Nikhef

Radboud University

ATLAS EXPERIMENT

Photon identification in ATLAS: a revolution

Luca Franco

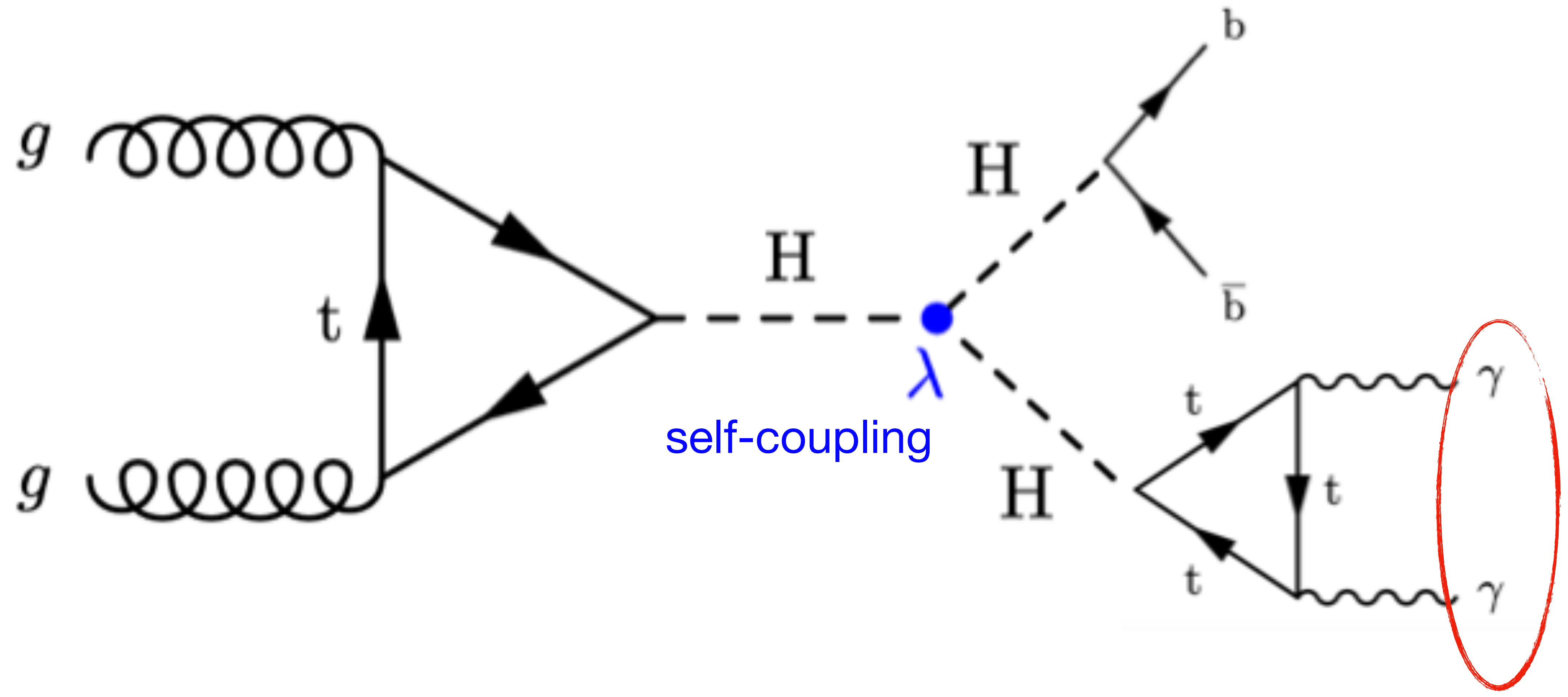
Nikhef ATLAS weekly meeting - 02/05/2025

A few remarks

- This project is done in the ATLAS combined performance EGamma group
- I am collaborating with the current EGamma photon identification (ID) convener (Mohamed Belfkir, UAEU)
- Much of what I will show today is very WIP, with mostly ideas and plans
- For some things I will follow a “pedagogical” order, rather than a chronological order

Motivation

- HH→bbγγ analysis
  - Current Photon ID efficiency ~85%
  - 100% Ph ID → 15% more events → ~7% larger HH significance
- H→Zγ analysis: ~80%
- All analyses with photons which are statistically limited

