize S vs B
- If S0/B0 => 0, S1/B1 => 1, then reduced to fully supervising learning
- Novelty is evaluated essentially based on local density => clustering-based

☆ Tends to be scored low

☆ Tends to be scored high

More Novelty Evaluators

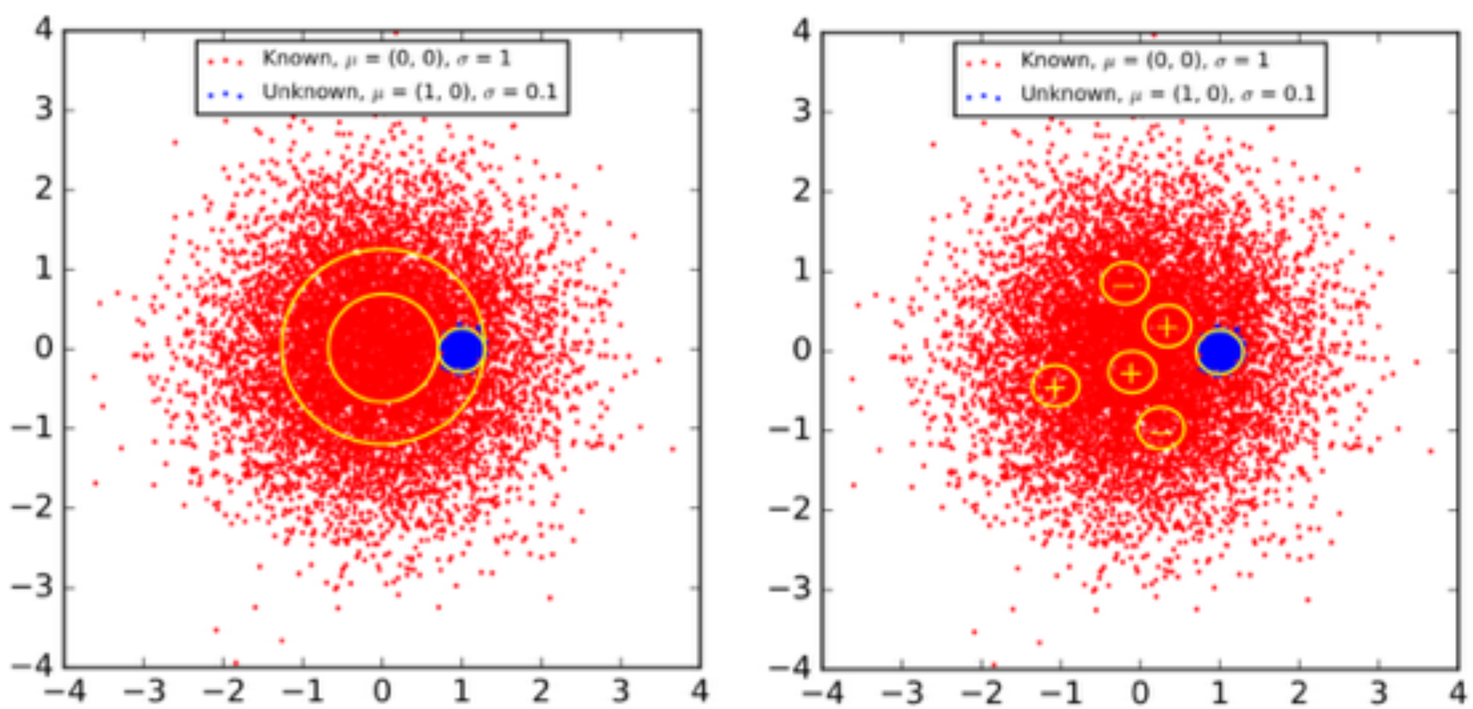
[arXiv:2202.02165; X.-H. Jiang, A. Juste, Y.-Y. Li, TL]

$\mathcal{O}_{\rm iso}$	k-nearest-neighbors(kNN)-based $\mathcal{O}_{\rm iso}$ [1]
	Autoencoder(AE)-based [12–21]
	Graph [22], classical k-means clustering [23]
$\mathcal{O}_{\rm clu}$	kNN-based $\mathcal{O}_{\rm clu}$ [1], TS [24]
	t-score [2, 25], ANODE [26], Poissonian Mixture Model [27]
	CWoLa [28–31], TNT [32], SALAD [33]
	SULU [34]
	UCluster [35]

Designed in different manners, but to be either isolation-based or clustering-based

Why Synergize O_iso and O_clu?

