Testing and Validation

Example Data Set

Streaming Readout Workshop

August 20, 2026

Oliver Partridge, Summer Master Student

Marco Meyer-Conde, Assistant Professor

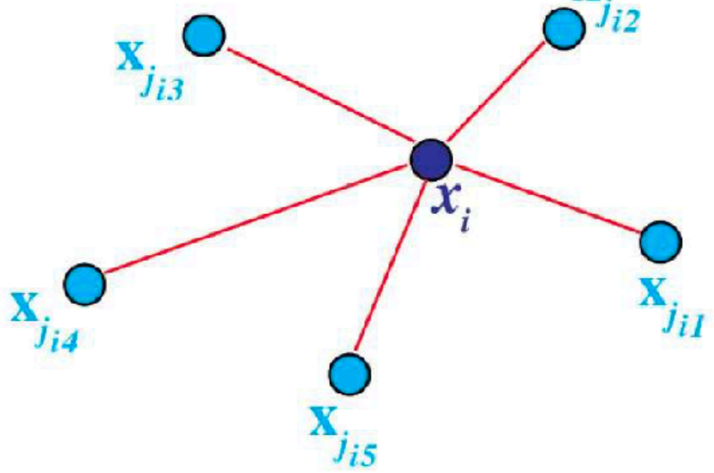
Advanced Research Laboratories, Tokyo City University