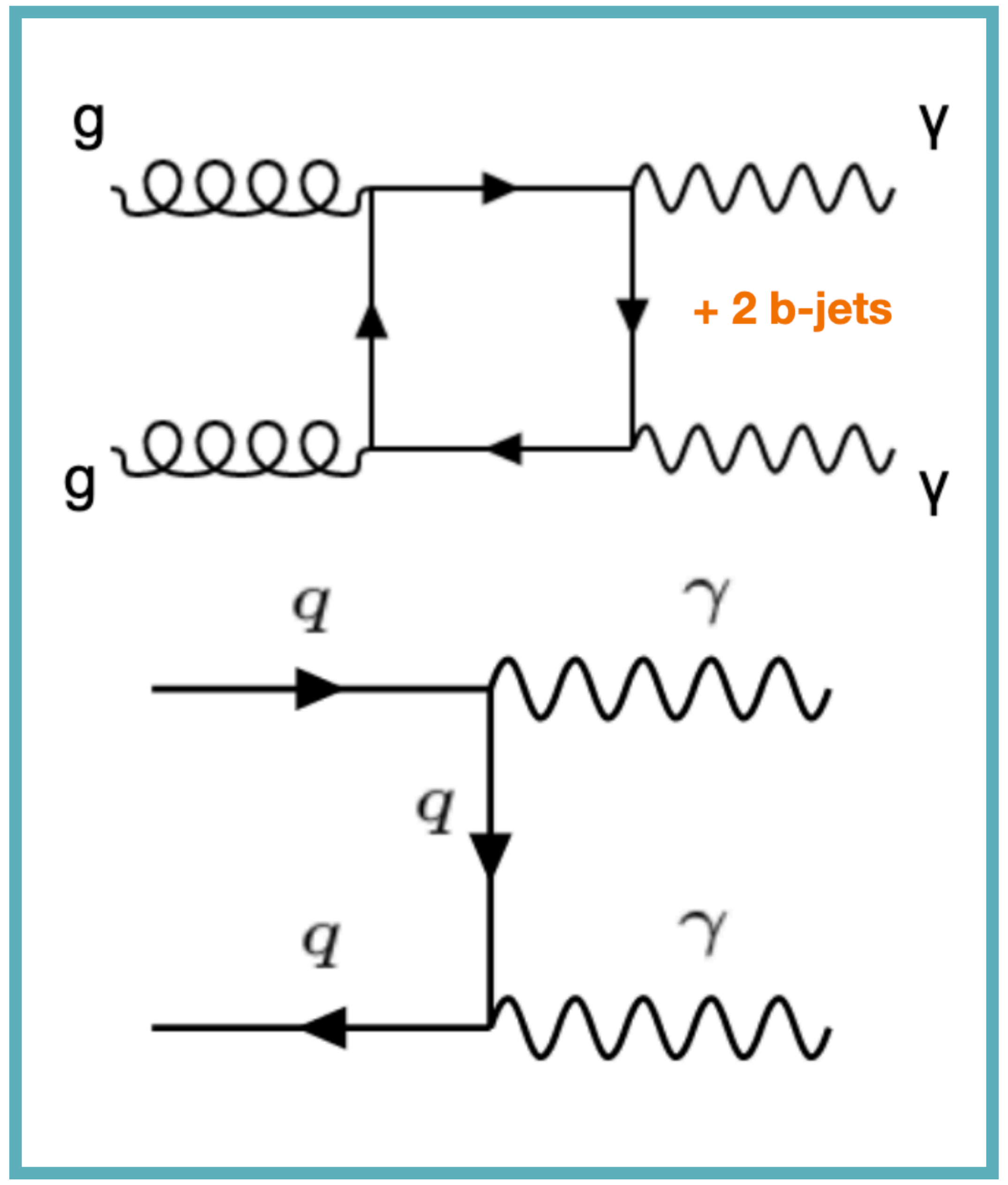


Cutflow of the Run2 HH→bbγγ analysis

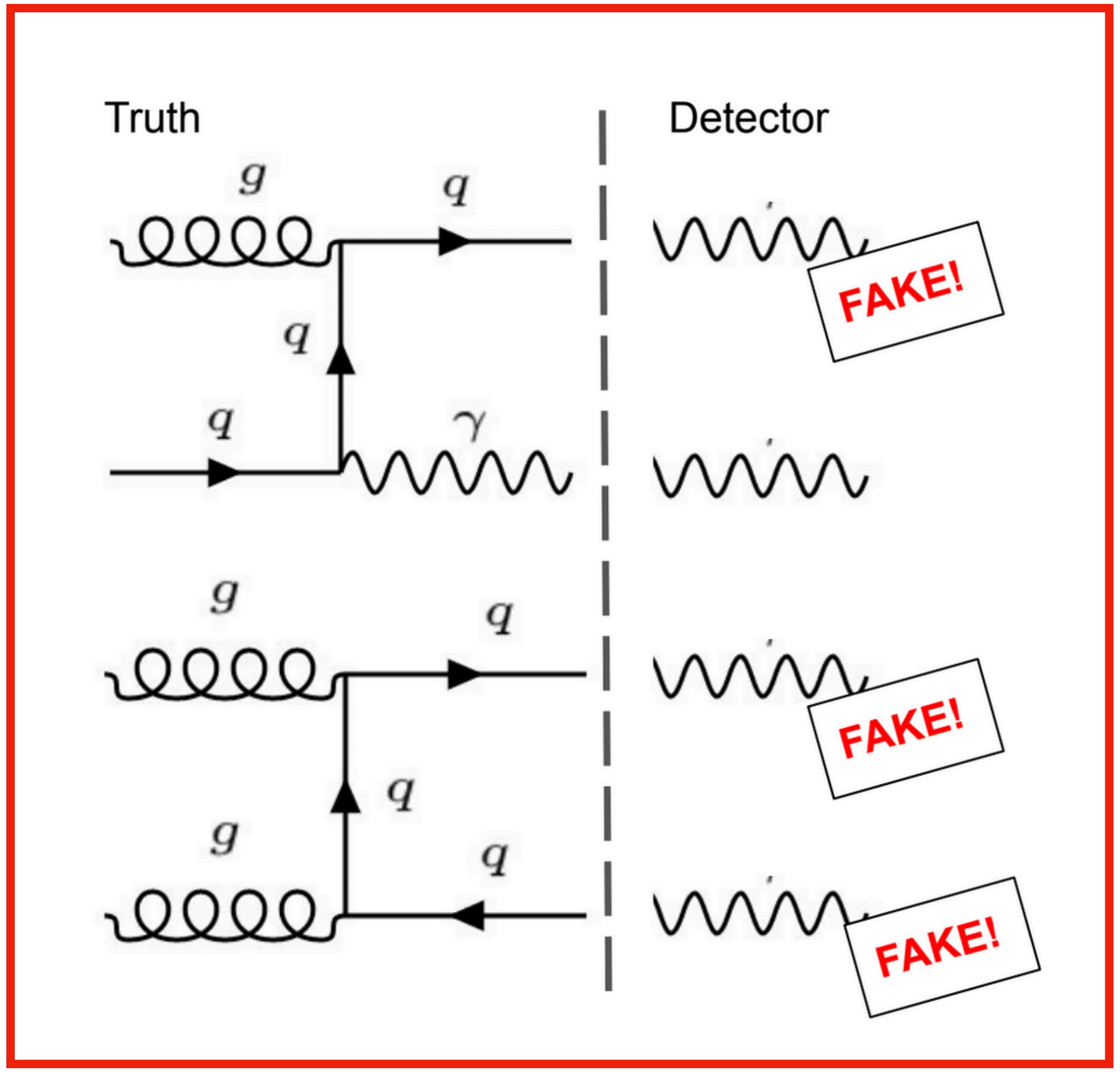
Cuts	SM ggF HH Yield	SM ggF HH Efficiency [%]
All events	12.600	100.00
Pass trigger	9.368	74.35
Has primary vertex	9.368	74.35
2 loose photons	6.875	54.57
$e-\gamma$ ambiguity	6.872	54.54
Trigger match	6.833	54.23
Photons tight ID cut	5.974	47.41
Photons isolation cut	5.332	42.31
Rel. p_T cuts	4.788	38.00
$m_{\gamma\gamma} \in [105, 160]$ GeV	4.782	37.95
$N_{\text{lep}} = 0$	4.759	37.77
$N_{\text{jets}} \geq 2$	4.221	33.50
$N_{\text{central jets}} < 6$	4.118	32.68
2 b-jets with 77% WP	1.628	12.92

