Representation Learning for LHC Triggers

CERN openlab Summer Student Programme Lightning talks

Estefanía Moreno León

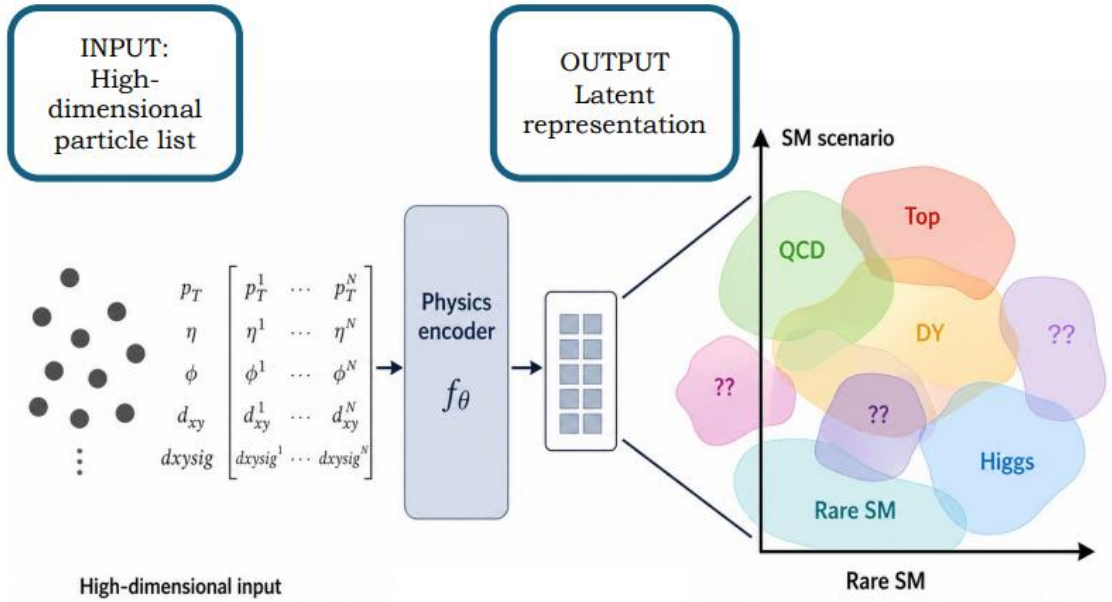
Supervisor: Maciej Mikolaj Glowacki and Sabrina Giorgetti

18/08/2026

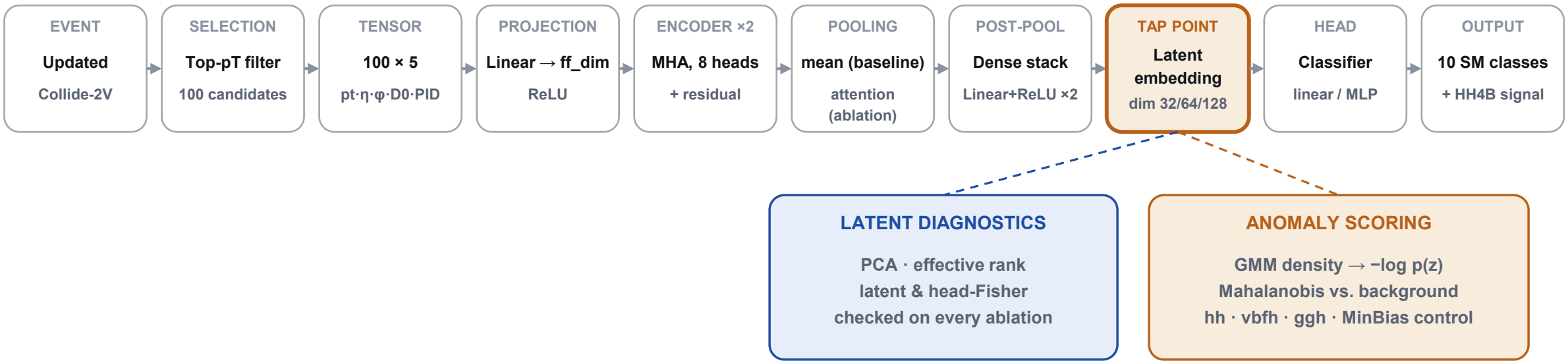
Motivation & Research Questions

NGT WP3 investigates whether ML can produce representations of LHC collision events that naturally cluster Standard-Model processes and remain sensitive to BSM signatures — and whether that capability follows predictable scaling laws.