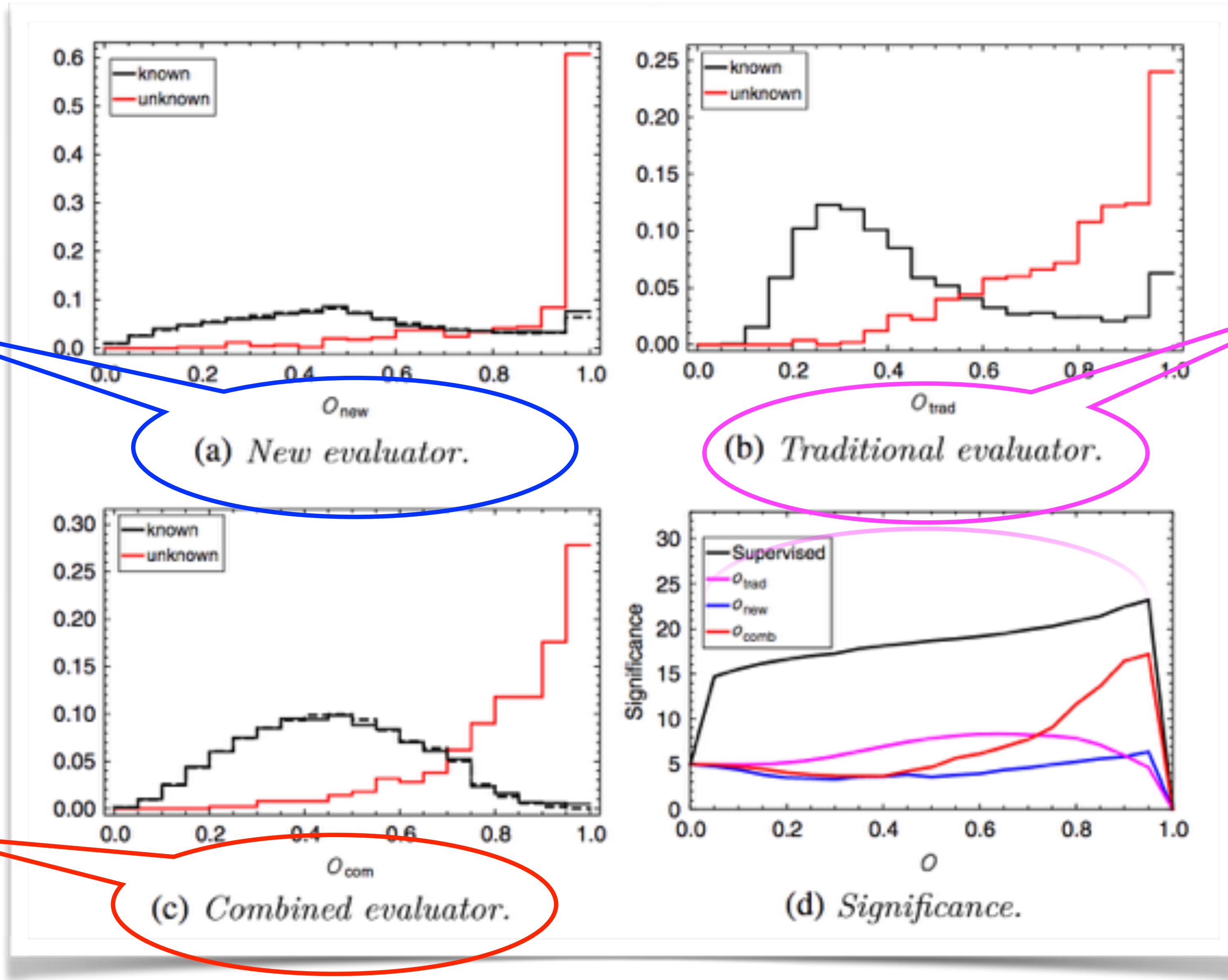
- Sensitive to the signals of different patterns
  - Isolation-based: signal events far away from the background bulk
  - Clustering-based: signal events creating a difference in the local density
- Subject to the reducible backgrounds of different sources
  - Isolation-based: events equidistant from the background bulk
  - Clustering based: statistical fluctuations
- Can broaden the coverage over signal patterns and suppress reducible backgrounds



Synergizing O_iso and O_clu - First Effort

kNN-based O_clu

kNN-based O_iso



kNN-based O_syn

$$\mathcal{O}_{\rm syn} = f(\mathcal{O}_{\rm iso}, \mathcal{O}_{\rm clu}) = \sqrt{\mathcal{O}_{\rm iso}\mathcal{O}_{\rm clu}}$$

[arXiv:1807.10261; J. Hajer, Y.-Y. Li, TL, and H. Wang]

Synergizing O_iso and O_clu - More Systematic Treatment

arXiv > hep-ph > arXiv:2202.02165

Search...

Help | Advance

High Energy Physics – Phenomenology

[Submitted on 4 Feb 2022]

Detecting New Physics as Novelty -- Complementarity Matters

Xu-Hui Jiang, Aurelio Juste, Ying-Ying Li, Tao Liu

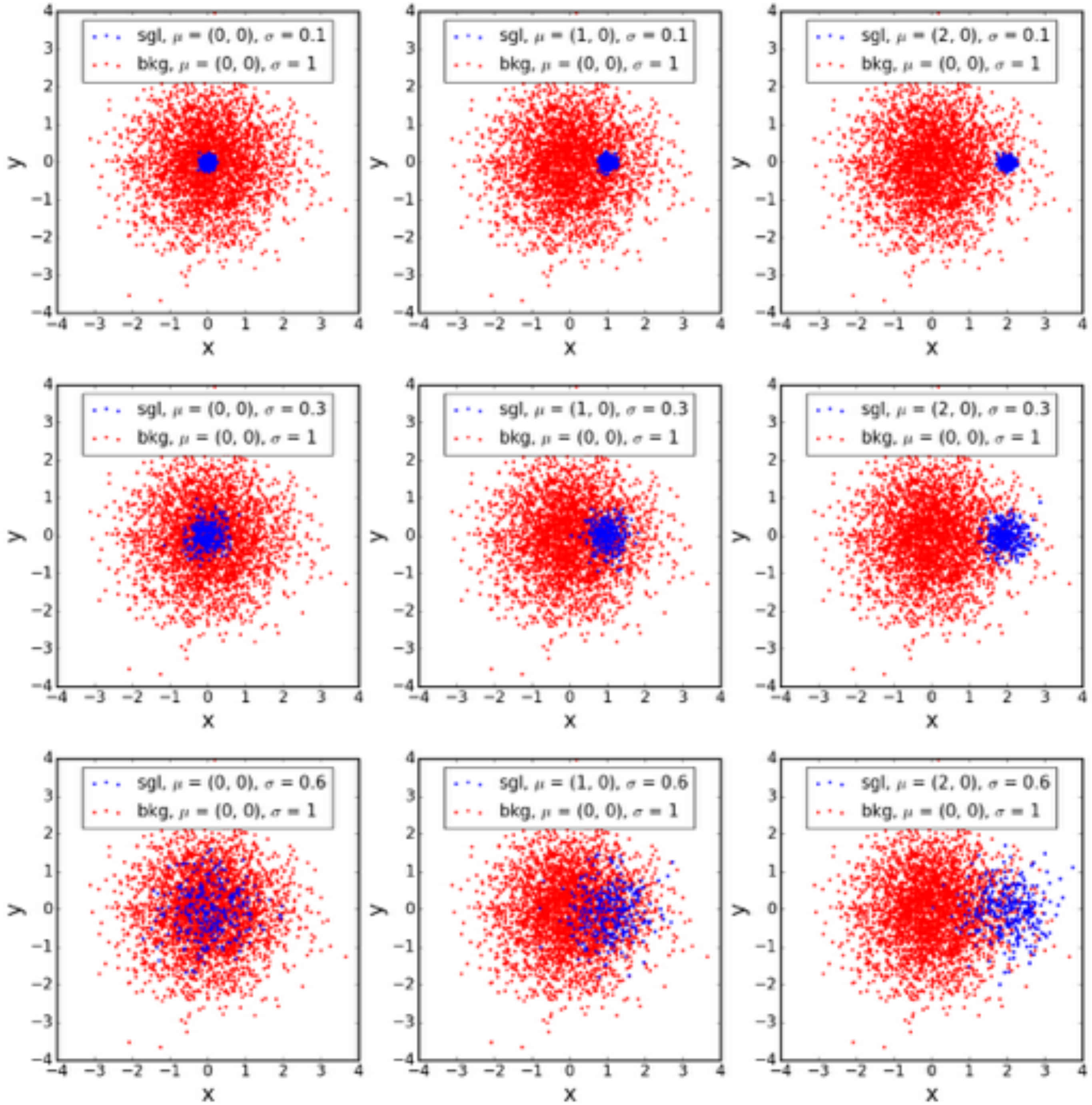
Novelty detection is a task of machine learning that aims at detecting novel events without a prior knowledge. In particular, its techniques can be applied to detect unexpected signals from new phenomena at colliders. In this paper, we develop an analysis scheme that exploits the complementarity, originally studied in Ref.~\cite{Hajer:2018kqm}, between isolation-based and clustering-based novelty evaluators. This approach can significantly improve the performance and overall applicability of novelty detection at colliders, which we demonstrate using a variety of two dimensional Gaussian samples mimicking collider events.

