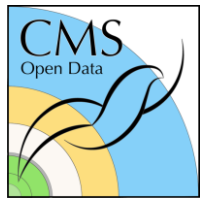


BETHEL
UNIVERSITY

Unravelling physics beyond the standard model with classical and quantum anomaly detection

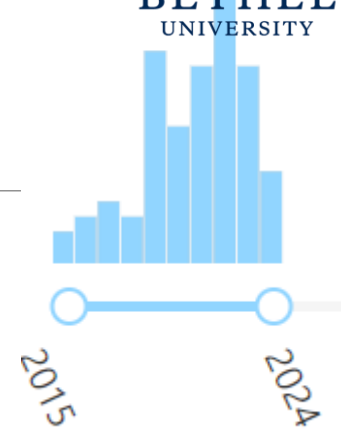
A PEAK INSIDE A CMS OPEN DATA PUBLICATION

JULIE HOGAN

7.30.24

Research with CMS Open Data

CMS Open Data datasets have DOI numbers: [InspireHEP search reveals citations]



Subject

- Experiment-HEP 27
- Phenomenology-HEP 25
- Computing 11
- Data Analysis and Statistics 9
- Other 3
- Theory-HEP 3
- Theory-Nucl 3
- Experiment-Nucl 2
- Quantum Physics 1

Finetuning foundation models for joint analysis optimization in High Energy Physics #1

Matthias Vigl (Tech. U., Munich (main)), Nicole Hartman (Tech. U., Munich (main)), Lukas Heinrich (Tech. U., Munich (main)) (Jan 24, 2024)

Published in: *Mach.Learn.Sci.Tech.* 5 (2024) 2, 025075 • e-Print: 2401.13536 [hep-ex]

pdf DOI cite claim reference search 5 citations

Jet Energy Calibration with Deep Learning as a Kubeflow Pipeline #2

Daniel Holmberg (U. Helsinki (main)), Dejan Golubovic (CERN), Henning Kirschenmann (Helsinki Inst. of Phys.) (Aug 23, 2023)

Published in: *Comput.Softw.Big Sci.* 7 (2023) 1, 9 • e-Print: 2308.12724 [hep-ex]

pdf DOI cite claim reference search 2 citations

Potential of the Julia Programming Language for High Energy Physics Computing #3

Jonas Eschle (U. Zurich (main)), Tamás Gál (Erlangen - Nuremberg U., Theorie III), Mosè Giordano (Imperial Coll., London), Philippe Gras (IRFU, Saclay), Benedikt Hegner (CERN) et al. (Jun 6, 2023)

Published in: *Comput.Softw.Big Sci.* 7 (2023) 1, 10 • e-Print: 2306.03675 [hep-ph]

pdf DOI cite claim reference search 8 citations

Jet fragmentation properties with CMS open-data #4

Saksevul Arias (CINVESTAV, IPN), Eleazar Cuautle (Mexico U.), Hermes León Vargas (Mexico U., ICN) (Feb 17, 2023)

Published in: *Phys.Scripta* 98 (2023) 3, 035305

pdf DOI cite claim reference search 2 citations

2023 Anomaly Detection paper

"Unravelling physics beyond the standard model with classical and quantum anomaly detection"