We want to improve this selection cut

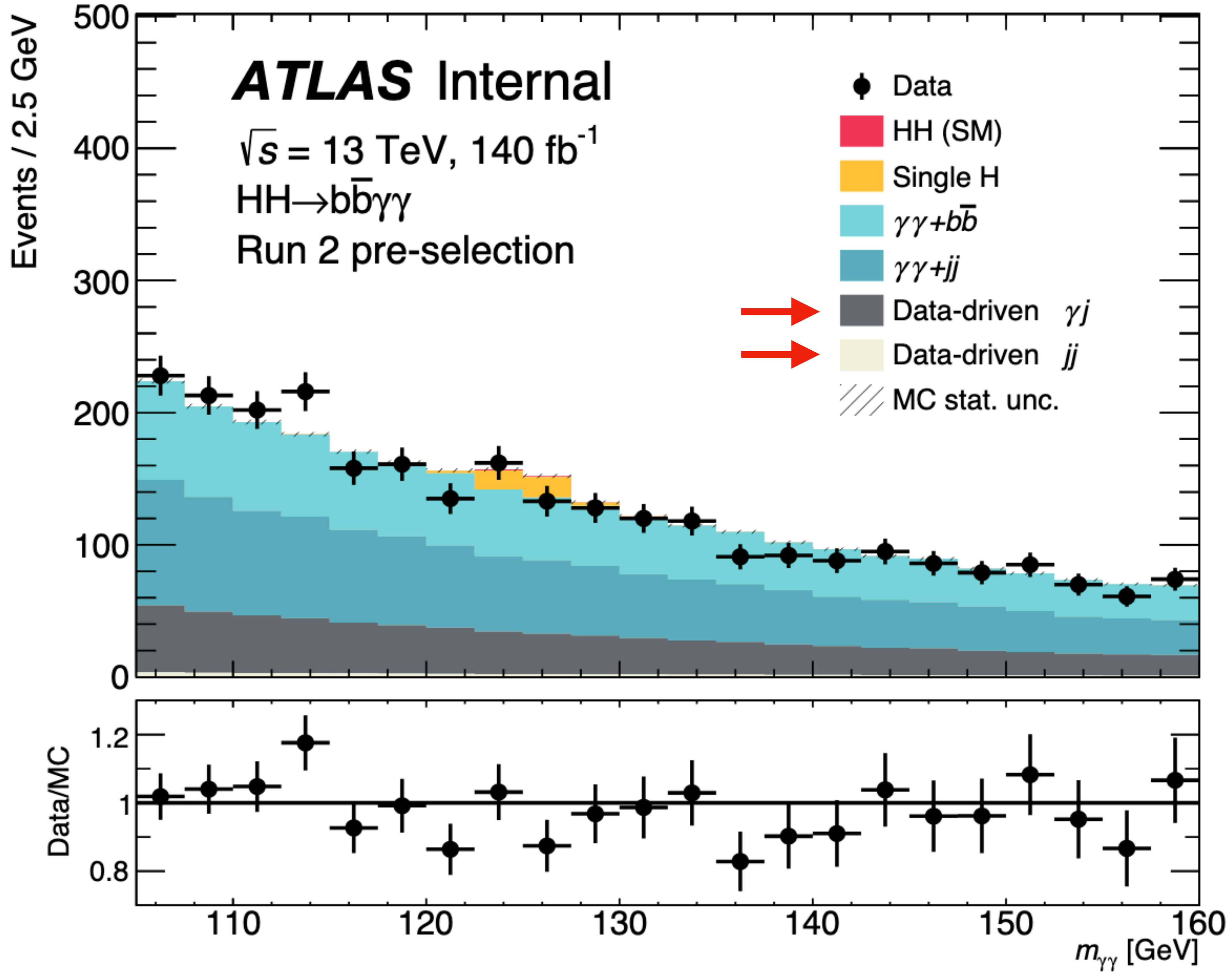
Jets faking photons



Irreducible γγ BKG

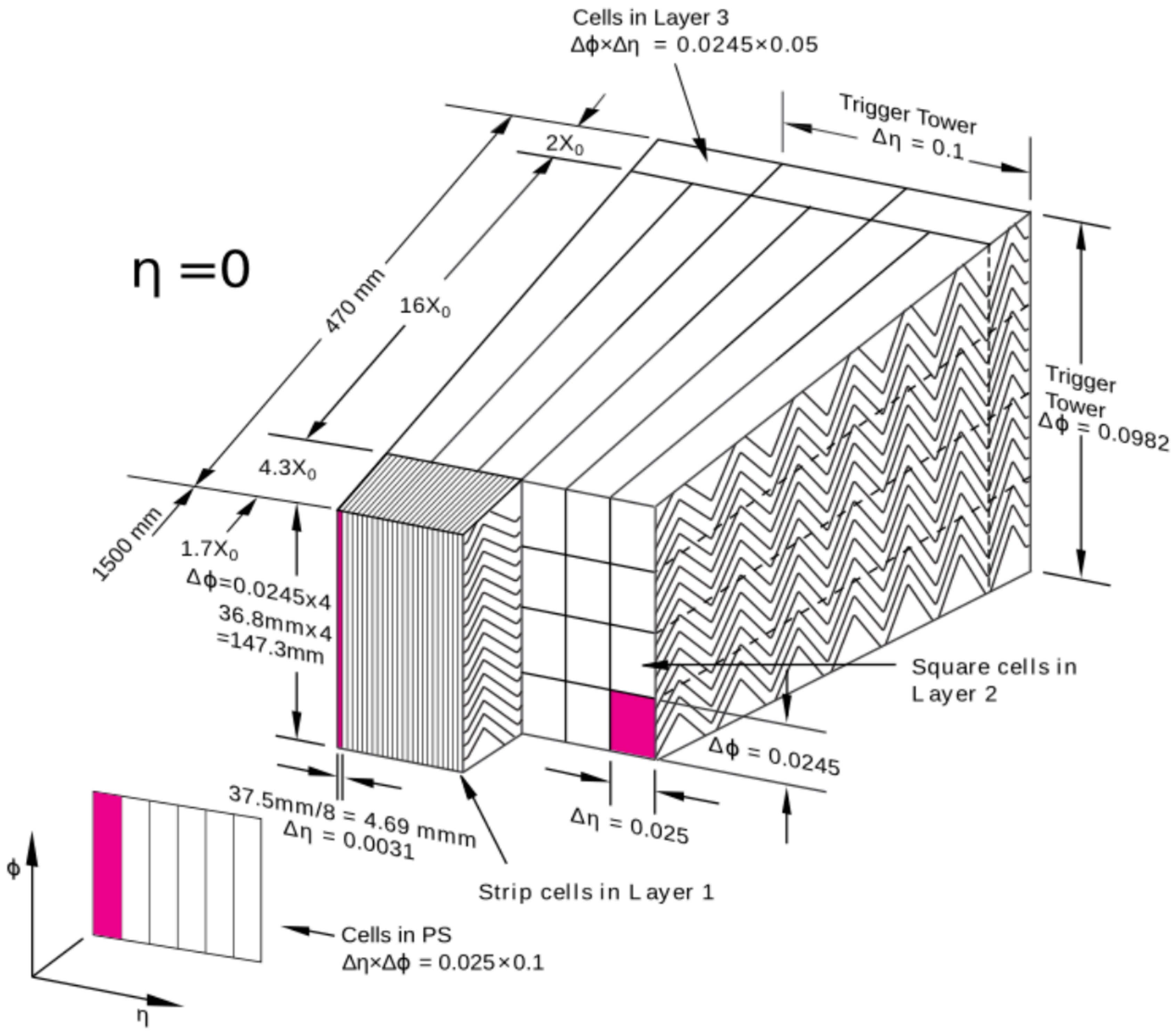