- Goal: latent representations of LHC L1 trigger events, useful for both physics clustering and anomaly sensitivity.
- Paradigms in scope: supervised classification (today's evidence) → Autoencoder / BERT-style / VICReg / DINO / JEPA (planned).
- Three research questions anchor everything that follow:
  a. Which learning process yields the optimal representation of LHC collider data?
  b. How does representation quality evolve with model and dataset size?
  c. Can scaling laws quantify anomaly-detection capability with scale?



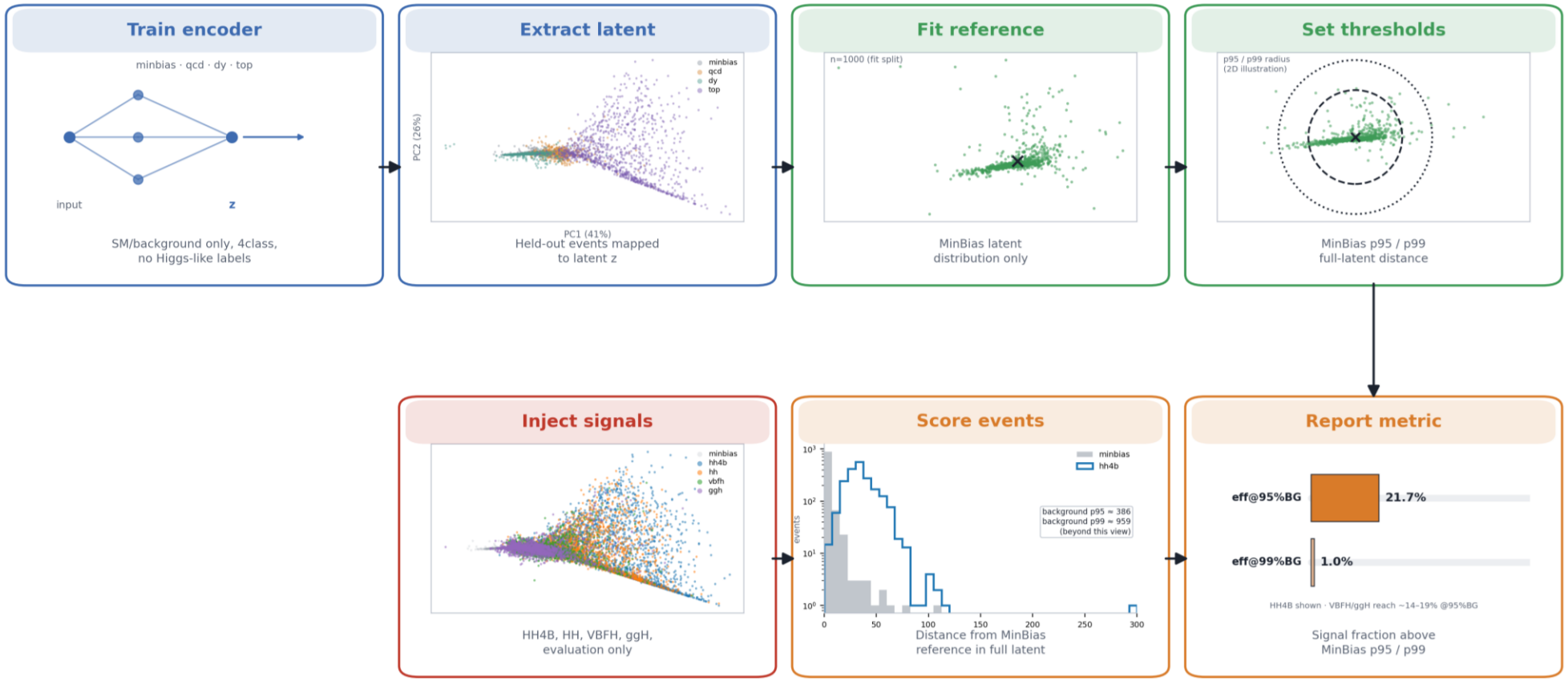
Dataset, Representation & Workflows



ABLATIONS SWEPT AT THIS TAP POINT

latent_dim 16 / 32 / 64 / 128 · pooling: mean / attention · probe: linear / MLP / MLP2
objective: CE / label-smooth / focal / SupCon / covariance / variance / VICReg-style
labels: 4class / 10plus_sm / granular cross-checks
scaling: 50k / 100k / 150k / 200k / 1M events per class + encoder depth/width