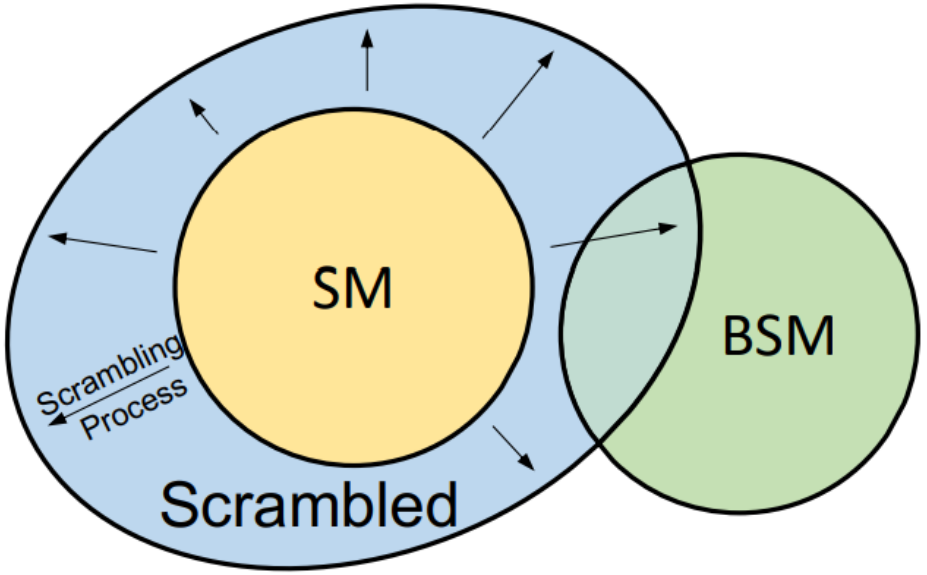
J. Schuhmacher, L. Boggia, V. Belis, E. Puljak, M. Grossi, M. Pierini, S. Vallecorsa, F. Tacchino, P. Barkoutsos, I. Tavernello

Authors from a mix of institutes: IBM Research, ETH Zurich (Theory & Particle Physics), Barcelona, CERN

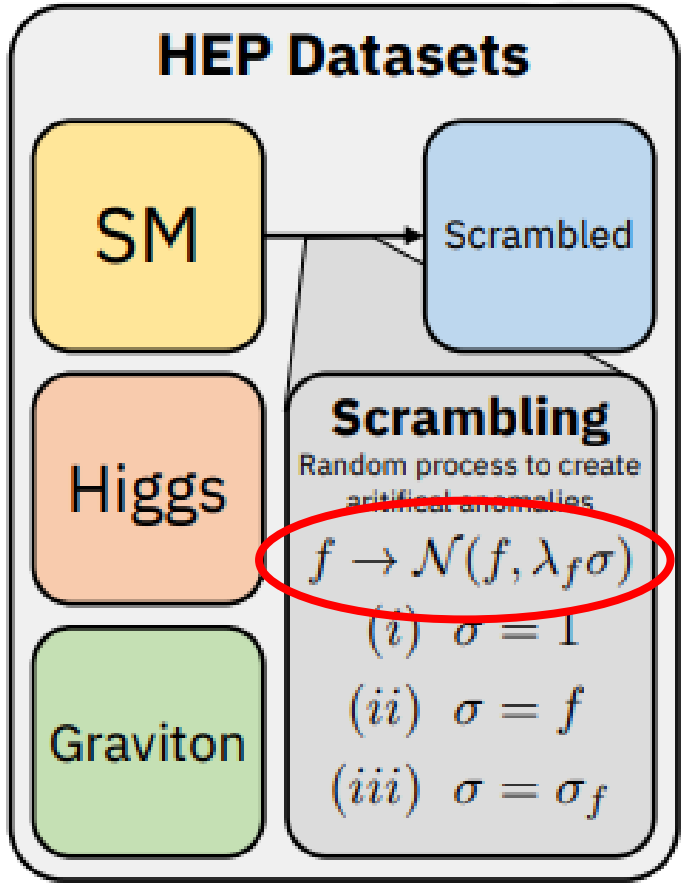
Much hope for finding new physics phenomena at microscopic scale relies on the observations obtained from High Energy Physics experiments, like the ones performed at the Large Hadron Collider … We propose a supervised learning setting in which the signal sample is built perturbing the background sample, without relying on any specific BSM theory

“SM but not quite” – data scrambling

- SM simulation processed for an earlier autoencoder paper from 2012 AOD Open Data
- Variables considered for random scrambling include
  - Momenta: jet scalar sum, lepton vector sum, highest pT lepton, MET components, lepton pseudorap.
  - Isolation: lepton w.r.t charged hadrons, neutral hadrons, photons
  - Counts: jets, b-tagged jets, leptons, 3 PF candidate types
- **Scrambling must obey conservation laws** and other “sensibility” constraints