2D Gaussian Benchmarks

$R1 : \sigma = 0.1$

$R2 : \sigma = 0.3$

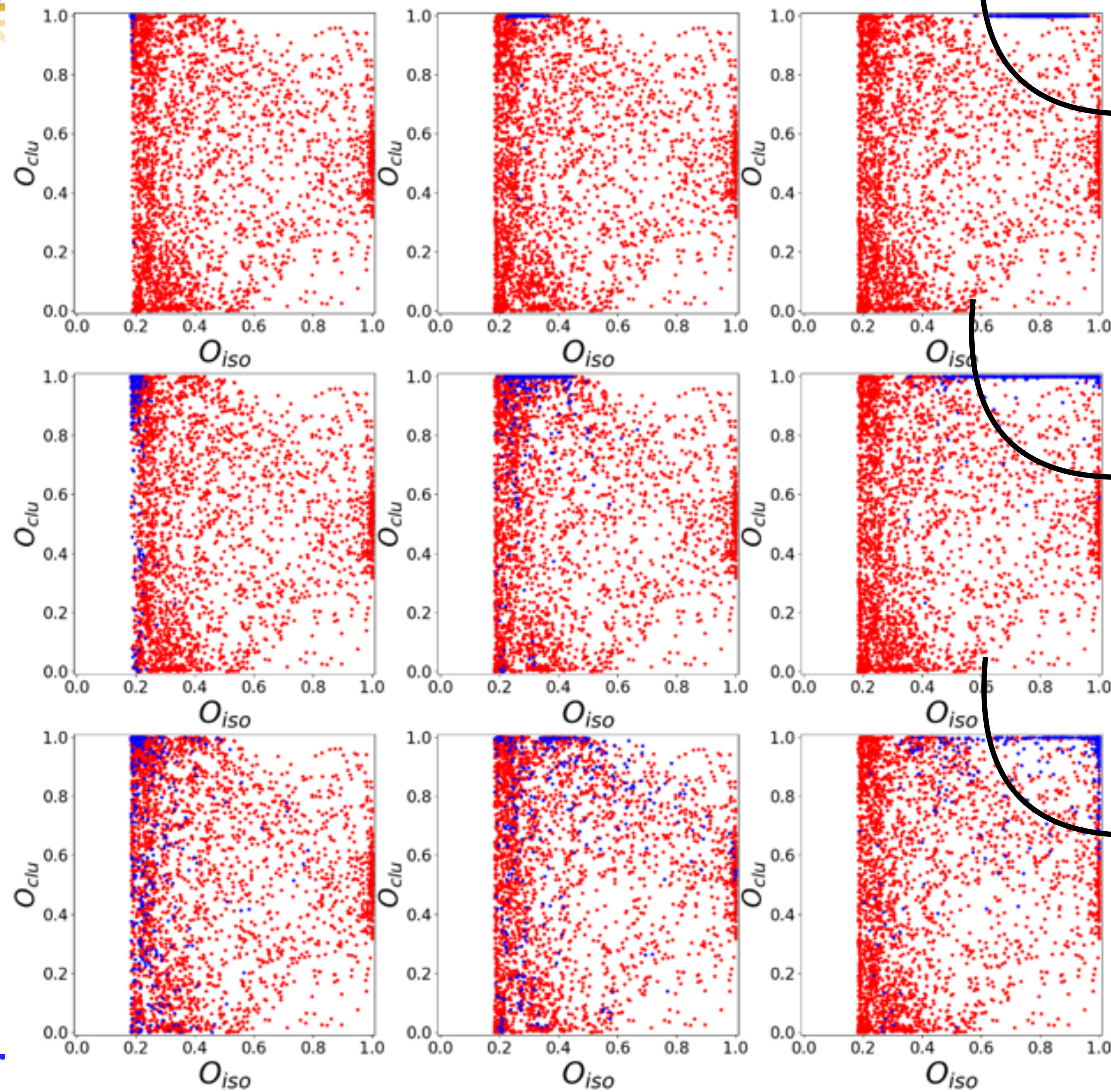
$R3 : \sigma = 0.6$



$C1 : \mu = (0, 0)$ $\quad C2 : \mu = (1, 0)$ $\quad C3 : \mu = (2, 0)$

$$B = 10000, S = 1000$$

Complementarity Between O_iso and O_clu



- Signals: sensitive to different patterns

  O_iso: C1 < C2 < C3

  O_clu: R3 < R2 < R1

- Reducible backgrounds: mutually suppressed to various extents

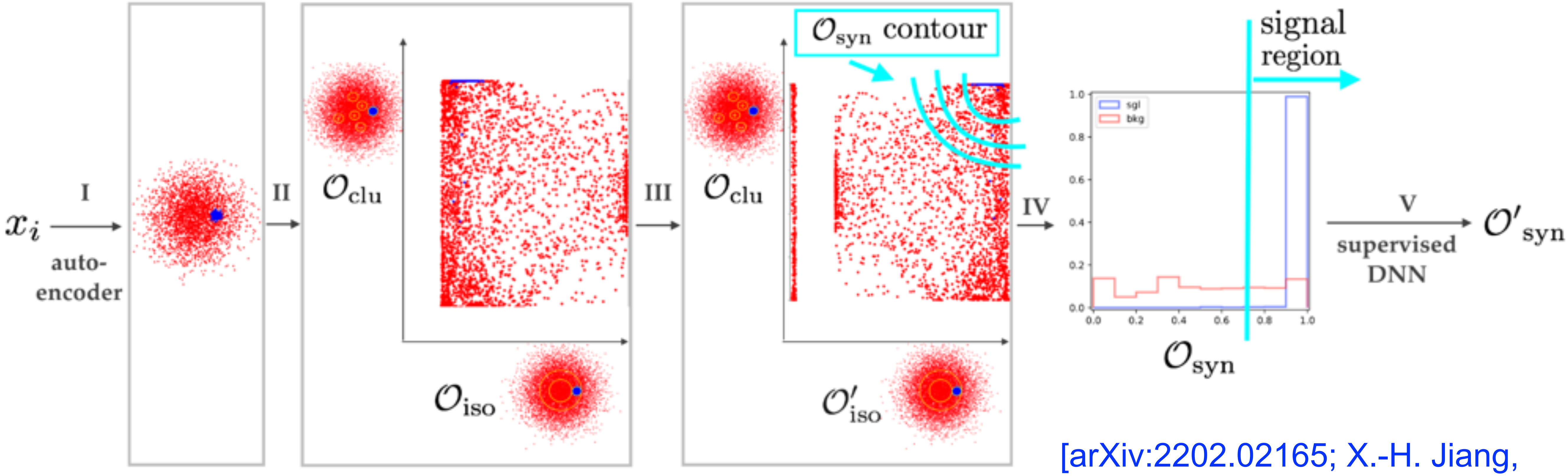