How the Anomaly Metric Is Computed

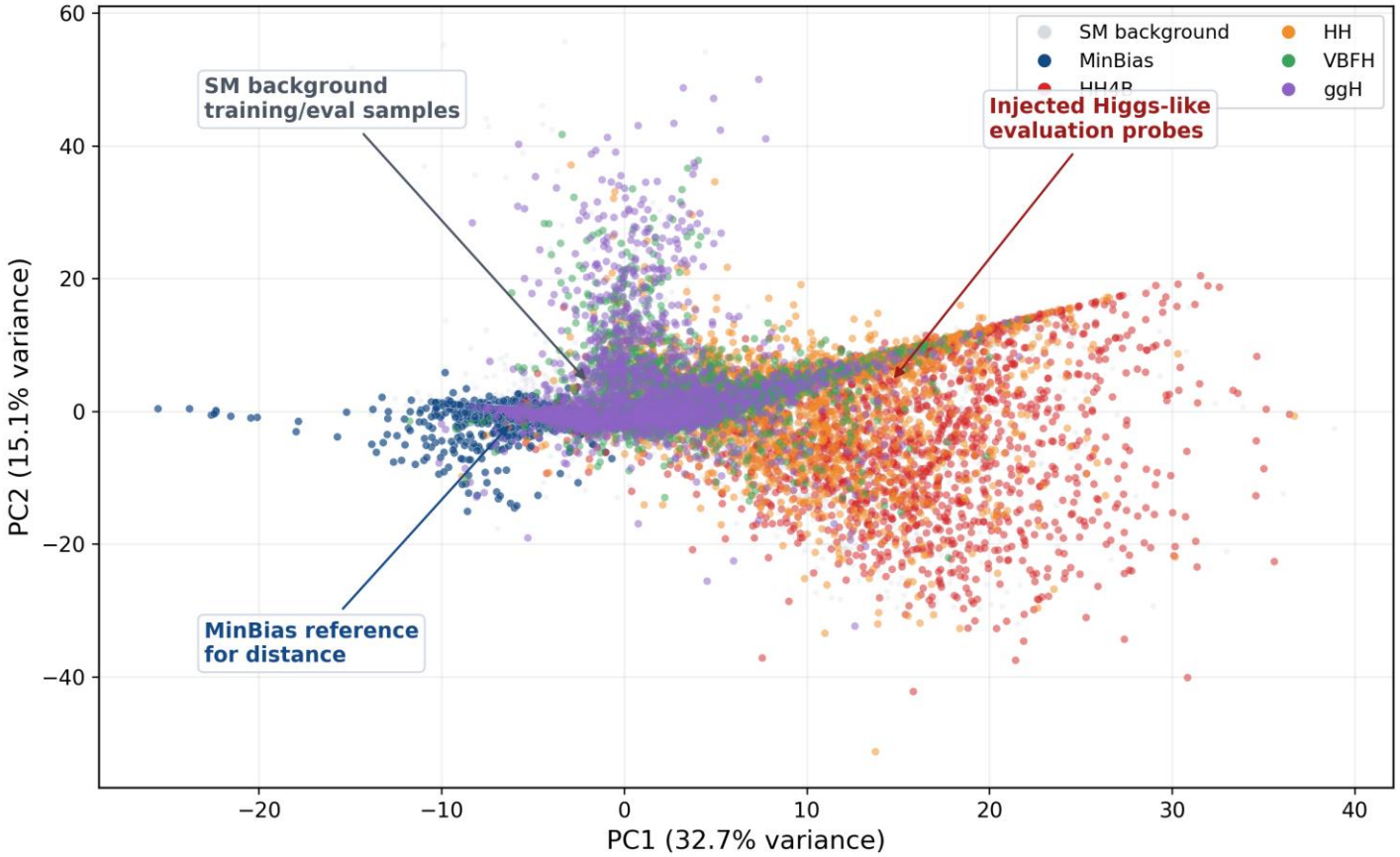


4class, 100k/clase, ff_dim=256, seed 42

Training class set changes by configuration; injected Higgs-like samples are evaluation-only, never training labels.

Why this matter?