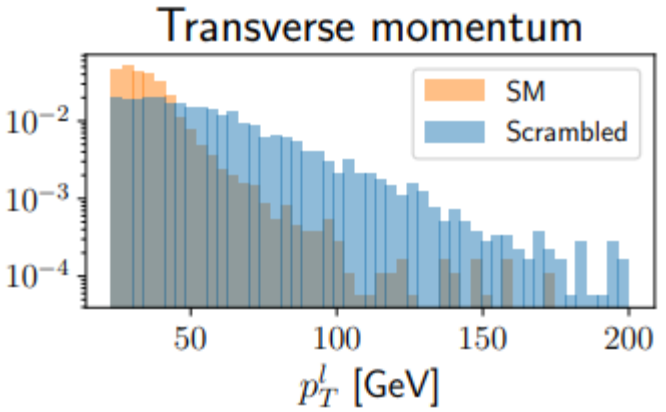
“SM but not quite” – data scrambling

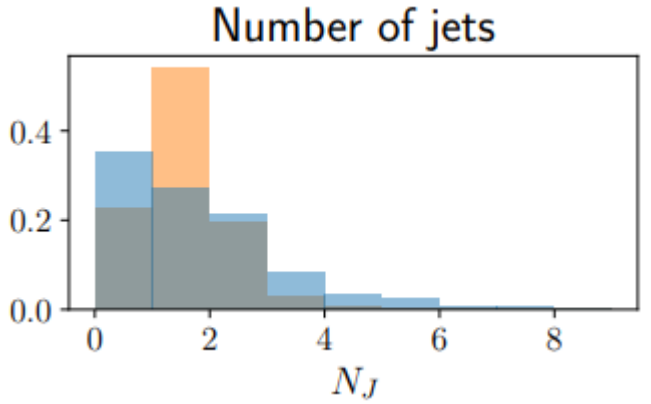


Each feature gets scrambled by sampling from a Gaussian with various methods of changing the std. dev.

3 scrambling strengths defined:

Factor	Low	Medium	High
λ_p	0.5	1.0	2.0
λ_ϕ	0.05	0.1	0.2
λ_η	0.05	0.1	0.2
λ_{H_T}	0.5	1.0	2.0
λ_{Iso}	0.25	0.5	1.0
λ_J	0.5	1.0	2.0
λ_b	0.5	1.0	2.0
$\lambda_{N,\text{charged}}$	0.25	0.5	1.0
$\lambda_{N,\text{neutral}}$	0.25	0.5	1.0
$\lambda_{N,\text{photons}}$	0.25	0.5	1.0
$\lambda_{N,\text{electrons}}$	0.5	1.0	2.0
$\lambda_{N,\text{muons}}$	0.5	1.0	2.0





SVC networks

Support Vector Classifier: embeds data in a high-N space for separability by hyperplanes

Classical kernel: Gaussian $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$