- O_syn works well for certain cases (right column) only

$$\mathcal{O}_{\mathrm{syn}} = \sqrt{\mathcal{O}_{\mathrm{iso}}\mathcal{O}_{\mathrm{clu}}}$$

Question:

How to recognize well the signal events in a general context, based on the O_iso and O_clu evaluations?

Improved Workflow for Novelty Detection

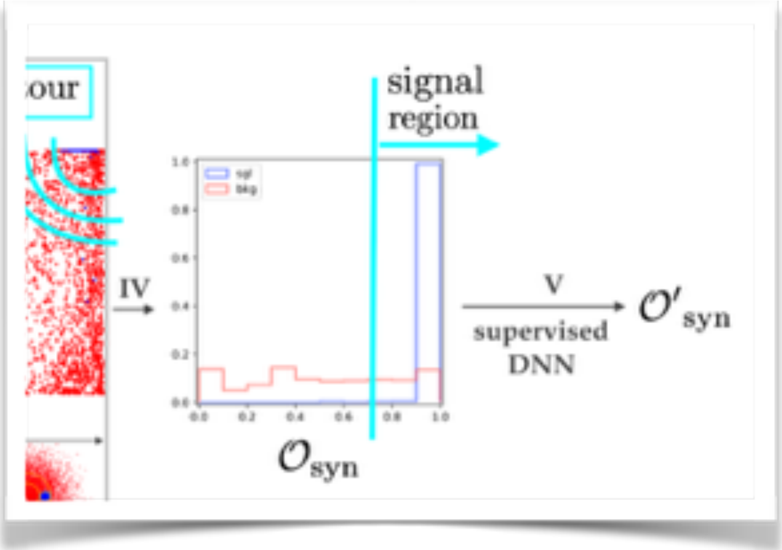


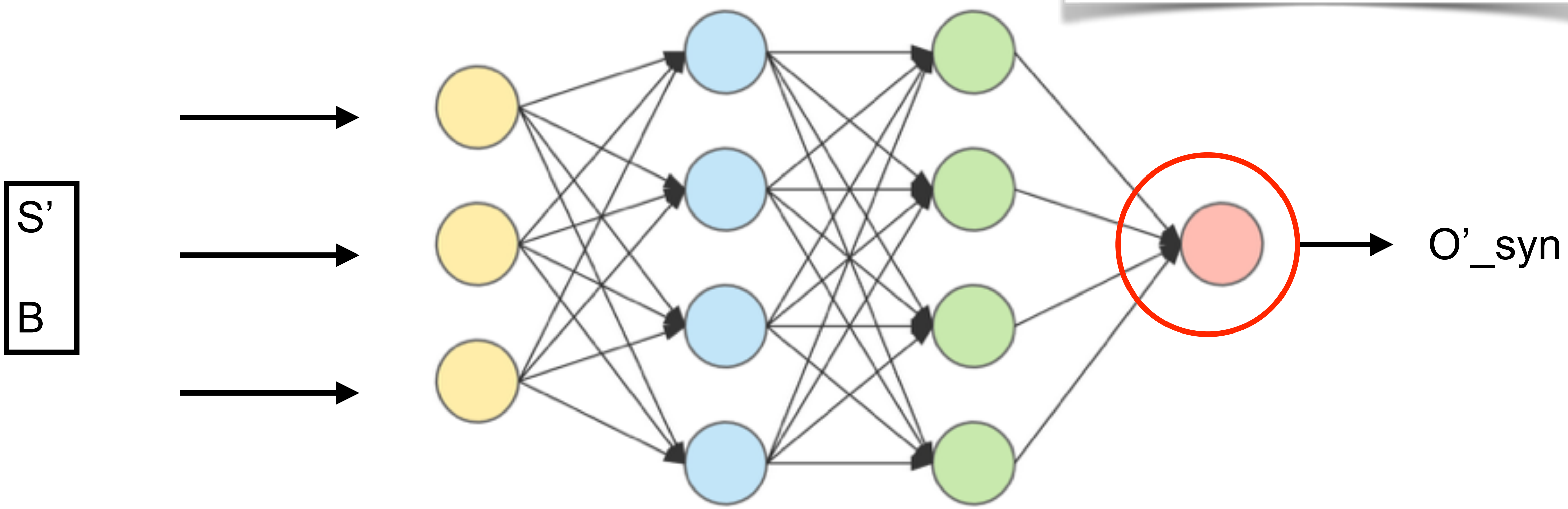
[arXiv:2202.02165; X.-H. Jiang, A. Juste, Y.-Y. Li, TL]