ONNX Implementation in EICRecon

Technical Subtlety

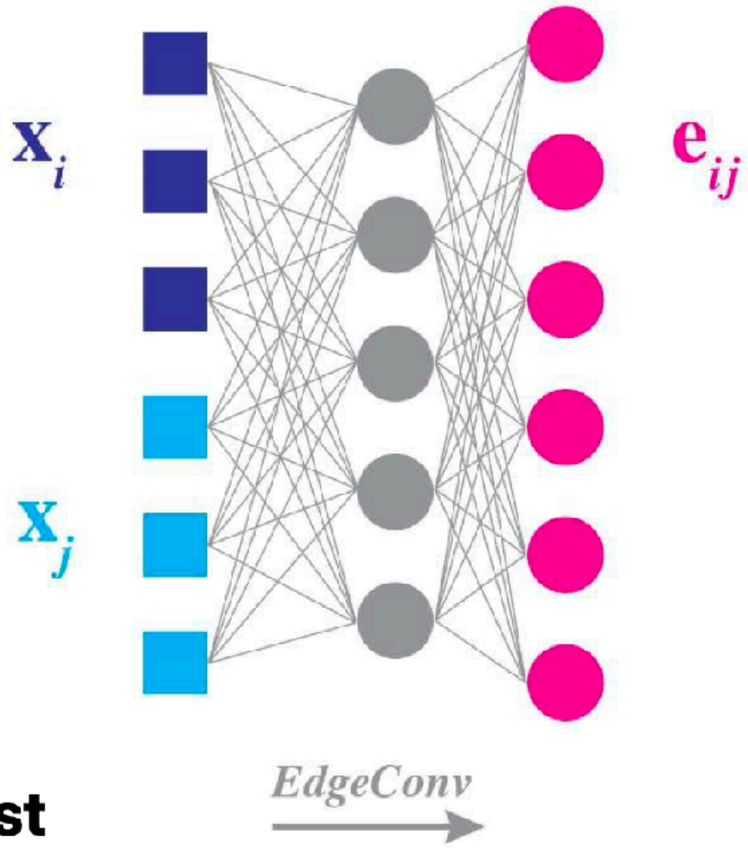
Dynamic KNN

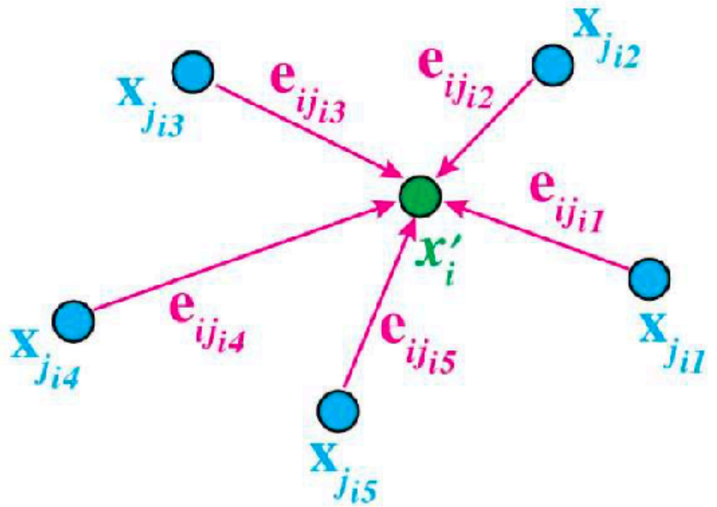
Using nanoflann



Draw relations with nearest neighbours

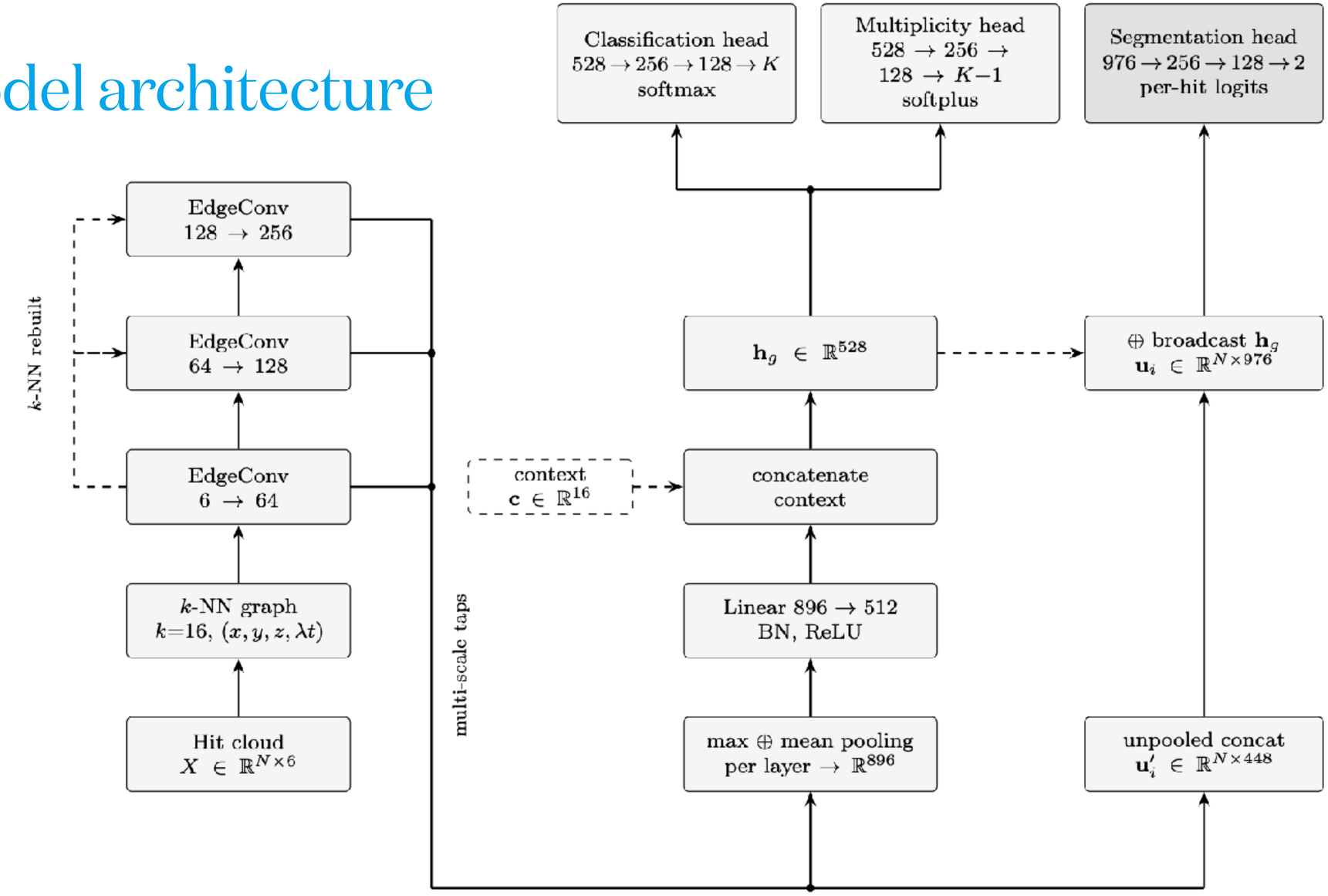
Static graph





Represent each hit by its interaction with KNN

Model architecture



GNN Architecture

Test Production

Data

~8000 simulated collisions events split into 75/15/15 training, validation and test

Optimization

Epochs 20-200
Optimizer Adam
Batch size

Loss function

Cross-entropy loss

Stability

Curriculum strategy
Dropout, weight decay, gradient clipping

Hardware/compute

Training runs on a 4-GPU workstation (NVIDIA Blackwell)

Hyperparameter selection

3-4 layer depth
Graduated width 64, 64, 128, 256

Goal

2 heads:
- Classification
- Multiplicity

Model dimensionning

Dataset being considered

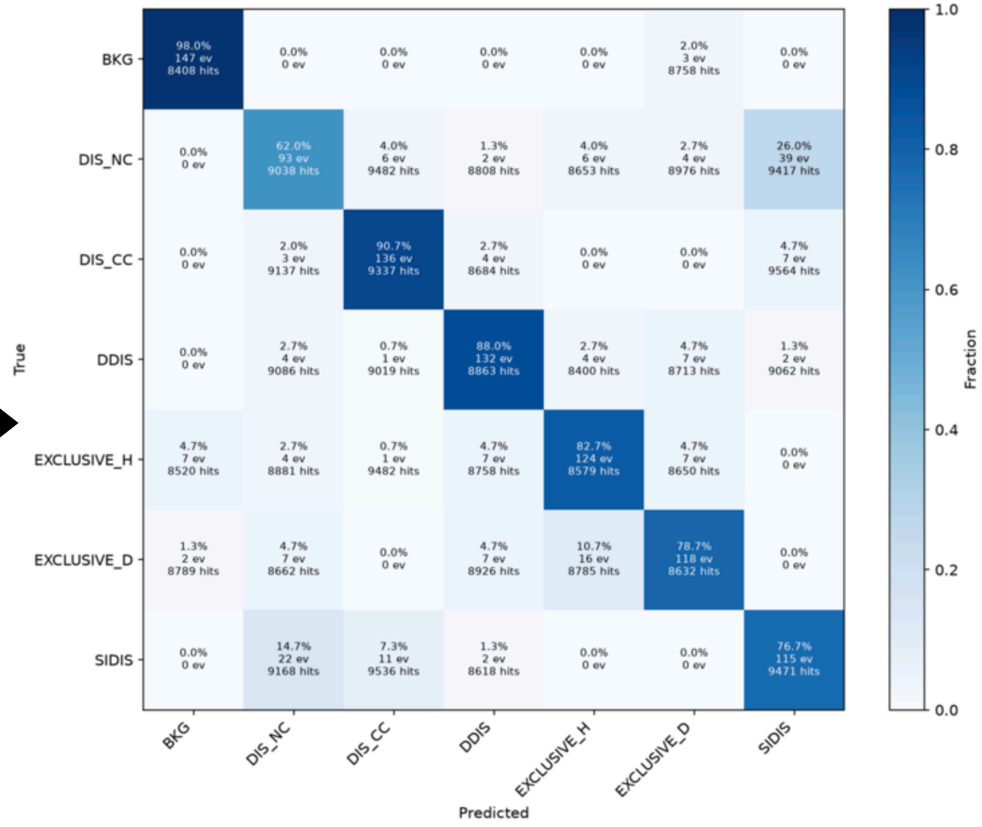
Initial Classification

simulation_campaign_datasets Public — Watch

main — 115 Branches — 0 Tags — Go to file — Add file — Code

3 people — Split DIS ep NC and CC jobs by generator family (#52) — 903e031 · yesterday — 443 Commits

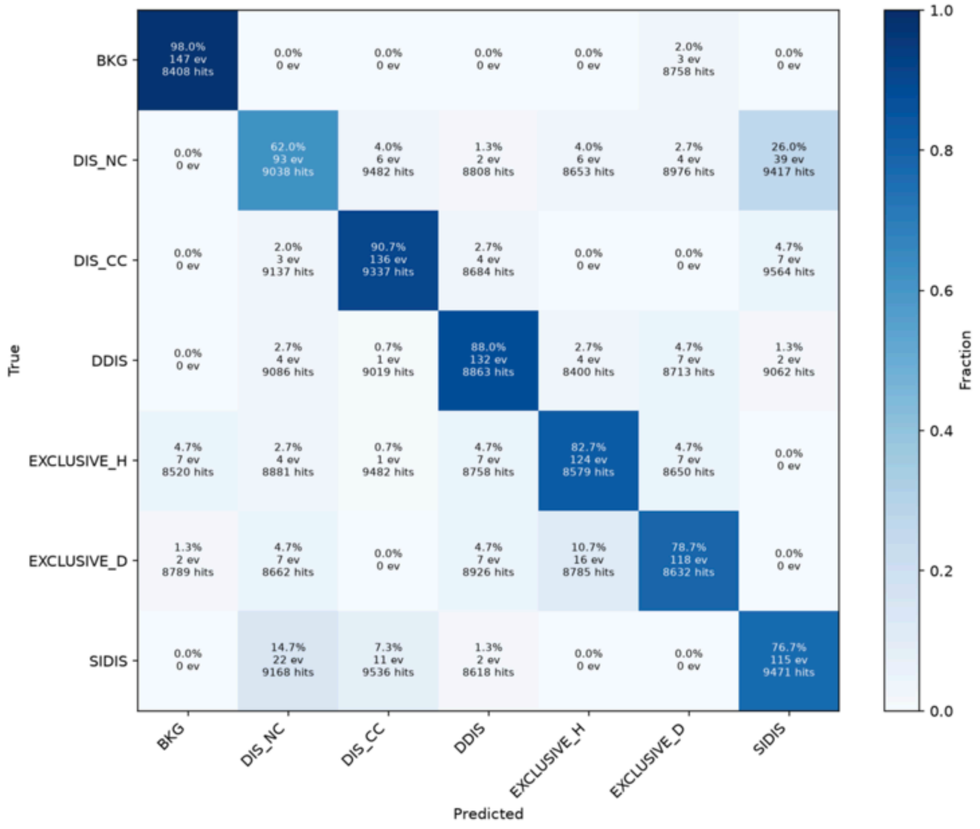
.github/workflows	Add configurable detector version support for datasets (#...	9 months ago
BACKGROUNDS	Update background configurations (#5)	10 months ago
DDIS	Add rapgap3.310-2.0 DDIS ep datasets and nest DDIS lay...	3 weeks ago
DIS	Split DIS ep NC and CC jobs by generator family (#52)	yesterday
EW_BSM	Add EW_BSM physics category with ALP datasets (#41)	3 months ago
EXCLUSIVE	Add DEMPgen-1.2.5 pi+ datasets for 9x130 and 9x275 (#...	4 days ago
SIDIS	Add pythia6-eic 1.3.0 ep_noradcor datasets for 9x130 an...	2 days ago
SINGLE	Collect timing data for single neutrons	2 years ago
SR	nevents needs gzip	5 years ago
config_data	Add background mixed configs for 10x130 GeV beam ener...	6 months ago
scripts	Respect prescribed nevents (#38)	3 months ago
.gitlab-ci.yml	Add EW_BSM physics category with ALP datasets (#41)	3 months ago
README.md	Initial commit	5 years ago