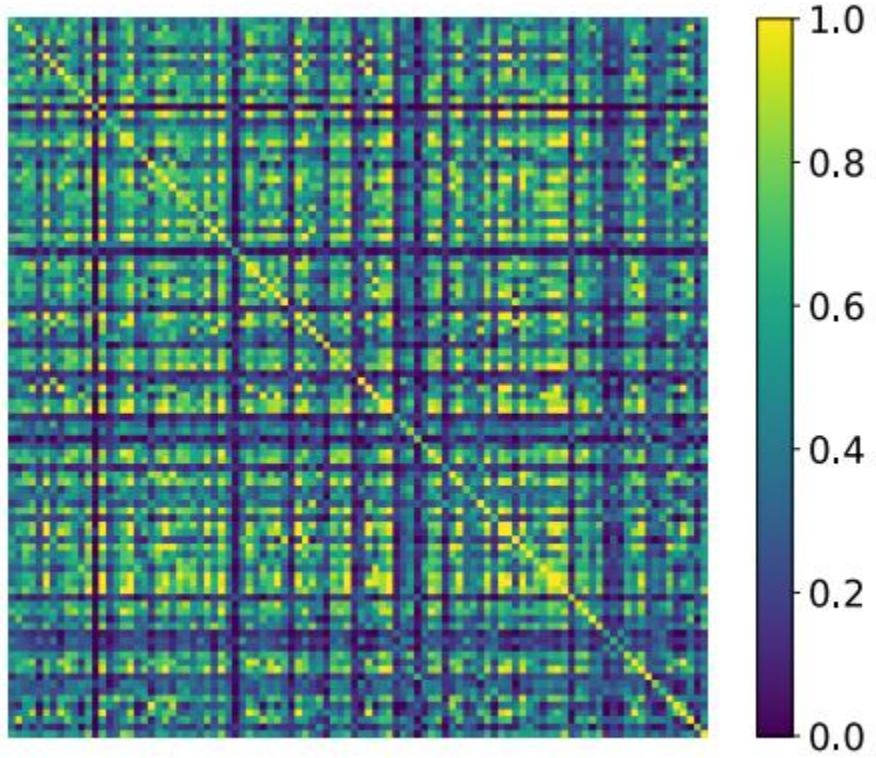
- Optimize the hyperparameter gamma

Quantum kernel:

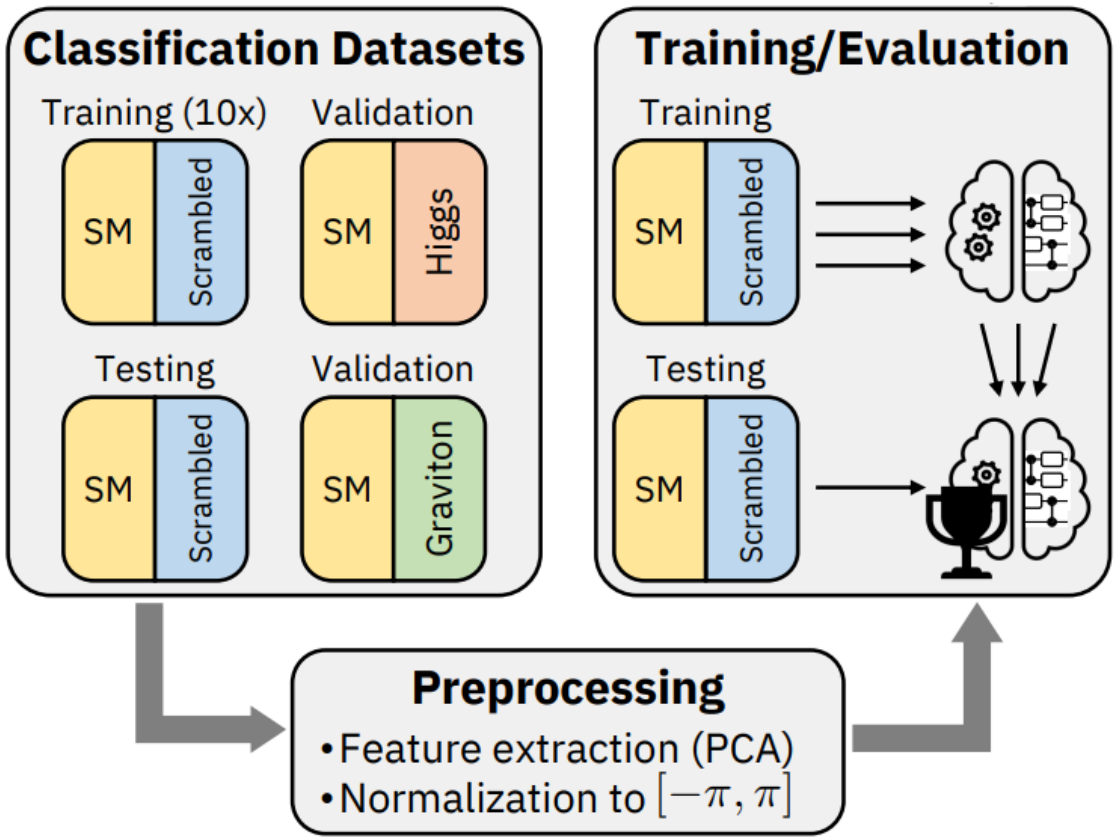
- classical features (**x**) mapped to a quantum state space
- Quantum feature map ϕ

$$K(\mathbf{x}_i, \mathbf{x}_j) = |\langle \phi(\mathbf{x}_i) | \phi(\mathbf{x}_j) \rangle|^2$$