Reducible γj - jj BKG



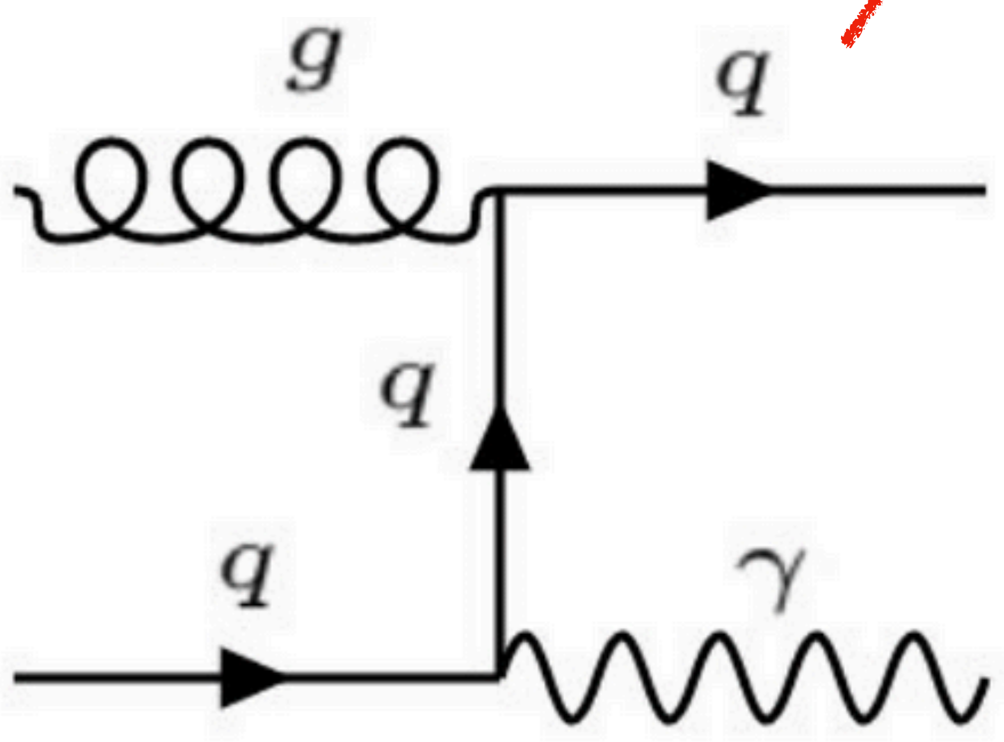
L. Franco - 02/05/2025

Photons in the detector

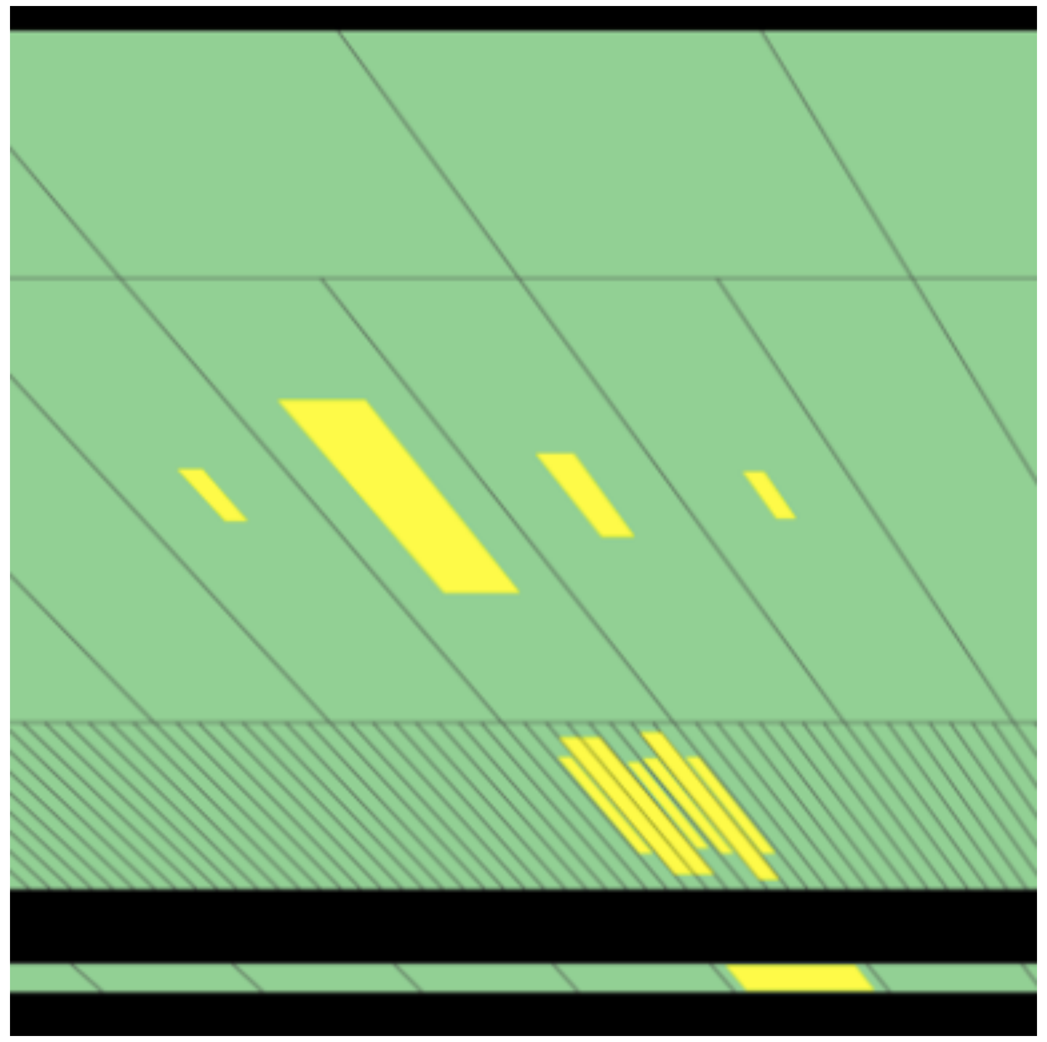