A useful trigger representation should keep ordinary MinBias events close together while placing rare signal-like events farther away.



Real 4class 100k run. PCA is qualitative; MinBias-distance metrics are computed in full latent space.

Scaling Study: Two Independent Axes

CLASS DIVERSITY TESTED

Result: more SM granularity did not improve the current MinBias-vs-anomaly metric

4class — best current focus → **10plus_sm** — 11 classes → **granular_sm13** — 13 classes

Conclusion: fix 4class for scaling

ACTIVE SCALING AXIS: 4CLASS SAMPLE SIZE + ENCODER CAPACITY

Controlled setup: heads=8, linear probe, CE + latent var/cov regularization

Scaling varies input size and encoder feed-forward capacity

50k	100k	150k	200k	1M	next
ff=128 pilot	ff=256	ff=256	ff=256 / ff512 comparison	ff=256 / ff512 lr=3e-5	distributed large-scale run

Injected-anomaly results use these 4class scaling configurations

full evaluation: HH4B, HH, VBFH, ggH | next plot focuses on HH4B efficiency | metric follows the MinBias-reference procedure

Scaling Sample Size & Model Capacity

After testing class diversity, we focus on the 4-class setup because it gives the clearest MinBias-vs-anomaly separation for the current objective.

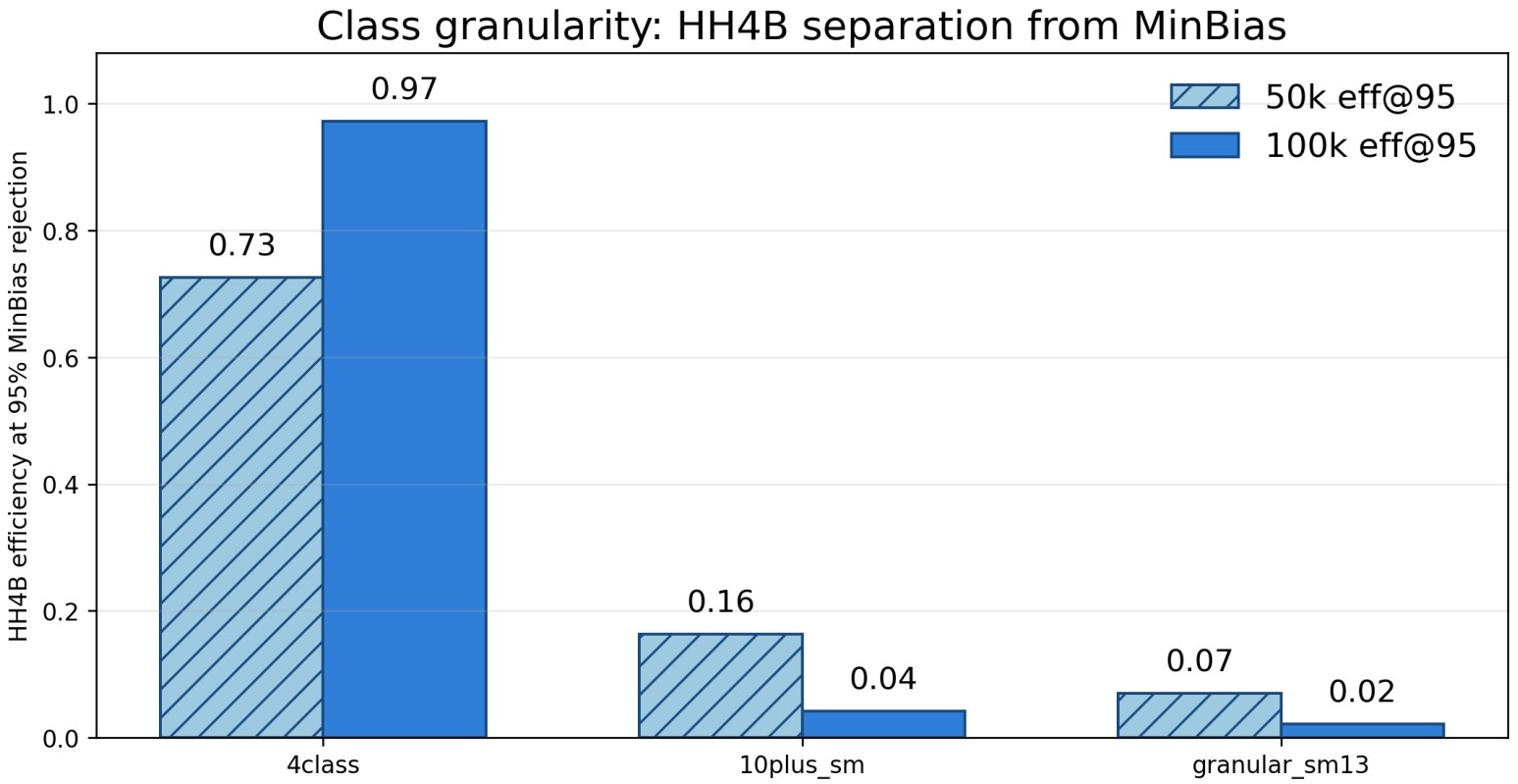
The next question is whether this separation improves with scale:- more events per class- controlled encoder capacity- fixed heads - fixed linear probe