- Kernel evaluated on imb_cairo quantum processor, 6 qubits
- Distinction between **numerical** and **hardware** experiments



Training & Validation scheme



Goal is to determine whether a network trained on the "artificial" anomalies can identify "real" anomalies in the form of known BSM theories

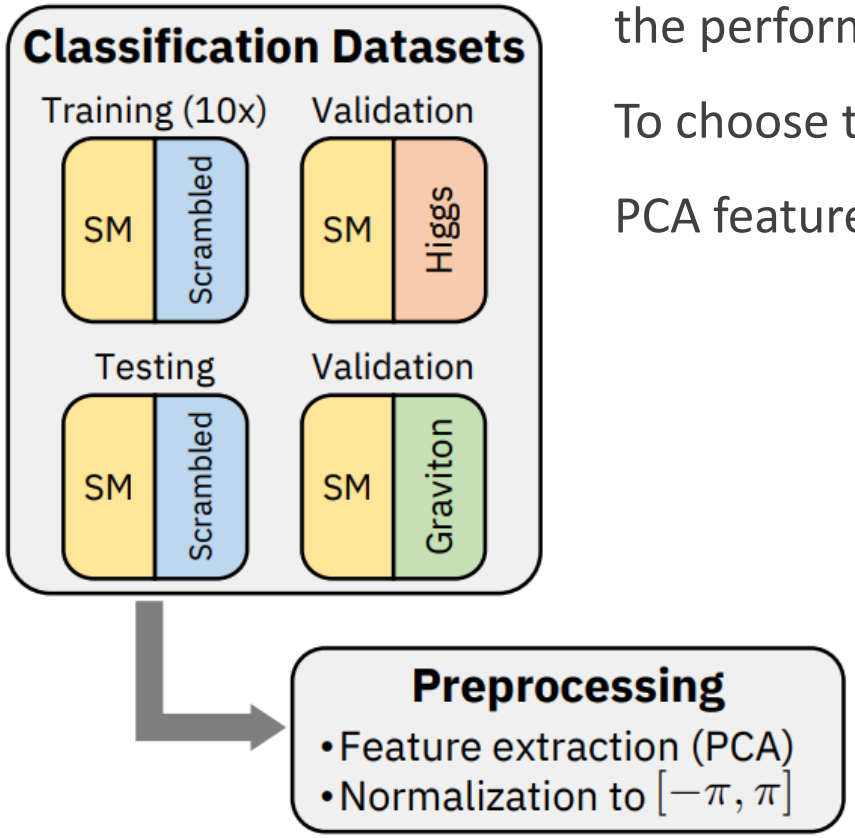
Training & testing performed with:

- Background = SM mixture of W, Z, ttbar, and QCD
- Signal = scrambled data
- 1000 events / class for each of 10 training sets