Section of the liquid Argon (LAr) electromagnetic calorimeter

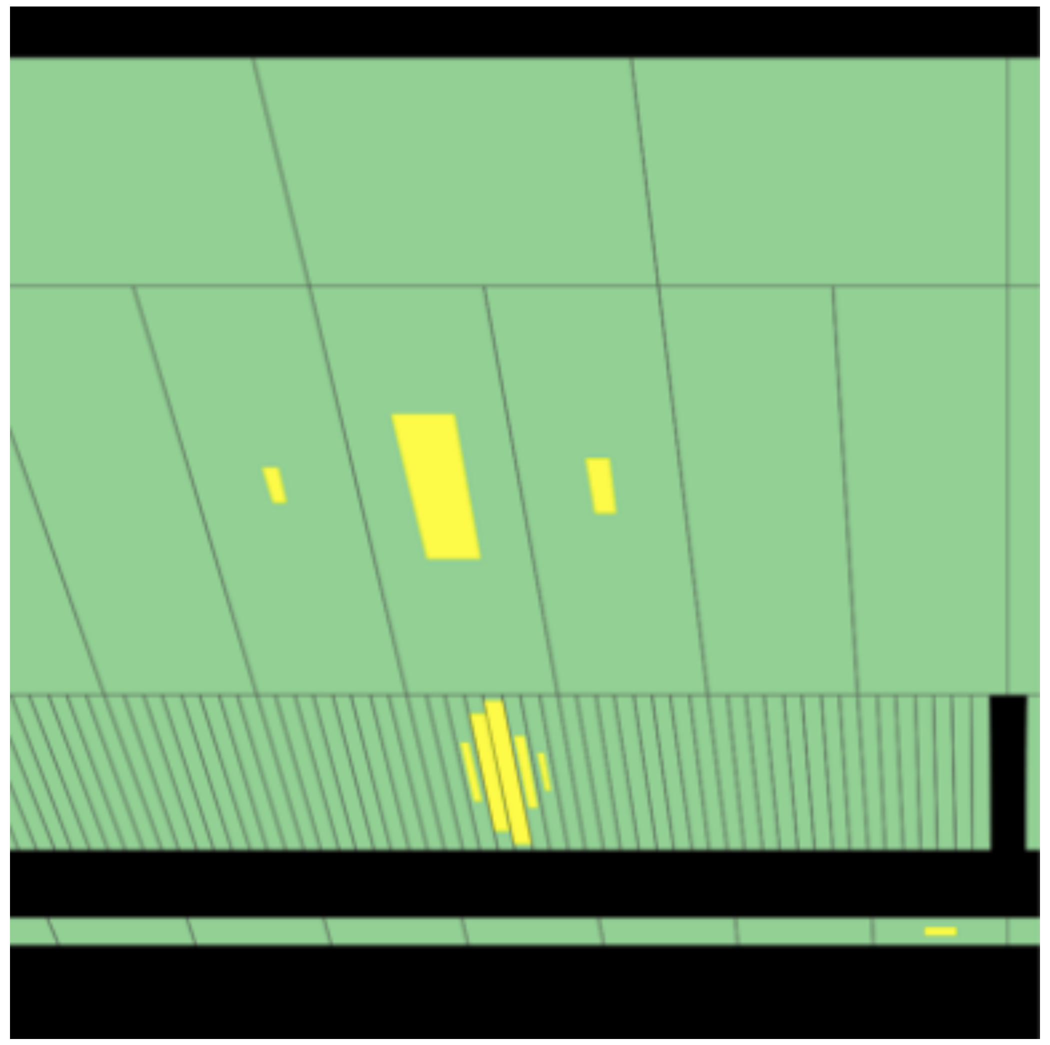
Quark from the **hard-scattering** will hadronize and form a jet with a π0