Main metric: signal efficiency at fixed background rejection, especially eff@95%BG.

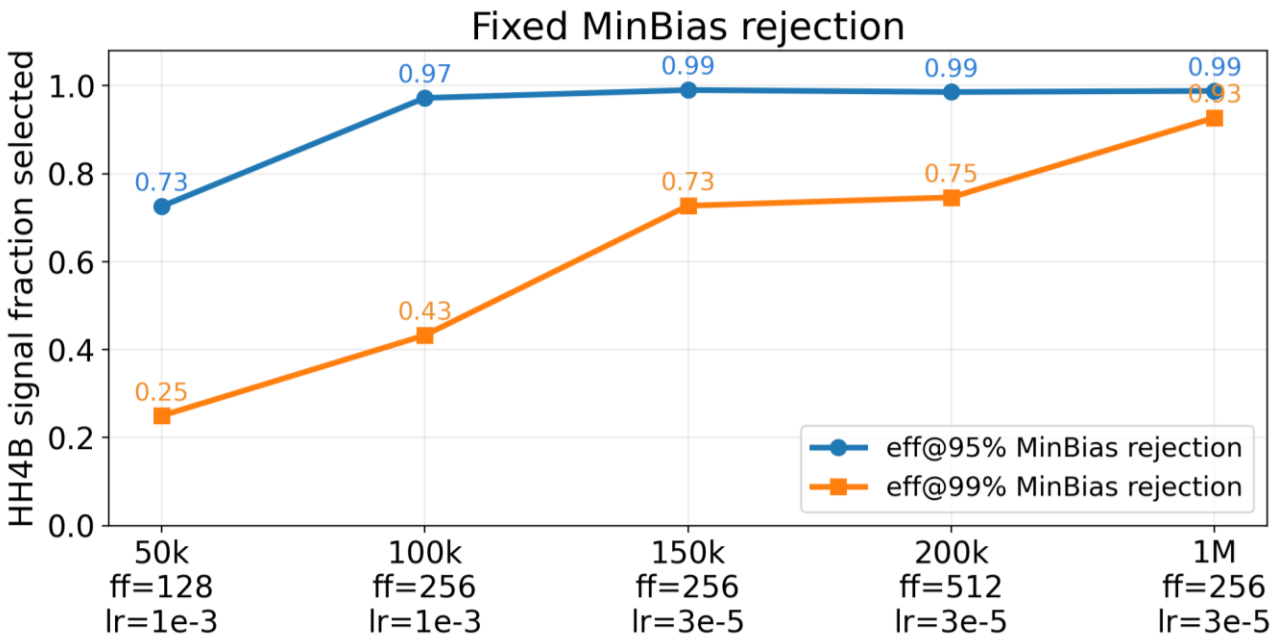
SETUP (LINEAR)	50K AUC	100K AUC	CONCLUSIONS
4class	**0.964**	**0.986**	best current MinBias-vs-HH4B separation
10plus_sm	0.898	0.852	more diversity did not improve this metric
granular_sm13	0.888	0.792	finer granularity did not improve this metric