Highest performer in test dataset is used for the BSM samples

- High-mass VBF Higgs
- Rkk Graviton → ZZ → 4 leptons

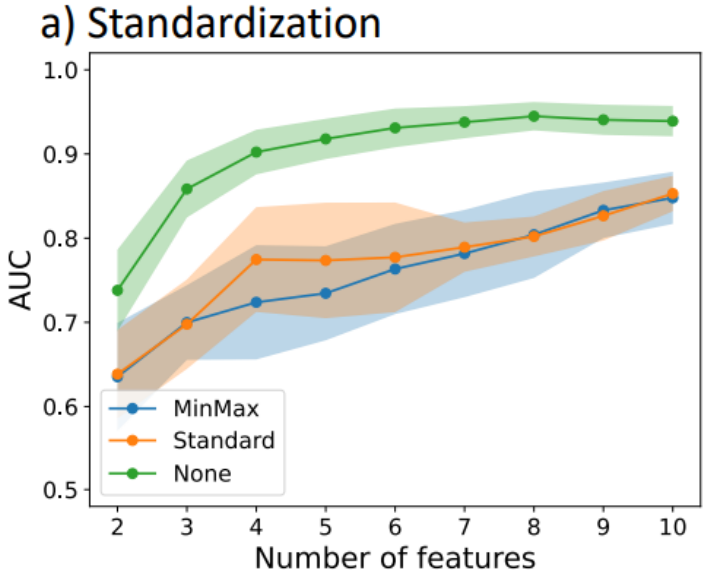
Training & Validation scheme

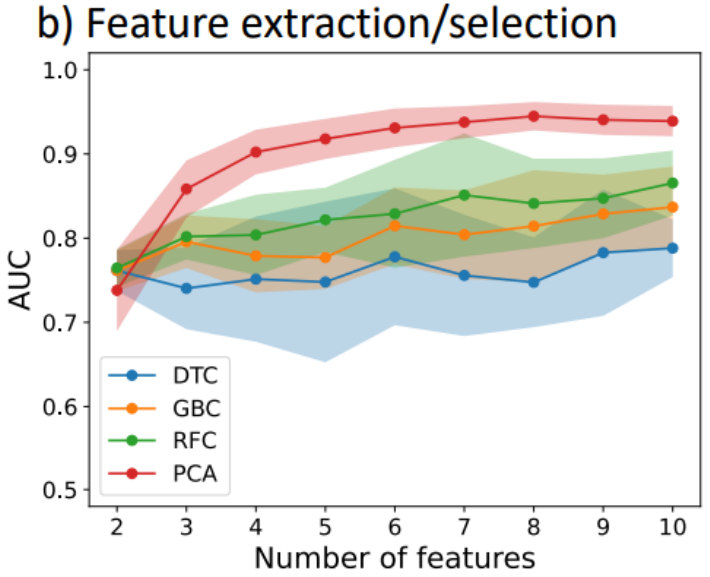


Tests showed that standardizing the features before down-selecting harmed the performance significantly.

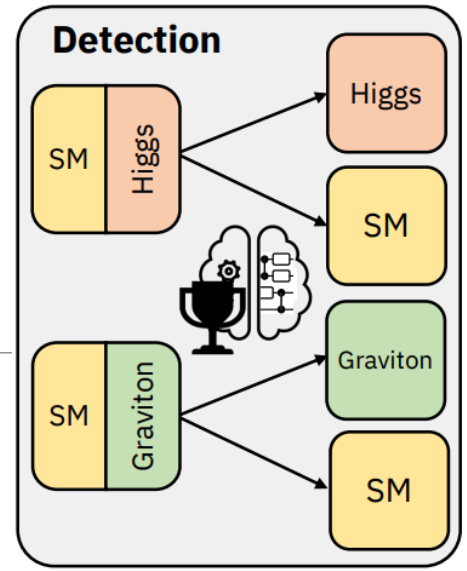
To choose the final training features, PCA outperformed various decision trees.