- Step III: bin resorting of O_iso => O'_iso
- Step IV: signal-like region identifying: $\mathcal{O}_{\rm syn} = \sqrt{\mathcal{O}'_{\rm iso}\mathcal{O}_{\rm clu}}$
- Step V: novelty re-evaluating (using the DNN score of weakly supervised learning as O'_syn), as a further optimization

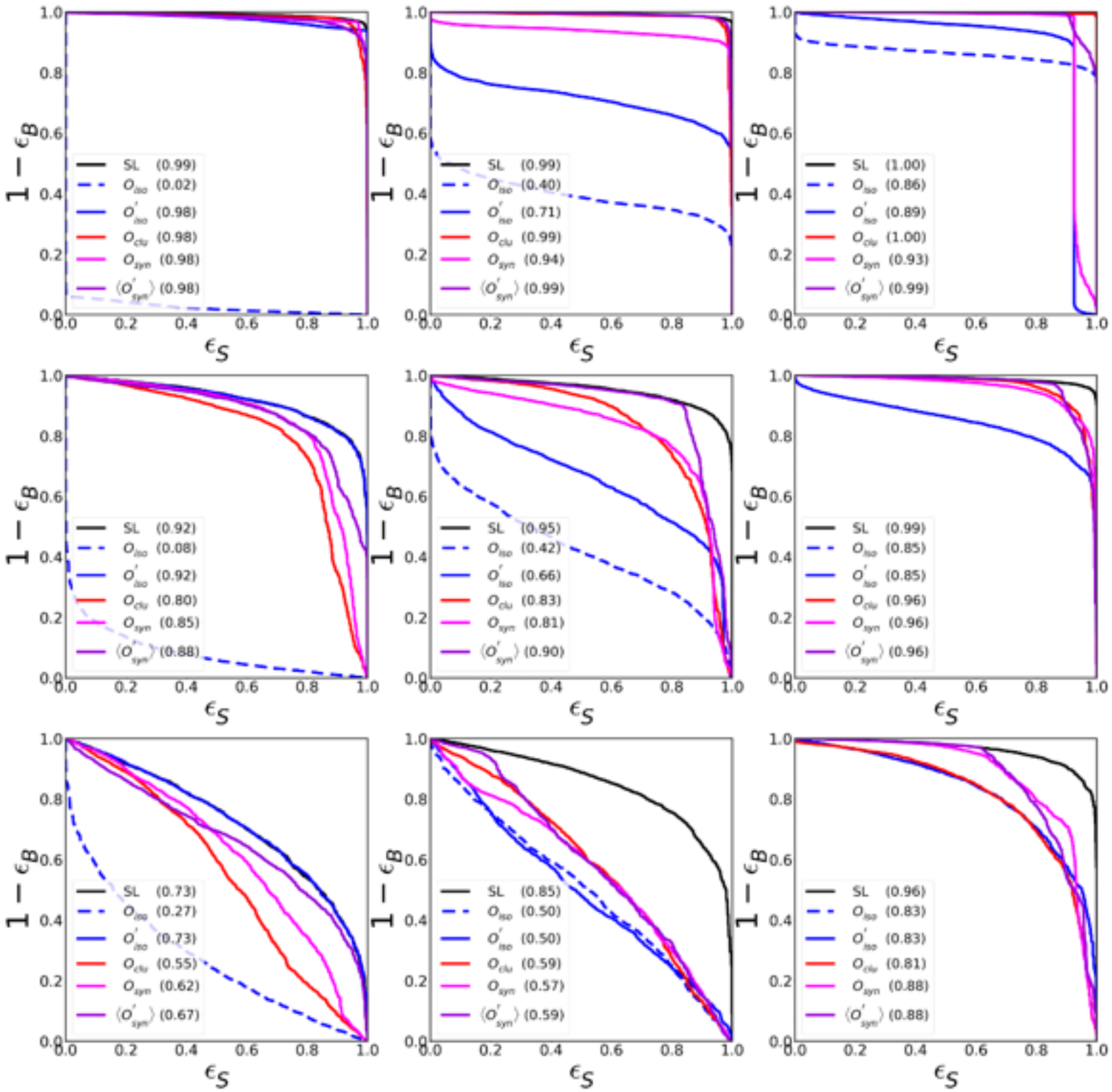
Synergy-based Evaluator O'_syn

- Build up a new NN of supervised learning, and use its output neuron as a synergy-based evaluator (O'_syn)
  - Signal-like (S'): O_syn > r
  - Background (B): from simulation





ROC Curves and AUC Values



kNN-based O_iso

kNN-based O_clu

kNN-based O_syn

kNN-based O'_syn

Supervised (ref)

- O_syn: performs universally better than O_iso or O_clu or both
- O'_syn: performs the best or among the best in almost all cases
- The sensitivity gap between O'_syn and SV is small (except for BP 8)