https://github.com/eic/simulation_campaign_datasets

Dataset being considered

Initial Classification

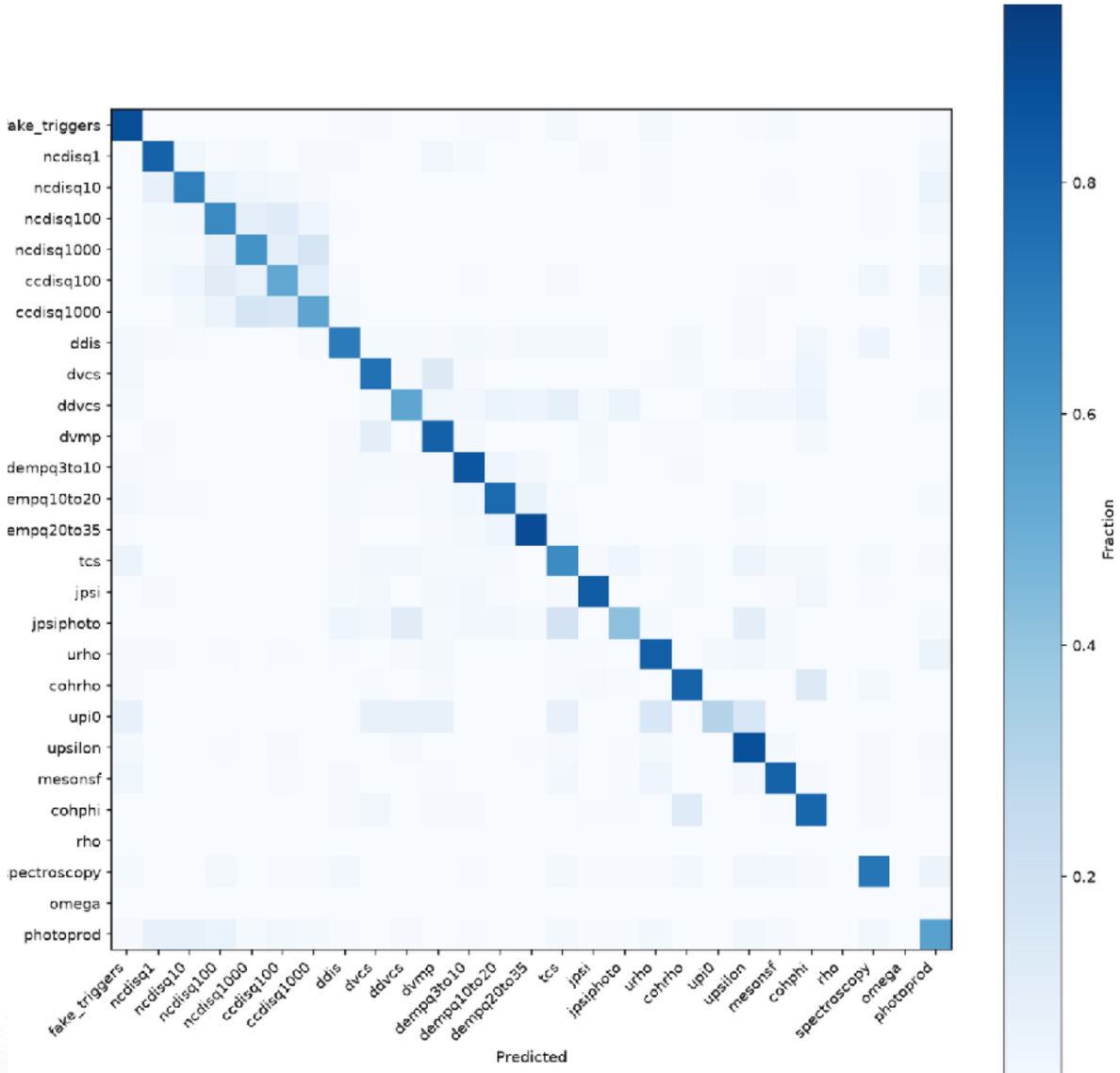


#	Class	Physics process	Generator
0	`fake_triggers`	fake trigger (no true collision matched)	—
1	`ncdisq1`	neutral-current DIS, $Q^2 \geq 1$	Pythia 8
2	`ncdisq10`	neutral-current DIS, $Q^2 \geq 10$	Pythia 8
3	`ncdisq100`	neutral-current DIS, $Q^2 \geq 100$	Pythia 8
4	`ncdisq1000`	neutral-current DIS, $Q^2 \geq 1000$	Pythia 8
5	`ccdisq100`	charged-current DIS, $Q^2 \geq 100$	Pythia 8
6	`ccdisq1000`	charged-current DIS, $Q^2 \geq 1000$	Pythia 8
7	`ddis`	diffractive DIS	RAPGAP
8	`dvcs`	deeply virtual Compton scattering	EpIC
9	`ddvcs`	double deeply virtual Compton scattering	EpIC
10	`dvmp`	deeply virtual meson production	EpIC
11	`dempq3to10`	deeply exclusive meson production, $3 < Q^2 < 10$	DEMPgen
12	`dempq10to20`	deeply exclusive meson production, $10 < Q^2 < 20$	DEMPgen
13	`dempq20to35`	deeply exclusive meson production, $20 < Q^2 < 35$	DEMPgen
14	`tcs`	timelike Compton scattering	—
15	`jpsi`	diffractive J/ψ production	lAger
16	`jpsiphoto`	photoproduced J/ψ	eSTARlight
17	`urho`	u-channel ρ photoproduction	—
18	`cohrho`	coherent ρ photoproduction	Sartre
19	`upi0`	u-channel π^0 photoproduction	—
20	`upsilon`	Υ photoproduction	—
21	`mesonsf`	meson structure function	eicMesonSFGen
22	`cohphi`	coherent ϕ photoproduction	Sartre
23	`rho`	exclusive ρ production	—
24	`spectroscopy`	exotic spectroscopy	—
25	`omega`	ω production	—
26	`photoprod`	minimum-bias photoproduction	Pythia 8