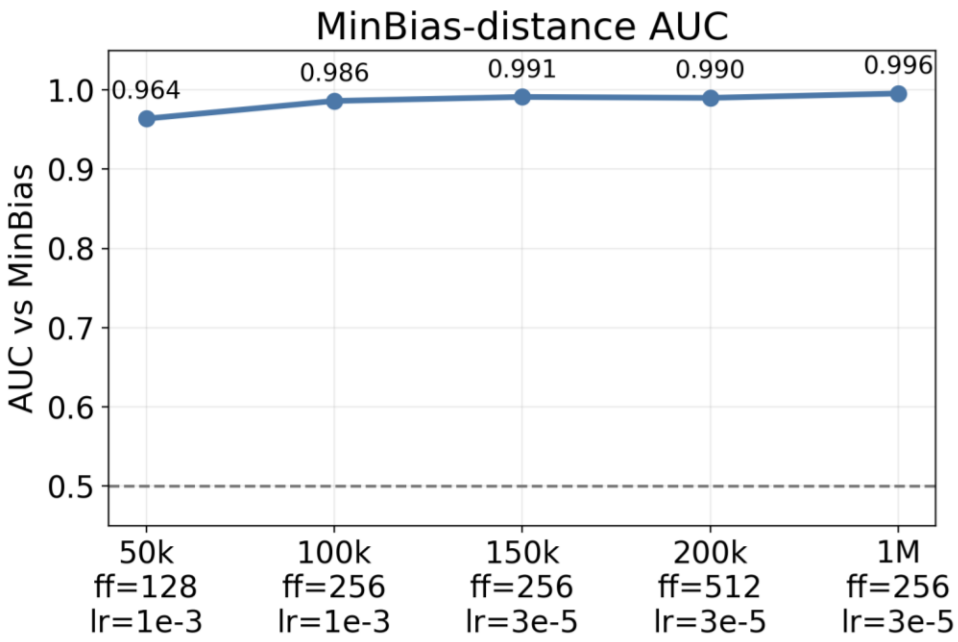
SAMPLES/CLASS	LATENT_DIM	FF_DIM	ATTENTION BLOCKS	HEADS	PROBE	LEARNING RATE
50k	64	128	2	8	Linear	Default
100k	128	256	2	8	Linear	Default
150k	192	384	3	8	Linear	Default
200k	256	512	4	8	Linear	Default
1M	128	512	2	8	linear	3e-5

How We Measure Representation Quality



How We Measure Representation Quality





Thanks

CERN openlab Summer Student Programme Lightning talks

Estefanía Moreno León

Supervisor: Maciej Mikolaj Glowacki and Sabrina Giorgetti

18/08/2026

Backups

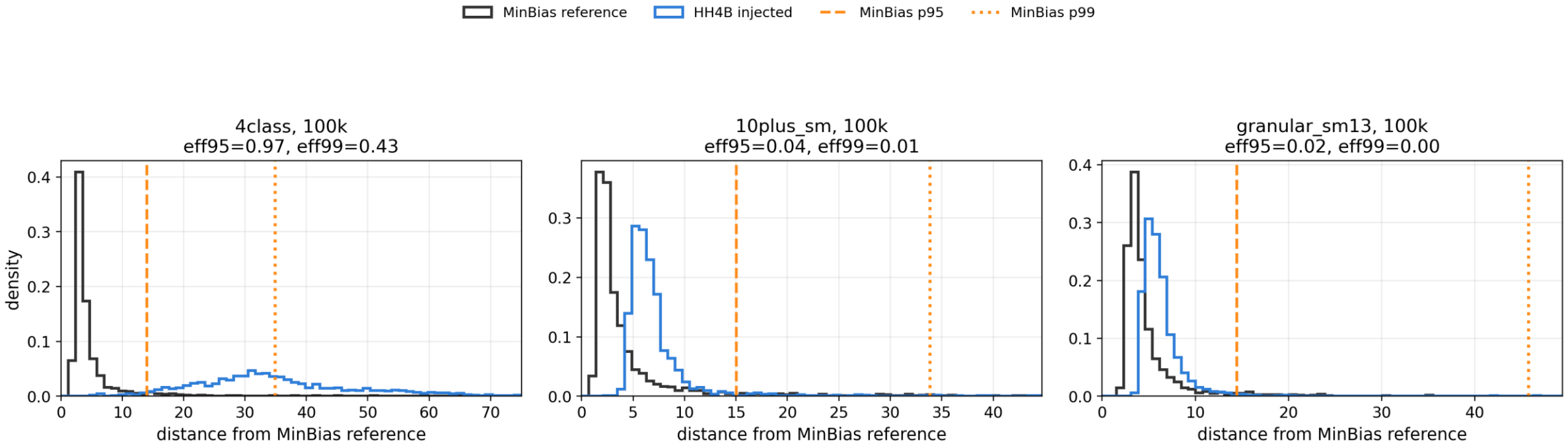
CERN openlab Summer Student Programme Lightning talks