Generalization of Our Analysis Scheme

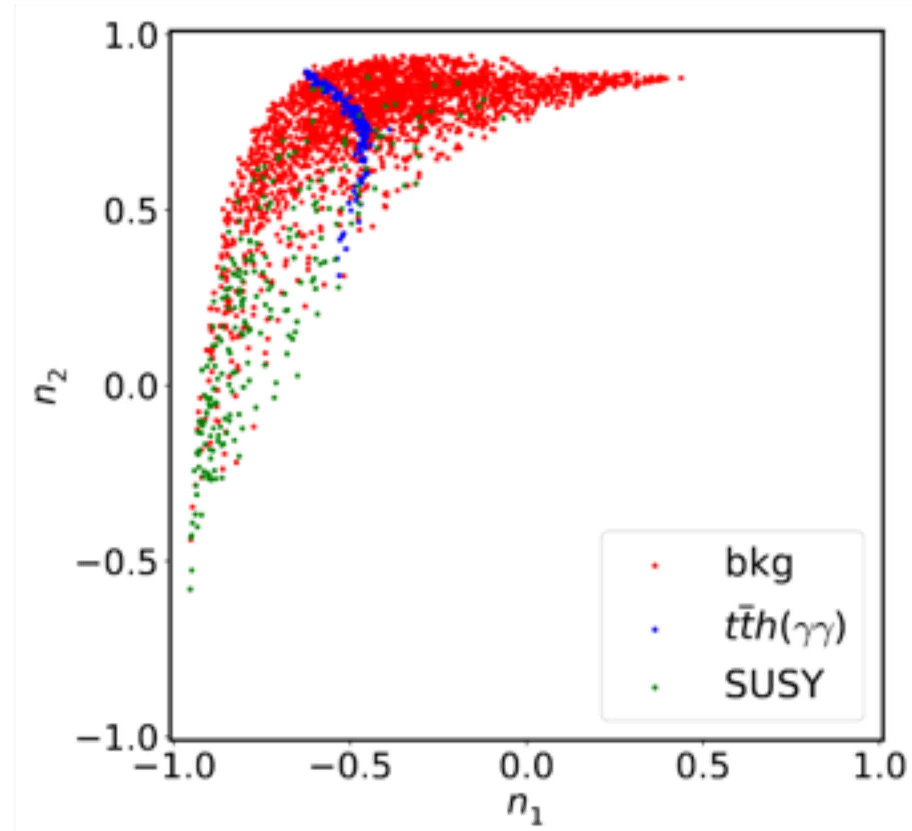
$\mathcal{O}_{\rm iso}$	k-nearest-neighbors(kNN)-based $\mathcal{O}_{\rm iso}$ [1]
	Autoencoder(AE)-based [12–21]
	Graph [22], classical k-means clustering [23]
$\mathcal{O}_{\rm clu}$	kNN-based $\mathcal{O}_{\rm clu}$ [1], TS [24]
	t-score [2, 25], ANODE [26], Poissonian Mixture Model [27]
	CWoLa [28–31], TNT [32], SALAD [33]
	SULU [34]
	UCluster [35]

The proposed analysis scheme is very general (although it is demonstrated using the kNN-based evaluators):

- one can pair any of the isolation-based and clustering-based evaluators to define O_syn and O'_syn, with the expectation of similar improvement for detection sensitivities
- one can even develop a clustering-based "partner" evaluator for each isolation-based evaluator, as it occurs to the kNN-based designs, and then embed them into this scheme (see the backups for such a realization using AE reconstruction error)

Application to tt+diphoton at LHC

	Process
Backgrounds	$t\bar{t}\gamma\gamma$
	$t\bar{t}\gamma$
	$t\bar{t}$
	Continuum $\gamma\gamma$
$t\bar{t}h$	$t\bar{t}h(\gamma\gamma)$
SUSY	$\tilde{t}\bar{\tilde{t}} \to t\bar{t}\gamma\gamma + 2\tilde{G}$



2D latent space of AE

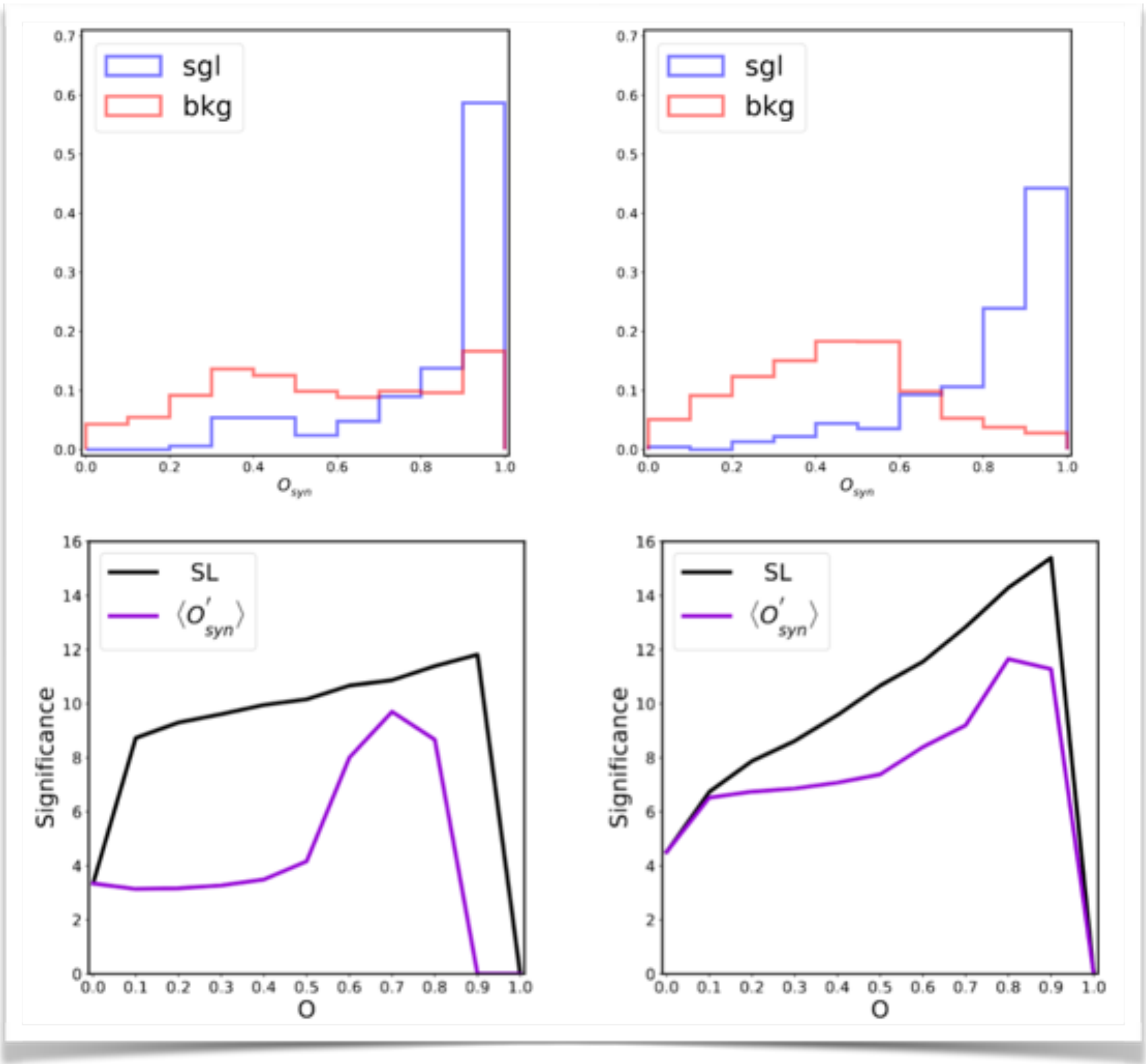
Two types of ``NP'': the same final state, but different signal patterns