https://github.com/eic/simulation_campaign_datasets

Event Classification Taxonomy

#	Class	Physics process	Generator
0	`fake_triggers`	fake trigger (no true collision matched)	—
1	`ncdisq1`	neutral-current DIS, $Q^2 \geq 1$	Pythia 8
2	`ncdisq10`	neutral-current DIS, $Q^2 \geq 10$	Pythia 8
3	`ncdisq100`	neutral-current DIS, $Q^2 \geq 100$	Pythia 8
4	`ncdisq1000`	neutral-current DIS, $Q^2 \geq 1000$	Pythia 8
5	`ccdisq100`	charged-current DIS, $Q^2 \geq 100$	Pythia 8
6	`ccdisq1000`	charged-current DIS, $Q^2 \geq 1000$	Pythia 8
7	`ddis`	diffractive DIS	RAPGAP
8	`dvcs`	deeply virtual Compton scattering	EpIC
9	`ddvcs`	double deeply virtual Compton scattering	EpIC
10	`dvmp`	deeply virtual meson production	EpIC
11	`dempq3to10`	deeply exclusive meson production, $3 < Q^2 < 10$	DEMPgen
12	`dempq10to20`	deeply exclusive meson production, $10 < Q^2 < 20$	DEMPgen
13	`dempq20to35`	deeply exclusive meson production, $20 < Q^2 < 35$	DEMPgen
14	`tcs`	timelike Compton scattering	—
15	`jpsi`	diffractive J/ψ production	lAger
16	`jpsiphoto`	photoproduced J/ψ	eSTARlight
17	`urho`	u-channel ρ photoproduction	—
18	`cohrho`	coherent ρ photoproduction	Sartre
19	`upi0`	u-channel π^0 photoproduction	—
20	`upsilon`	Υ photoproduction	—
21	`mesonsf`	meson structure function	eicMesonSFGen
22	`cohphi`	coherent ϕ photoproduction	Sartre
23	`rho`	exclusive ρ production	—
24	`spectroscopy`	exotic spectroscopy	—
25	`omega`	ω production	—
26	`photoprod`	minimum-bias photoproduction	Pythia 8



Event Builder Performances

Each class is mixed with ”gold-coating” — N true physics per 2us (~5MHz)

No lost on the efficiency

Acceptable ACTS + Calo clustering

— datasets/gnn-20260721-eventbld-pileup/eicrecon/10x100-gold-coating classes —

Class			eicrecon	(time,hits/tf)	(hits)	STAGE 1: TRIGGER LOGIC time %eff %pur	STAGE 1: TRIGGER LOGIC track %eff %pur	STAGE 1: TRIGGER LOGIC calo %eff %pur	STAGE 1: TRIGGER LOGIC trigger %eff %pur	STAGE 2: ACTS track %eff %pur	STAGE 2: ACTS calo %eff %pur	pileup %eff (found/inj)
bkg			–	–	–	–	–	–	–		–	
ncdisq1	1f	0tf		–	–	–	–	–	–	–	–	–
ncdisq10	1f	100tf	4,377ev	(1.0s 24,182h/tf)	(621h/ev)	100% 3%	97% 34%	100% 3%	100% 3%	82% 85%	14% 49%	97.0% (766/790)
ncdisq100	1f	0tf		–	–	–	–	–	–	–	–	–
ncdisq1000	1f	100tf	4,388ev	(1.0s 24,365h/tf)	(661h/ev)	100% 3%	100% 37%	100% 3%	100% 3%	79% 86%	11% 50%	97.2% (806/829)
ccdisq100	1f	100tf	4,395ev	(1.0s 24,065h/tf)	(642h/ev)	100% 3%	99% 37%	100% 3%	100% 3%	76% 84%	11% 46%	97.7% (816/835)
ccdisq1000	1f	100tf	4,348ev	(1.4s 22,932h/tf)	(608h/ev)	100% 3%	99% 35%	100% 3%	100% 3%	75% 88%	10% 45%	97.5% (818/839)
ddis	1f	100tf	4,366ev	(1.4s 22,670h/tf)	(554h/ev)	100% 3%	91% 37%	100% 3%	100% 3%	78% 79%	4% 25%	96.8% (825/852)
dvcs	1f	100tf	4,387ev	(1.3s 22,825h/tf)	(524h/ev)	100% 3%	78% 33%	100% 3%	100% 3%	98% 64%	0% 28%	96.5% (821/851)
ddvcs			–	–	–	–	–	–	–		–	
dvmp	1f	100tf	4,369ev	(1.4s 22,256h/tf)	(528h/ev)	100% 3%	79% 34%	100% 3%	100% 3%	98% 71%	5% 30%	96.5% (807/836)
dempq3to10	1f	100tf	4,367ev	(1.4s 22,594h/tf)	(535h/ev)	100% 3%	97% 41%	100% 3%	100% 3%	96% 69%	–	96.4% (825/856)
dempq10to20	1f	0tf		–	–	–	–	–	–	–	–	–
dempq20to35	1f	100tf	4,402ev	(0.9s 24,126h/tf)	(554h/ev)	100% 3%	91% 35%	100% 3%	100% 3%	96% 67%	0% 35%	97.2% (869/894)
tcs	1f	0tf		–	–	–	–	–	–	–	–	–
jpsi	1f	0tf		–	–	–	–	–	–	–	–	–
jpsiphoto			–	–	–	–	–	–	–		–	
urho	1f	100tf	4,347ev	(1.0s 23,657h/tf)	(576h/ev)	100% 3%	40% 22%	100% 3%	100% 3%	100% 49%	–	96.4% (795/825)
cohrho	1f	100tf	4,384ev	(1.3s 23,112h/tf)	(569h/ev)	100% 3%	95% 37%	100% 3%	100% 3%	88% 71%	–	96.9% (873/901)
upi0			–	–	–	–	–	–	–		–	
upsilon	1f	100tf	4,350ev	(1.3s 22,609h/tf)	(540h/ev)	100% 3%	95% 38%	100% 3%	100% 3%	96% 73%	–	97.0% (770/794)
mesonsf	1f	0tf		–	–	–	–	–	–	–	–	–
cohphi	1f	100tf	4,362ev	(1.3s 22,445h/tf)	(553h/ev)	100% 3%	89% 34%	100% 3%	100% 3%	82% 72%	–	97.9% (806/823)
rho	1f	0tf		–	–	–	–	–	–	–	–	–
spectroscopy	1f	100tf	4,357ev	(0.9s 23,545h/tf)	(582h/ev)	100% 3%	97% 40%	100% 3%	100% 3%	84% 78%	–	98.2% (763/777)
omega			–	–	–	–	–	–	–		–	
photoprod	1f	0tf		–	–	–	–	–	–	–	–	–
TOTAL	22f	1,400tf	61,199ev	(1.2s 23,242h/tf)	(575h/ev)	100% 3%	89% 36%	100% 3%	100% 3%	81% 78%	10% 42%	97.1% (11360/11702)

(f=files tf=2 µs timeframes ev=eventbuilder candidates, real+fake ()=mean h=hits (h/ev counts

ML Testing & Validation for the GNN Event Pre-Filter

Signal and Background Mixing

- Machine background (expressed in kHz): https://github.com/eic/simulation_campaign_datasets/tree/main/config_data
- Physics signals: https://github.com/eic/simulation_campaign_datasets

Source	Status	Vacuum	Gold Coating	File size	Effect
Synchrontron Radiation	2000	36.608 GHz	13.277 GHz	~78 MB	1–2 photons per event; unique templates reused 10 000×
Electron bremsstrahlung radiation	3000	3.177 MHz	3.177 MHz	~94 MB	Few-particle QED elastic scatter
Electron Touscheck scattering	4000	29.56 kHz	29.56 kHz	~525 MB	Intra-bunch scatter Slightly more complex final state
Electron Coulomb scattering processes	5000	233.5 kHz	233.5 kHz	~6.2 GB	Full DIS final state with hadronisation
Proton beam gas interactions	6000	31.9 kHz	31.9 kHz	~16.7 GB	Soft/hard hadronic production Many secondary tracks

EIC Campaign

How data get wrapped into a simulation stream? (ideally)

- Bkg + Events #1...N using generatedStatus
- PhysicsEvent weight: Event level information
- Frame level information using EventBuilderFrameInfo
- AI ground truth encoded using generatedStatus in *HitLinks for segmentation

Generated Status	Meaning
0-99999	bkg/machine-background
100000-199999	ncdisq1
200000-299999	ncdisq10
....	....