Photon from the **hard-scattering** will reach the calorimeter



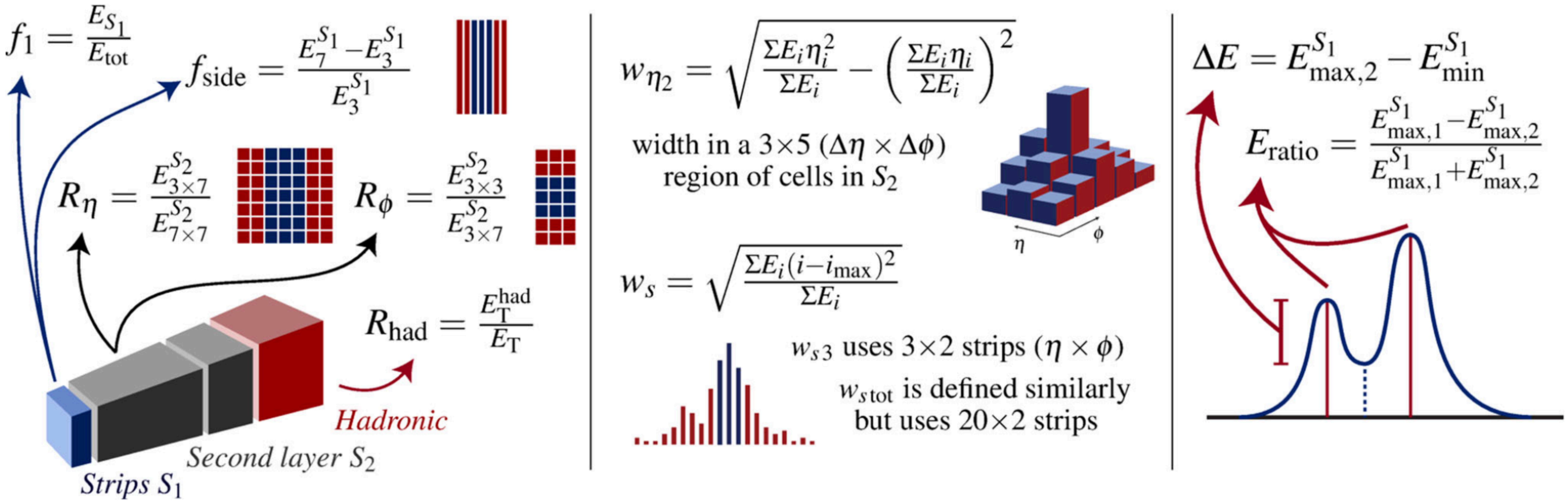
Non-prompt (fake) photon coming from π0→γγ



Prompt (real) photon in the EM calorimeter

Photon ID in ATLAS today

- Electromagnetic showers from real and fake photons have a different development in the calorimeter
- ~10 variables describing the properties of the *shower shape*