- tth: resonance
- SUSY: broad distribution

[arXiv:2202.02165; X.-H. Jiang, A. Juste, Y.-Y. Li, TL]

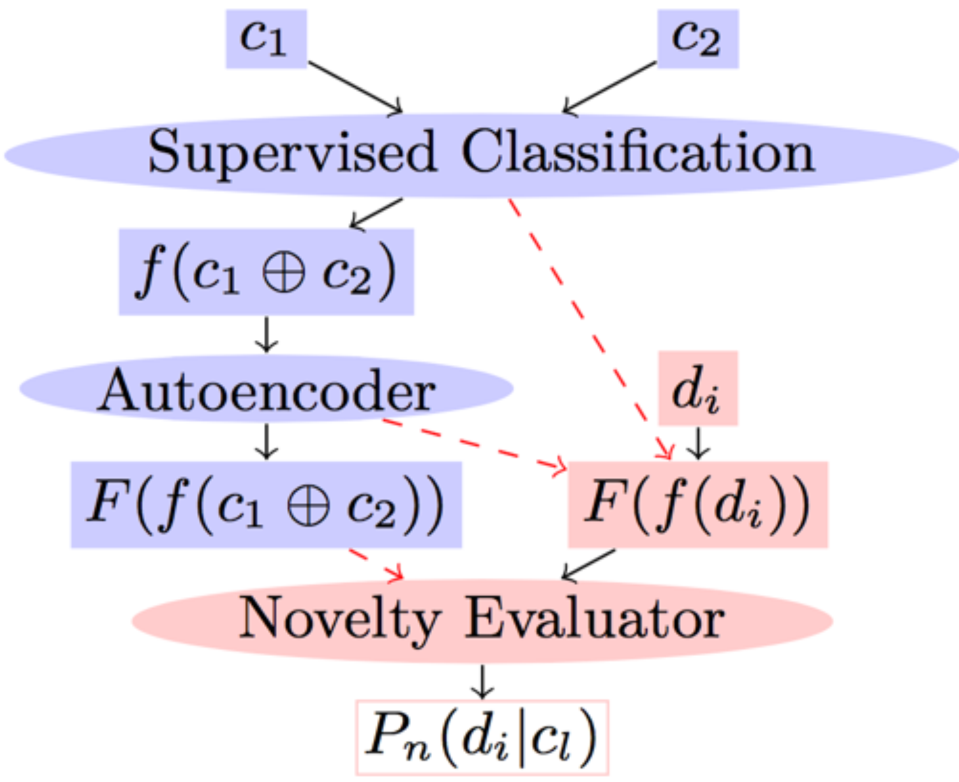
=> Good for testing the proposed analysis scheme

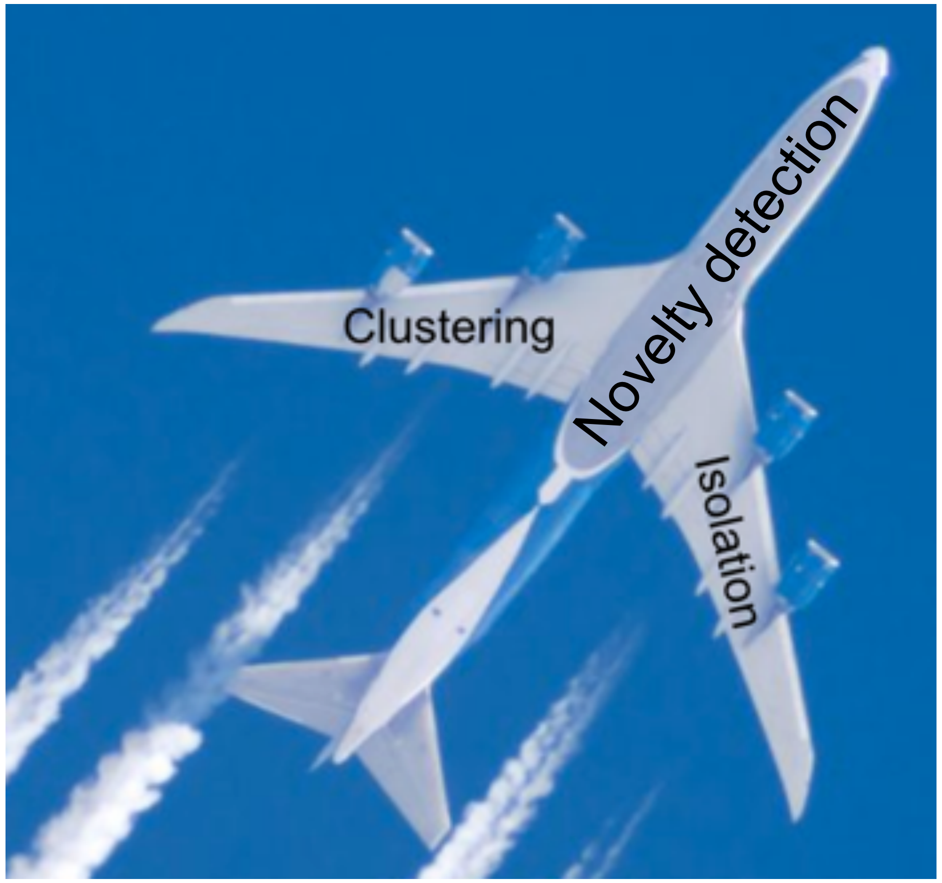
Application to tt+diphoton at LHC



tth SUSY

Take-home Messages!