Slot	Content
Background	Machine/background events
Slot 1	Classification + corresponding event information
Slot 2	Classification + corresponding event information
	…

$$\text{eventStatus} = \text{generatorStatus} + 10^5.\text{class} + 10^6.\text{slot} \sim \text{ground truth}$$

Testing & validation

Event builder for EICRecon — validating the GNN pre-filter

Follow-up to Partridge & Meyer-Conde — ePIC Streaming Readout Workshop, Aug 19 2026

What's being validated

Deterministic vs learned features

Stage A for Benchmarking

Rule-based event builder

- Trigger candidate (MPGD + TOF gating)
- Timing coincidence (per-detector σ gate)
- Tracklet coincidence (IP-pointing)
- Calorimeter energy threshold

Stage B: Testing & Validation

GNN pre-filter — 3 heads

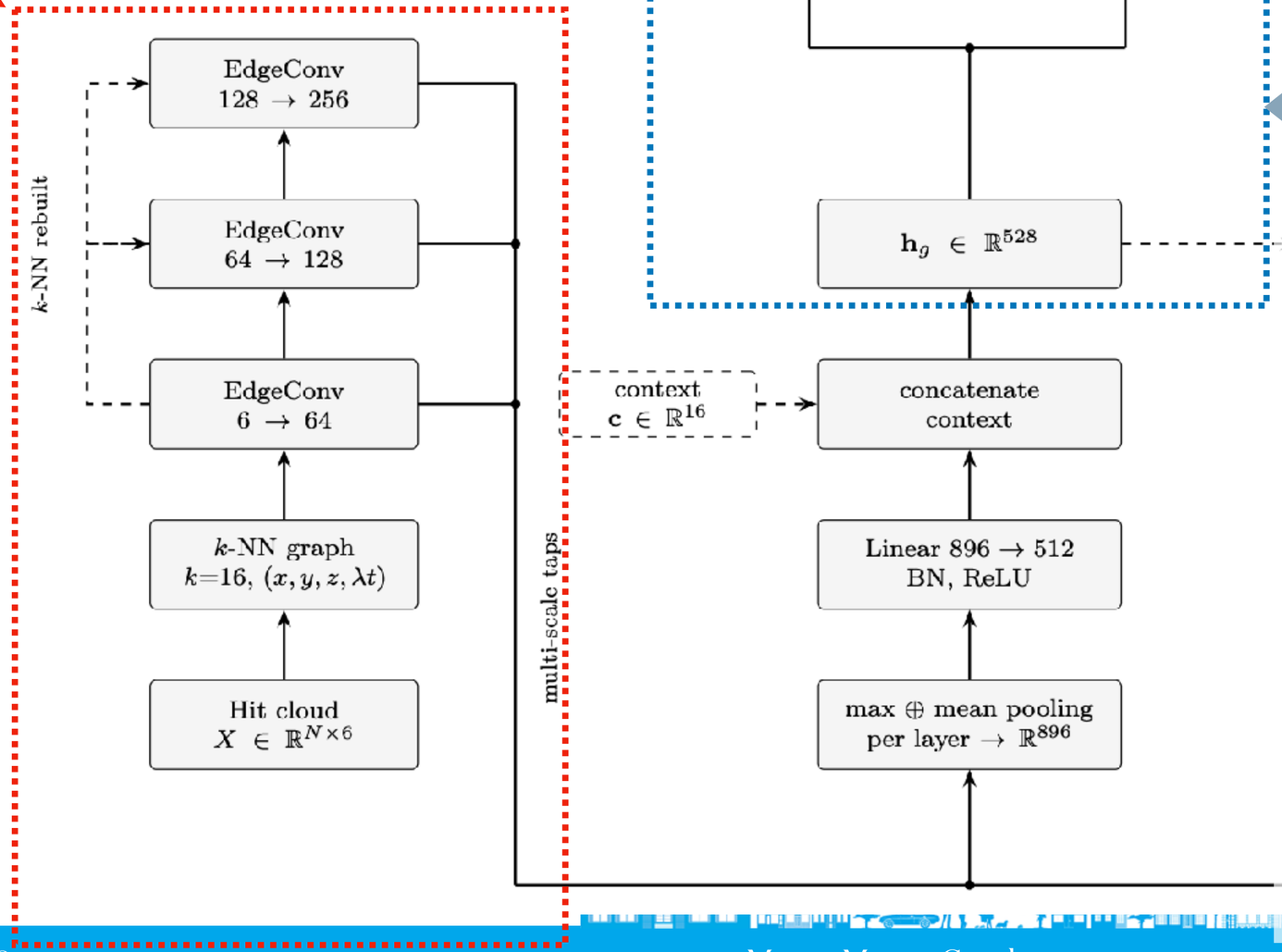
- Classification (background vs. signal, class)
- Segmentation (hit-level attribution)
- Multiplicity (pileup count)

Deterministic → learned

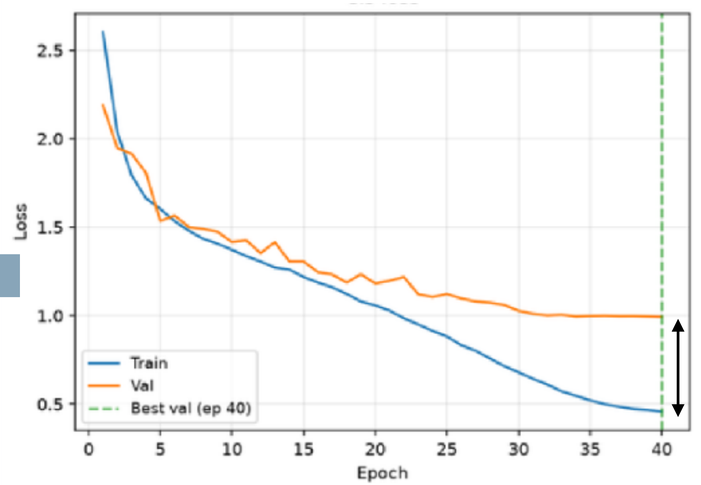
Model architecture

Event level classification

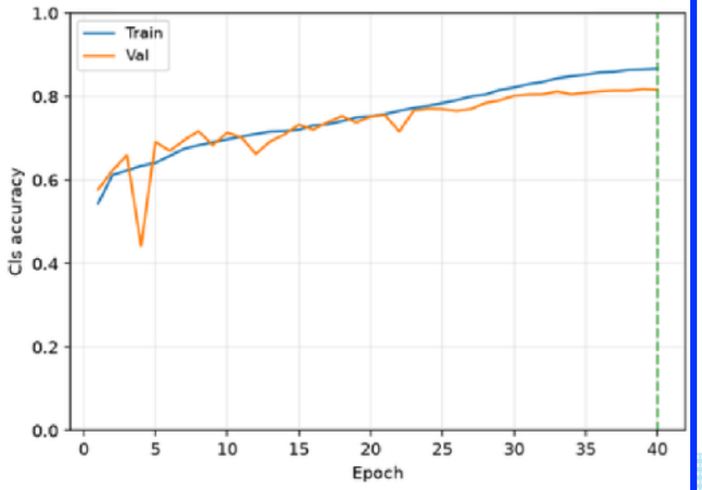
Feature extraction



Loss Estimate (Classification head only)