Estefanía Moreno León

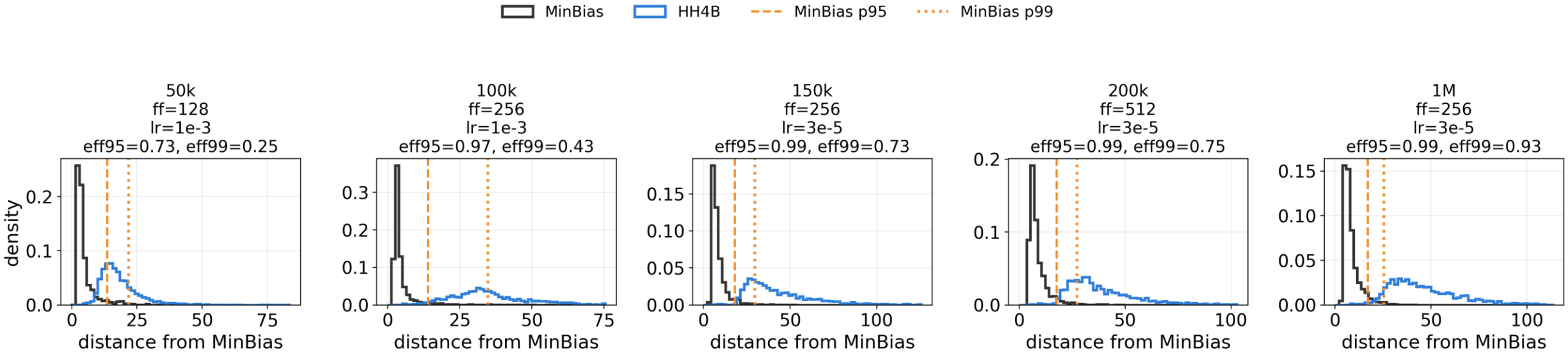
Supervisor: Maciej Mikolaj Glowacki and Sabrina Giorgetti

18/08/2026

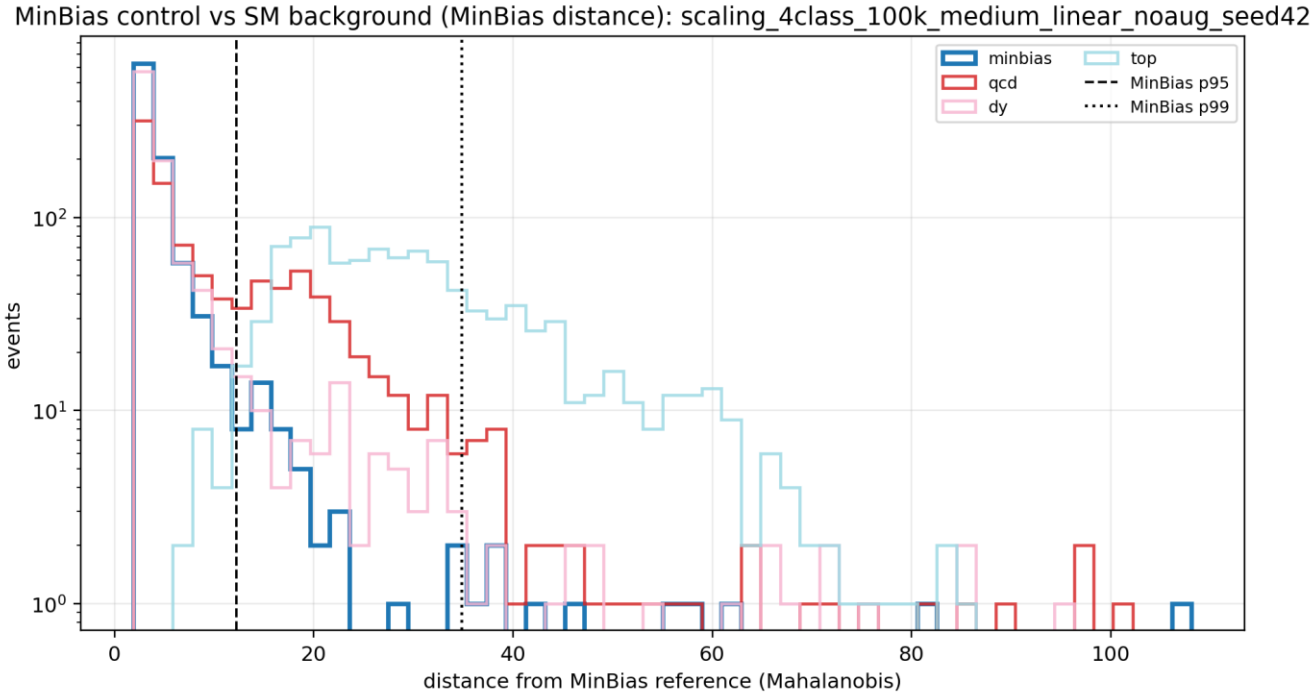
MinBias-reference distance: HH4B vs class granularity