- Searching for highly unexpected/elusive NP signals strongly motivates ML novelty (anomaly) detection at colliders
- Novelty evaluators (despite their so-many proposals) are mostly designed as being isolation-based or clustering-based
- Complementarity generically exists between O_iso and O_clu (in terms of sensitive signal patterns and reducible backgrounds)
- Synergy-based evaluators/methods can bring us to farther, …

UGC 大學教育資助委員會 University Grants Committee

GRF under grant No. 16312716

GRF under grant No. 16302117

GRF under grant No. 16304315

Thank you!

Backup: AE-Rec-Error-Based Evaluators

$$\Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'_{\text{train}} \rangle^{1/2}} \qquad \Delta_{\text{clu}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

- kNN: ``d'' is a distance measure, with Euclidean metric being assumed for the feature space
- AE: reconstruction error is a (squared) distance measure, namely $d = R_{\text{AE}}^{1/2}$ but with a more complex metric for the feature space

Introduce AE-A and AE-B: the same architecture

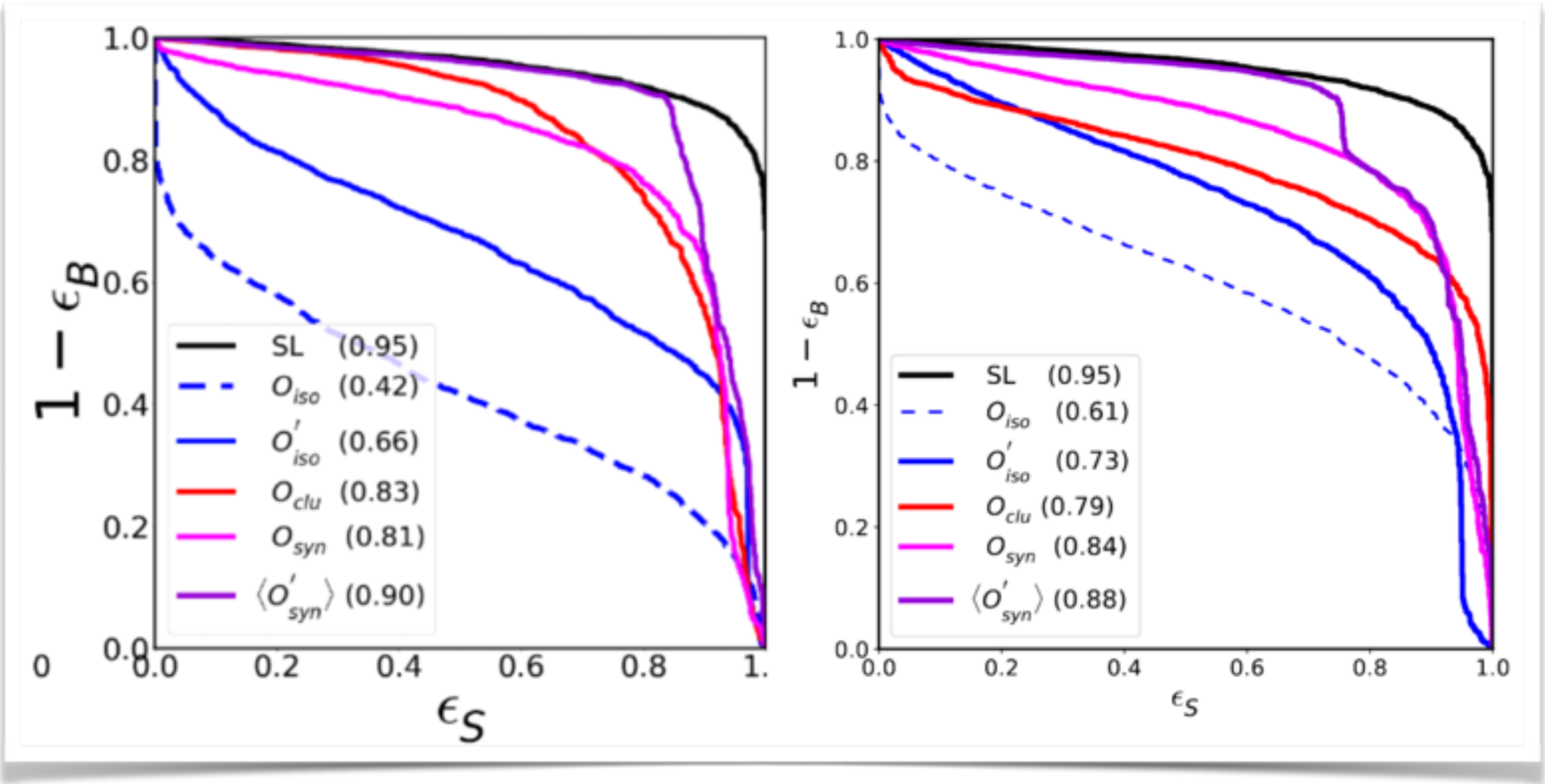
AE-A: trained by training sample

AE-B: trained by testing sample

$$d_{\text{train}} \to R_{\text{AE-A}}^{1/2} \qquad d_{\text{test}} \to R_{\text{AE-B}}^{1/2}$$

=> Reconstruction-error - based O_iso and O_clu

Backup: AE-Rec-Error-Based Evaluators



(For benchmark (C2, R2))

kNN-based

AE-rec-error-based