PCA features are then scaled to all share the same range





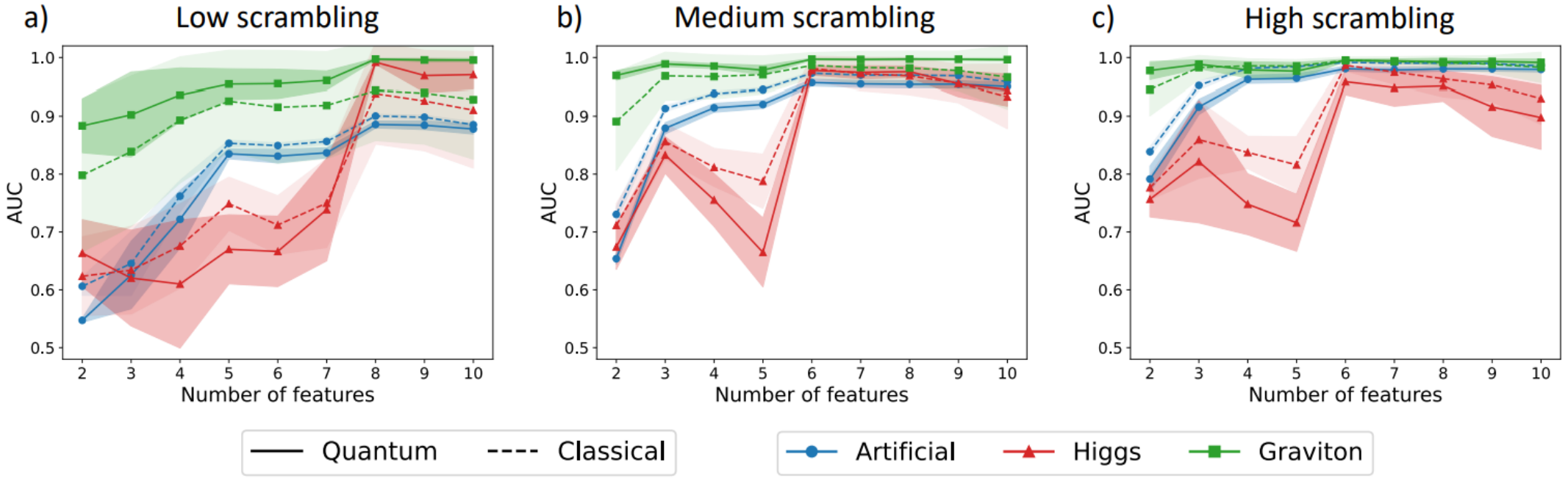
Training results



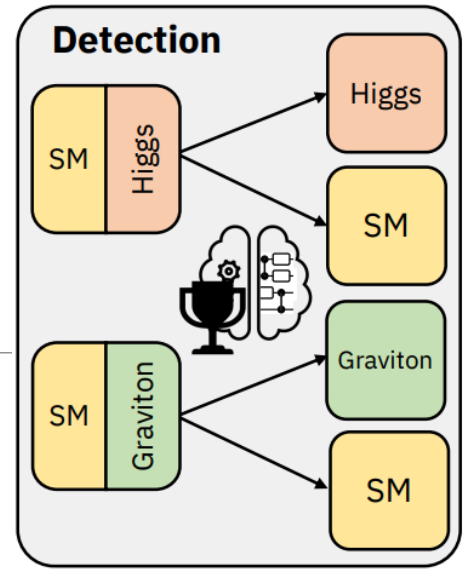
More features → better anomaly detection, including the BSM models

Fewer features needed if the scrambling is more extreme

Quantum struggles against the classical kernel for anomalies and Higgs

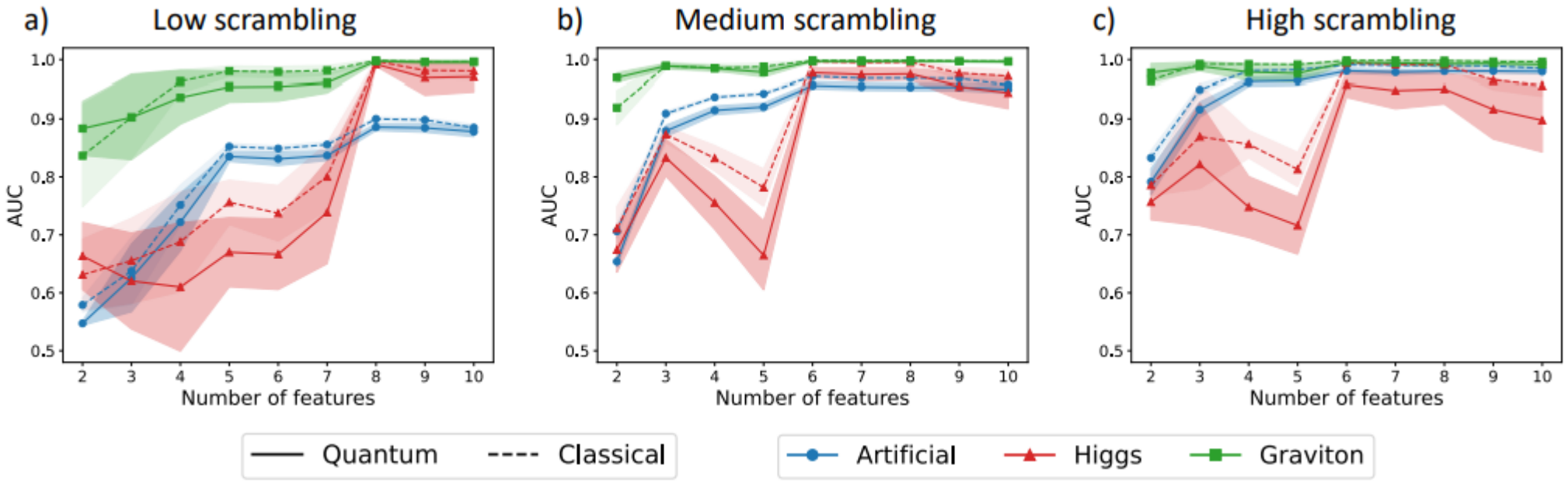


Training results



Does the QSVC appear less performant because the SVC is overfitted?