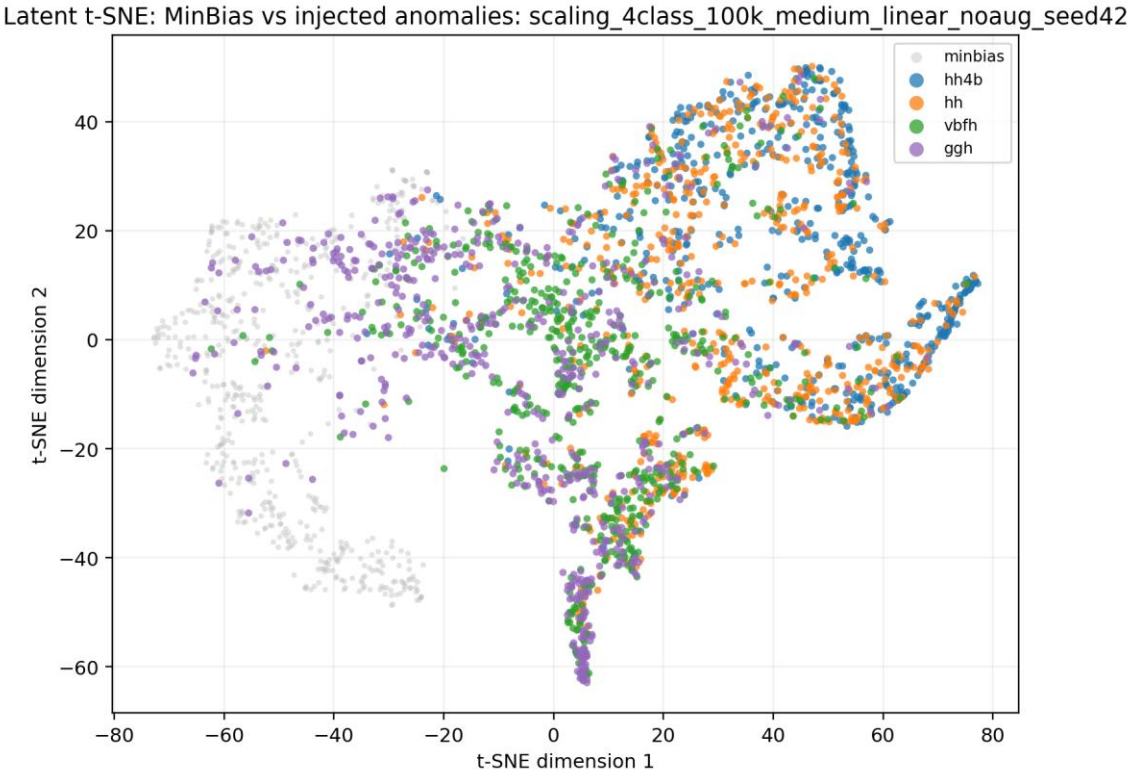


MinBias-reference distance: HH4B vs 4-class input size, no augmentation



Distances are computed in full latent space. X-axis is clipped at 99.5% combined quantile for readability.


MinBias control vs SM background (MinBias distance): scaling_4class_100k_medium_linear_noaug_seed42
minbias
qcd
dy
top
MinBias p95
MinBias p99
events
distance from MinBias reference (Mahalanobis)


Latent t-SNE: MinBias vs injected anomalies: scaling_4class_100k_medium_linear_noaug_seed42
minbias
hh4b
hh
vbfh
ggh
t-SNE dimension 2
t-SNE dimension 1