Classification head Accuracy

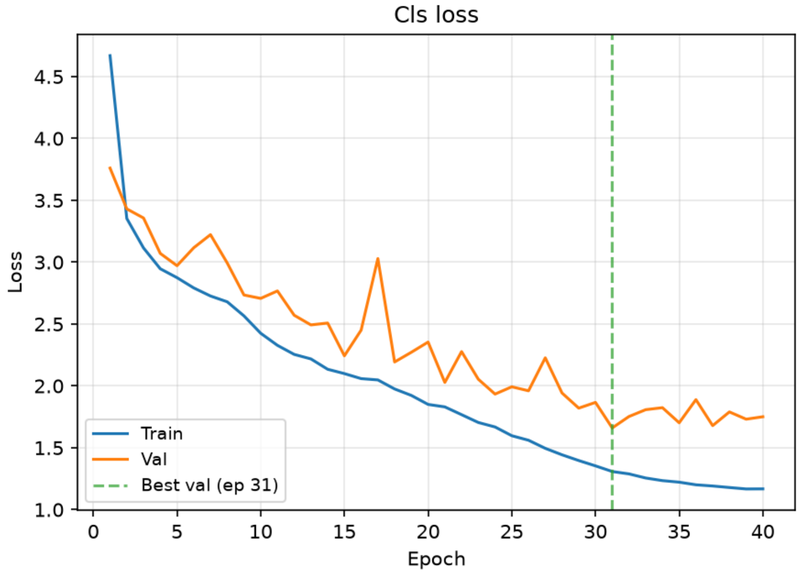


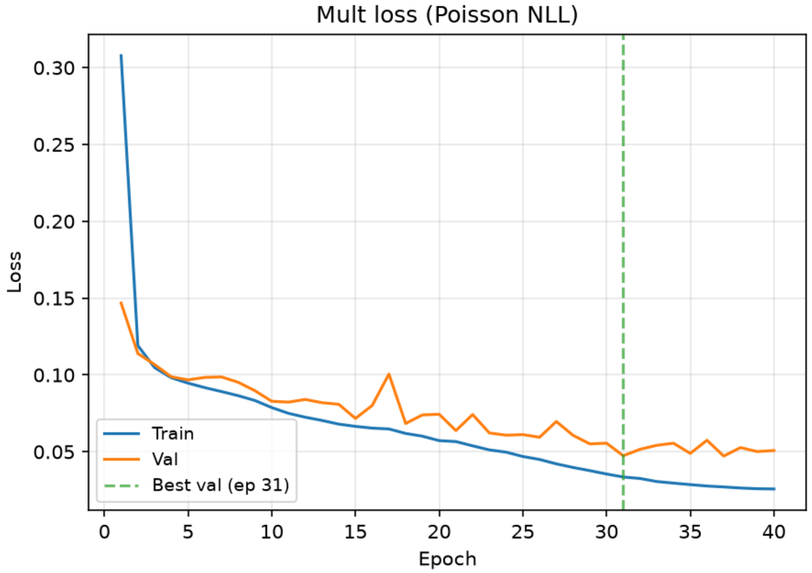
~85% with unoptimized training (!)