Doesn't appear so: adding a bias to the SVC doesn't change artificial anomaly finding, and allows the SVC to eclipse the QSVC even for gravitons



Quantum simulation vs hardware

The main results all use a simulated “perfect” quantum computer.

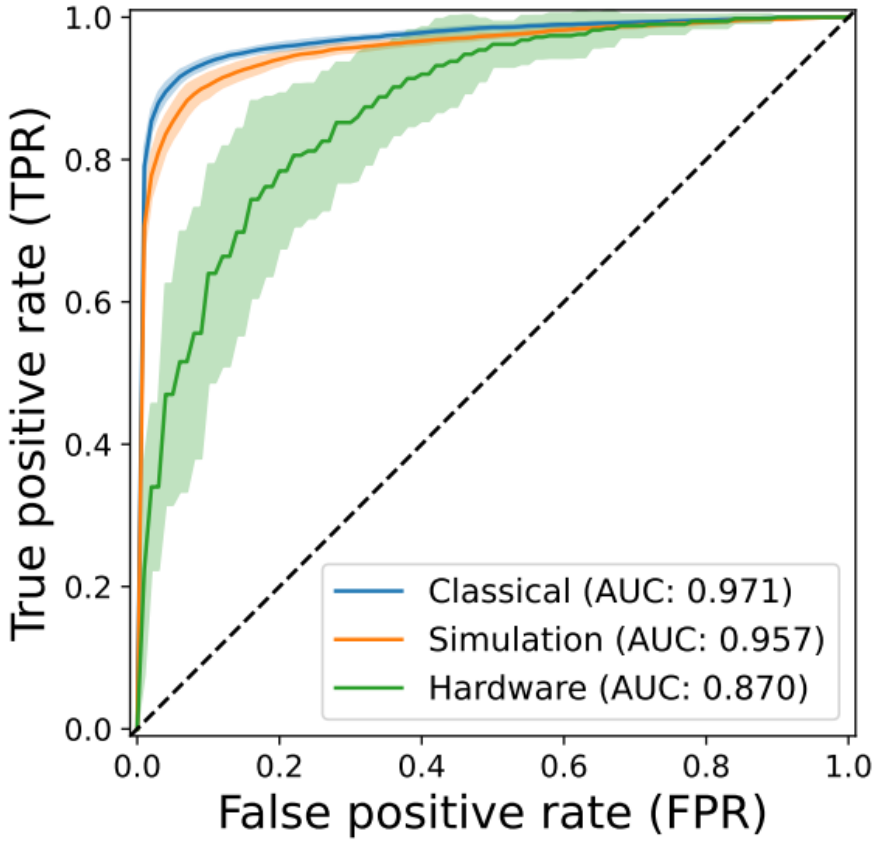
Hardware test performed with *ibm_cairo*

- 6 qubits used to make the kernel map
- Had to downscope to 50 events / class, and no external BSM samples
- Used previous optimized settings and chose 6 PCA features

Clearly the “noisy” hardware is not as performant!

- Unclear how the simulated SVC and QSVC would do with only 50 events / class

They conclude that this classification problem is not well suited to teasing out the future benefits of quantum computing



Summary

This paper is a great example of collaboration across disciplines!

Constructing this type of dataset from Open Data has now become even simpler

- 2012 AOD simulation
- This paper shared a dataset with an earlier paper
- MiniAOD and NanoAOD formats for 2015 and 2016 data are more accessible

For studies that don't rely on future detector simulation, we hope to see NanoAOD Open Data rise in popularity w.r.t fast simulations like Delphes!

We look forward to reading your work using Open Data!

All datasets have a unique DOI that you are requested to cite in any applications or publications.