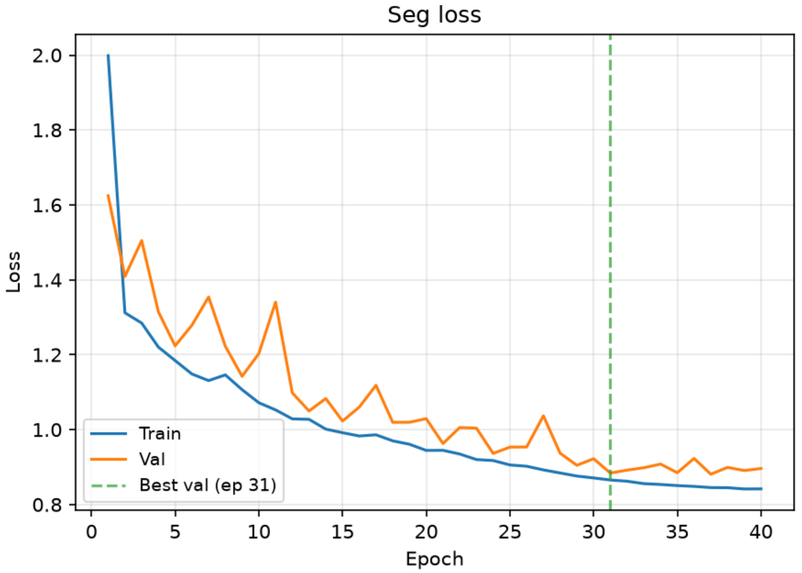
Results

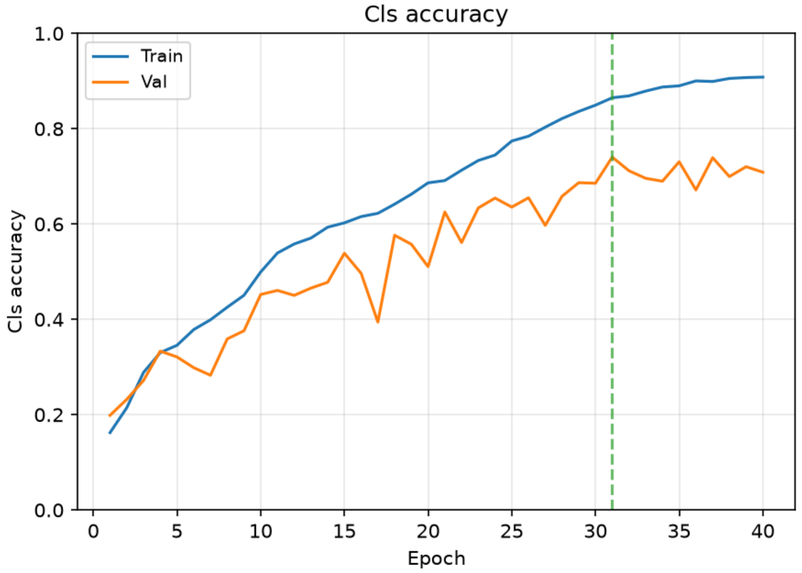
Training and intermediate results

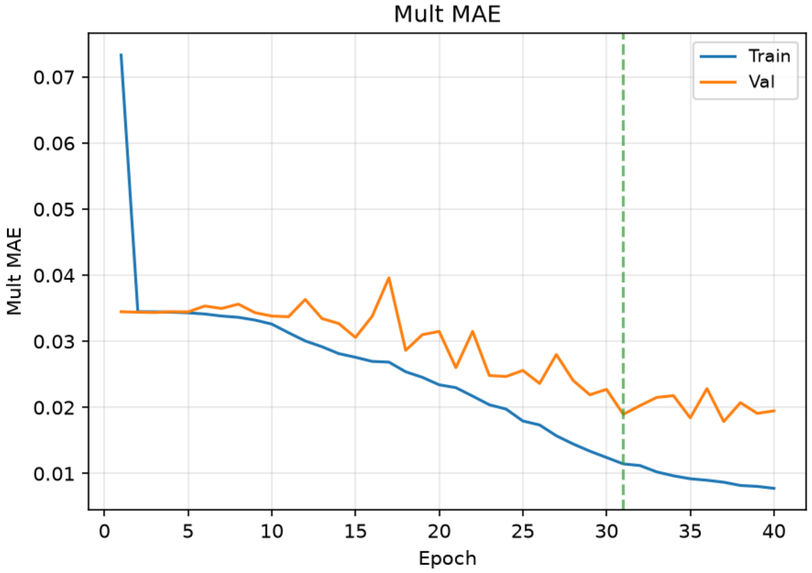
3 heads training

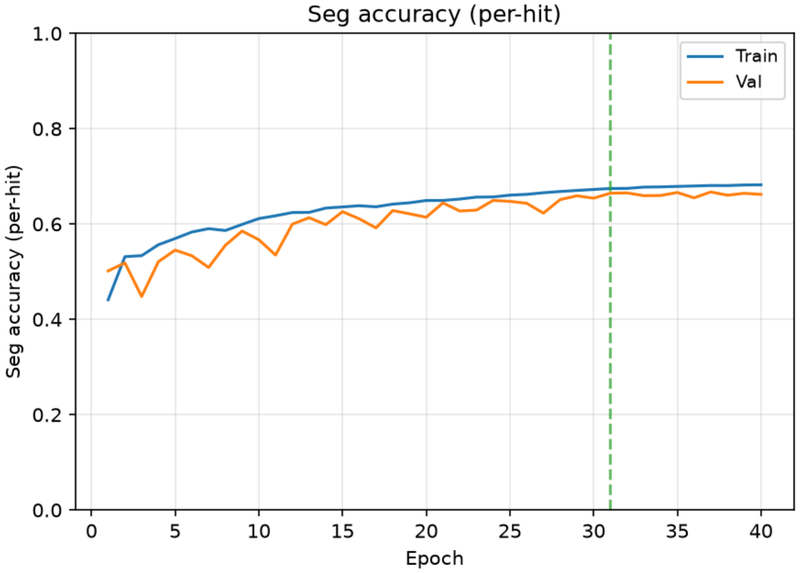












Results

Unresolved class

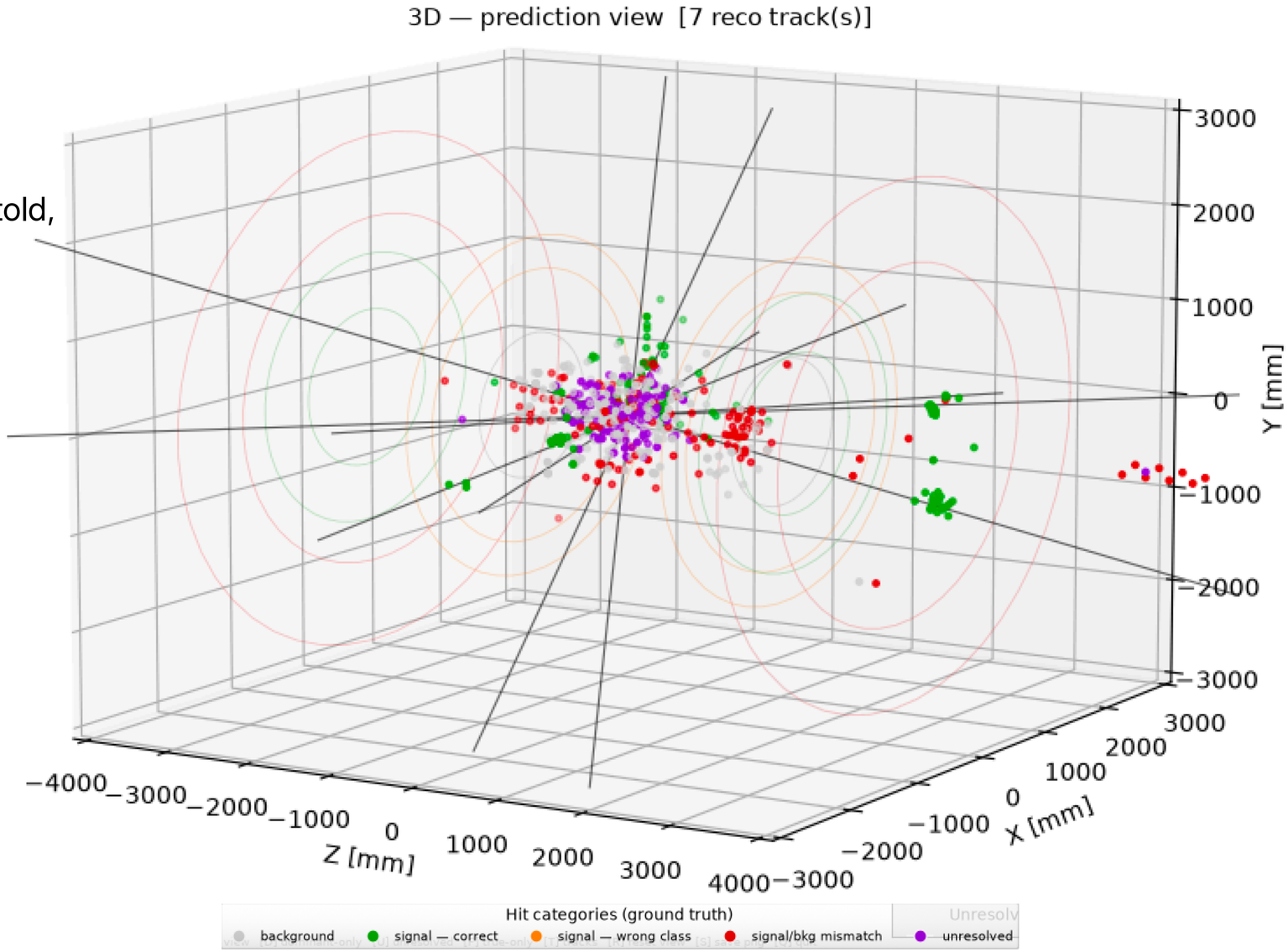
Unresolved class issue

"Unresolved" = a hit whose true origin the model can't be told, for one of two reasons:

- No truth link at all
- Undocumented generatorStatus gap [7000,10000]

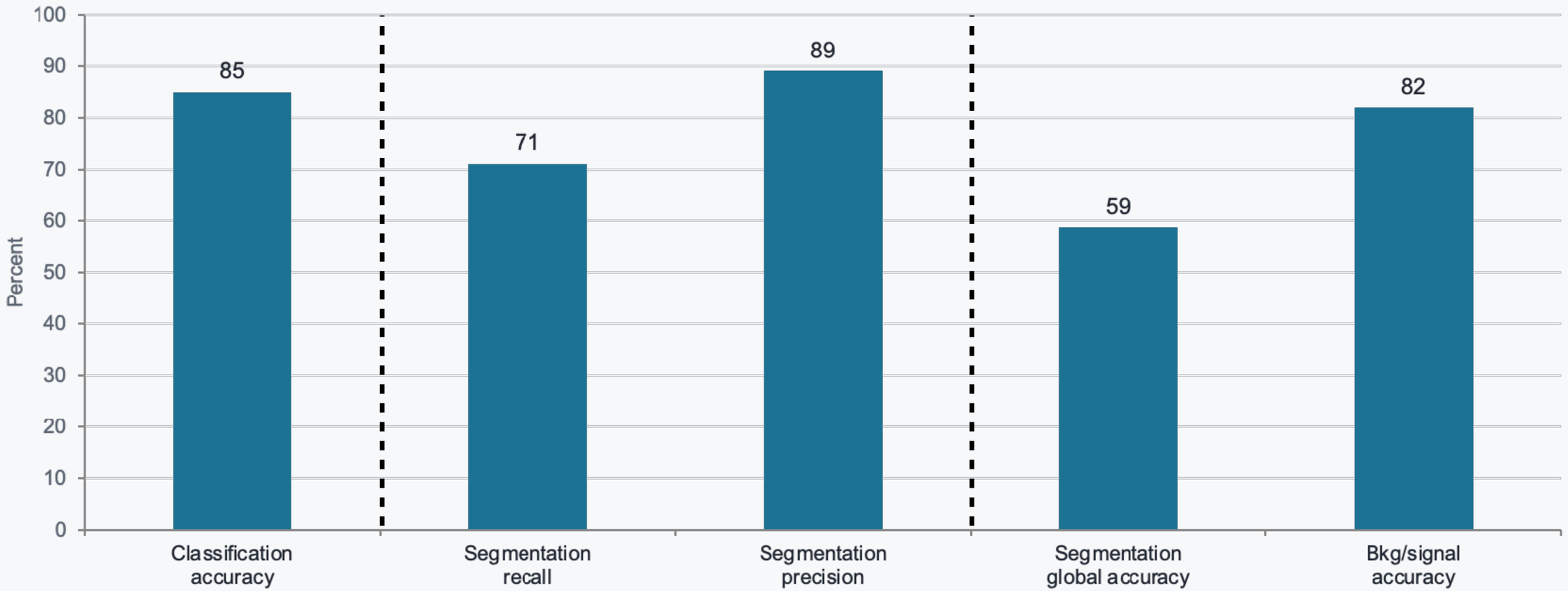
Potential root cause:

Integer overflow or signed/unsigned conversion error in physics generation software



Issue being discussed on GitHub: https://github.com/AIDASoft/DD4hep/issues/1657#issuecomment-5280917973

Current reported baseline



Caveat: these come from a single class, not an aggregate result across the full 27-class campaign.

A ground-truth problem was found

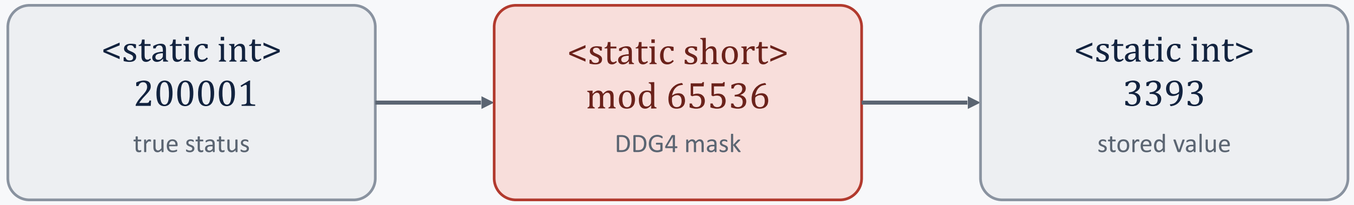
— Class labels come from generatorStatus ranges

0–99999 = background, 100000+ = physics class

— DDG4 stores this status in a 16-bit field

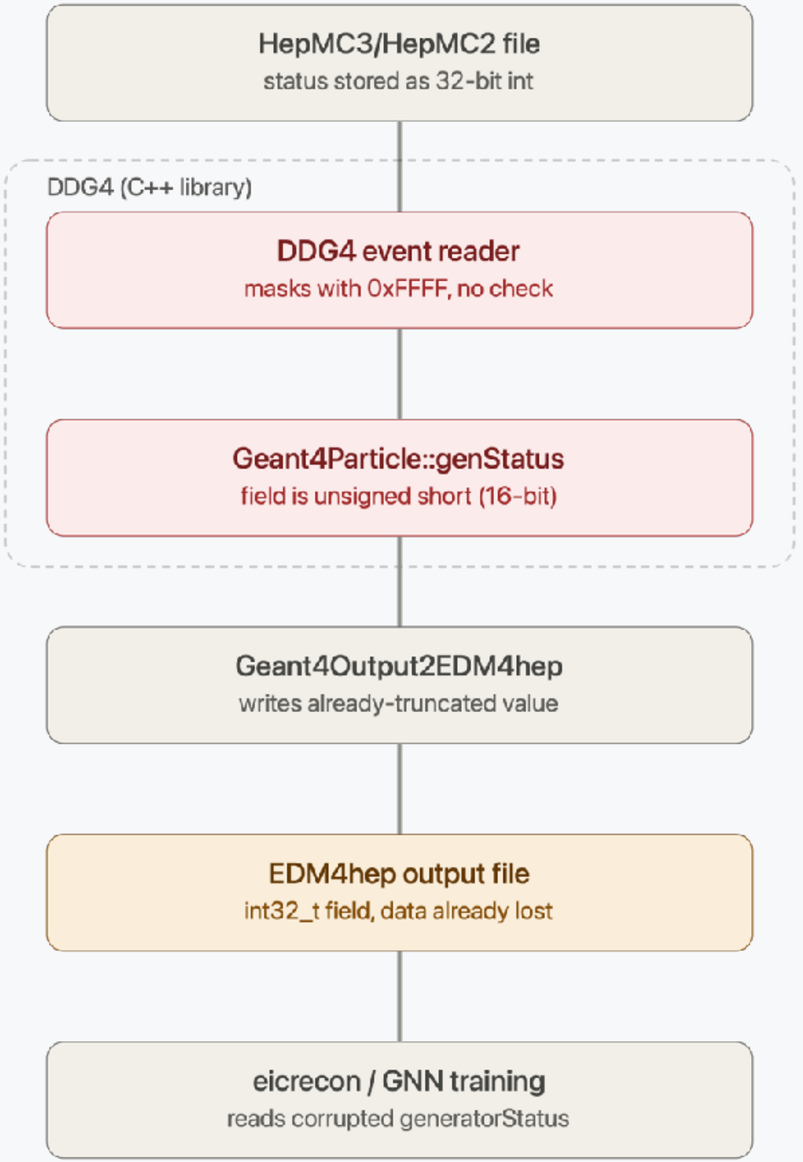
max value 65535, no bounds check

— Every physics-class label is guaranteed to truncate



ncdisq10 (a real physics event) truncates into the background numeric range

Where in the stack it happens



The two red boxes are the only 16-bit storage in an otherwise 32-bit pipeline — that's the whole bug.

The bug is invisible past this point — the file format is correct, only the value inside it is wrong.

Once the bug is fixed

1

Widen the field

Change genStatus to int32_t in DDG4 (the issue's suggested fix)

2

Regenerate & retrain

Rerun the simulation campaign, retrain from scratch — don't just patch downstream and reuse old weights

3

Add a regression test

Turn the GitHub issue's 3-step repro into a standing CI check so this can't silently reappear

Standard Event Builder and ML validation

Event-level splits

Train/val/test split by event, not by class — hit-level splits leak information. One hit may be shared across multiple events

Per-noise confusion matrix

Not just aggregate accuracy — rare classes can hide behind background dominance

Realistic pileup stress test

Validate at ~5MHz pileup, not just the easier single-event regime

Baseline benchmark

Compare against ACTS + Calo Island performance — already flagged as the pending next step; possibly turning ON/OFF factories TF->TF upstream

ONNX consistency check

Confirm CPU inference matches GPU-trained native model, especially through EdgeConv's dynamic kNN

Hit distribution

Confirm hit distribution: time, energy, positioning resolution verification

Backup slides

Hit distribution