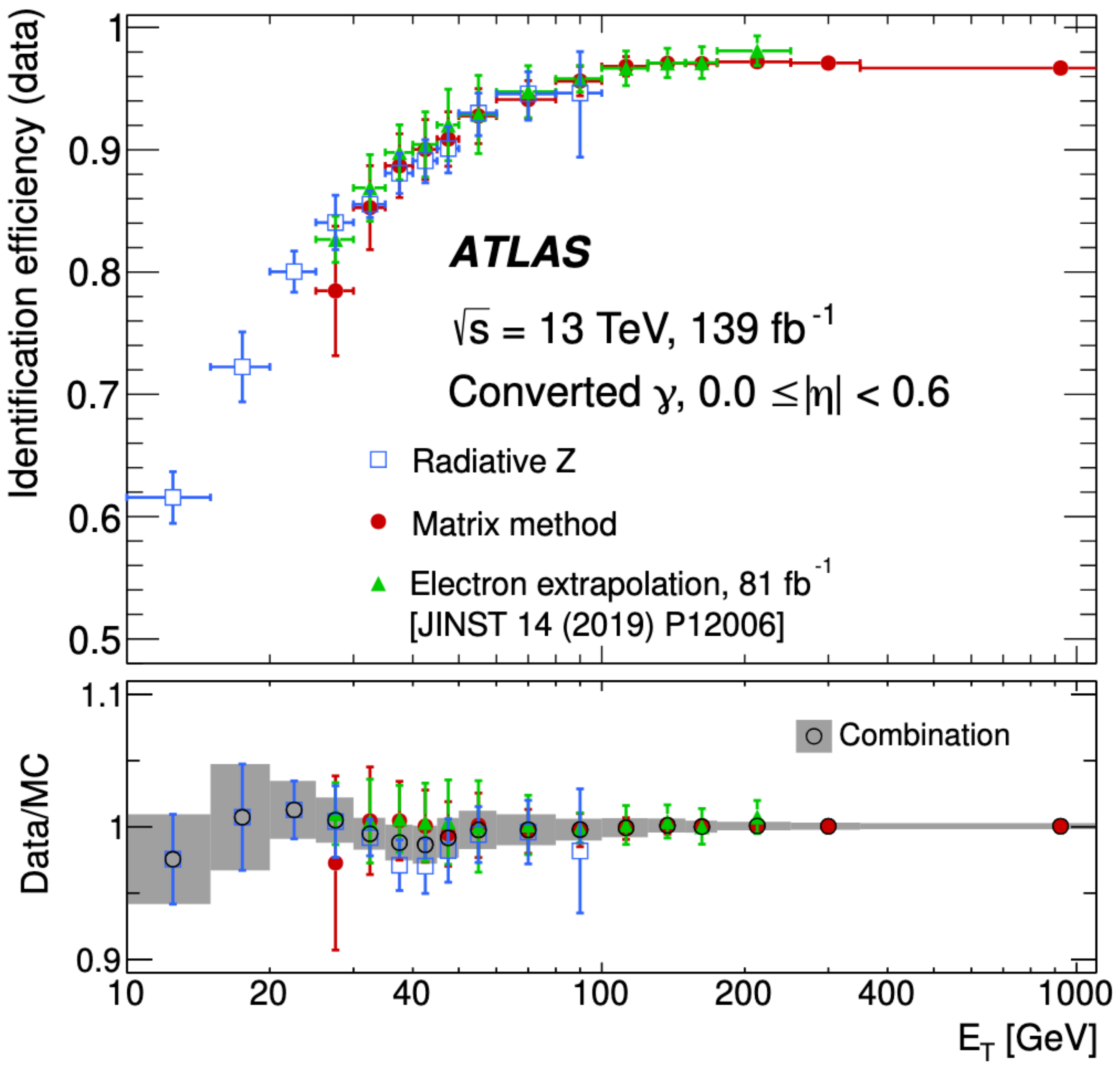


Photon ID in ATLAS today

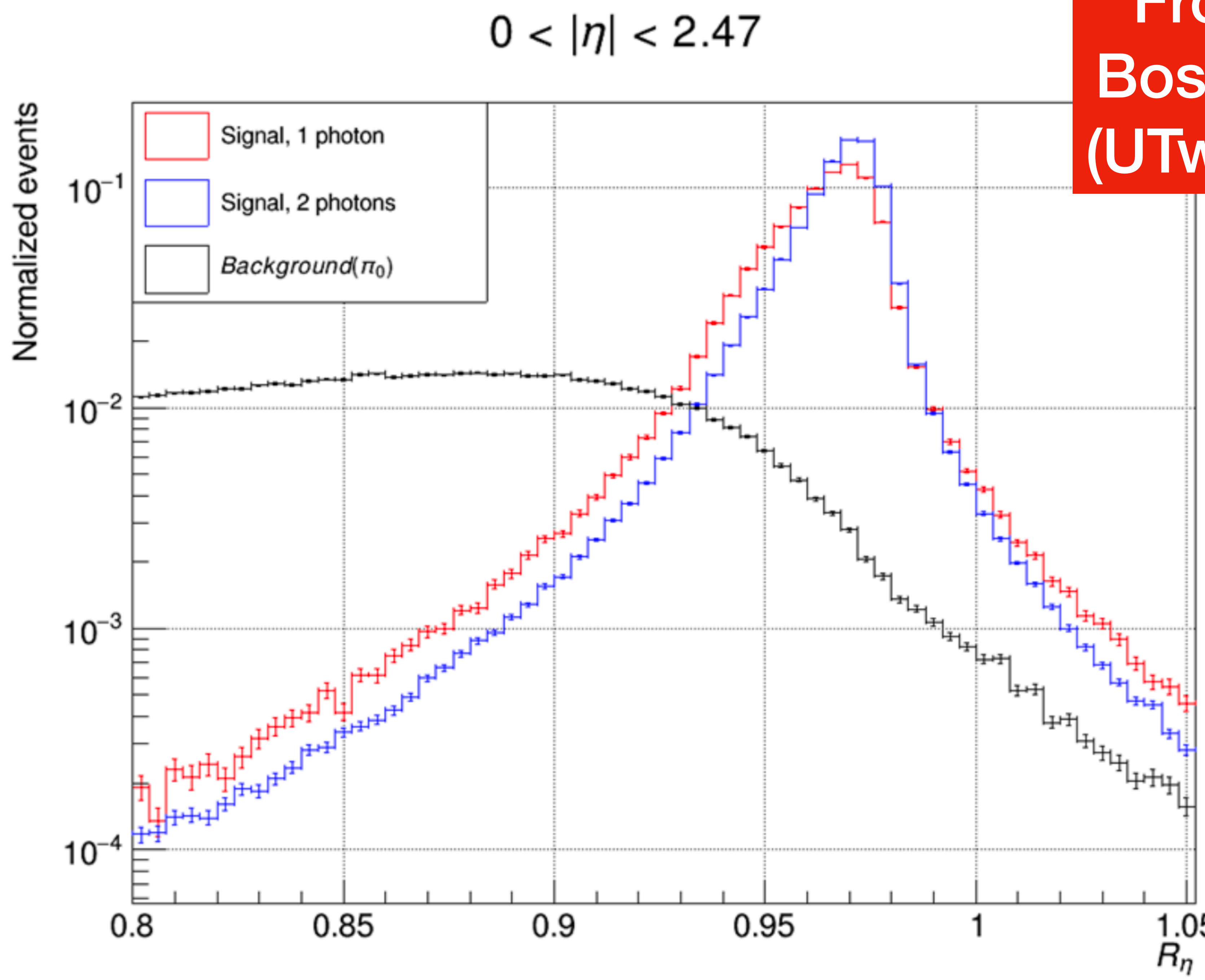
- These *shower shape* variables are used in a series of rectangular cuts (i.e. no Machine Learning)
- Optimization performed in different η-pT regions, separately for converted-unconverted photons

Eur. Phys. J. C 79, 205 (2019)

Efficiency at low pT can be as low as 60%



From Jasper Bos BSc thesis (UTwente, 2024)



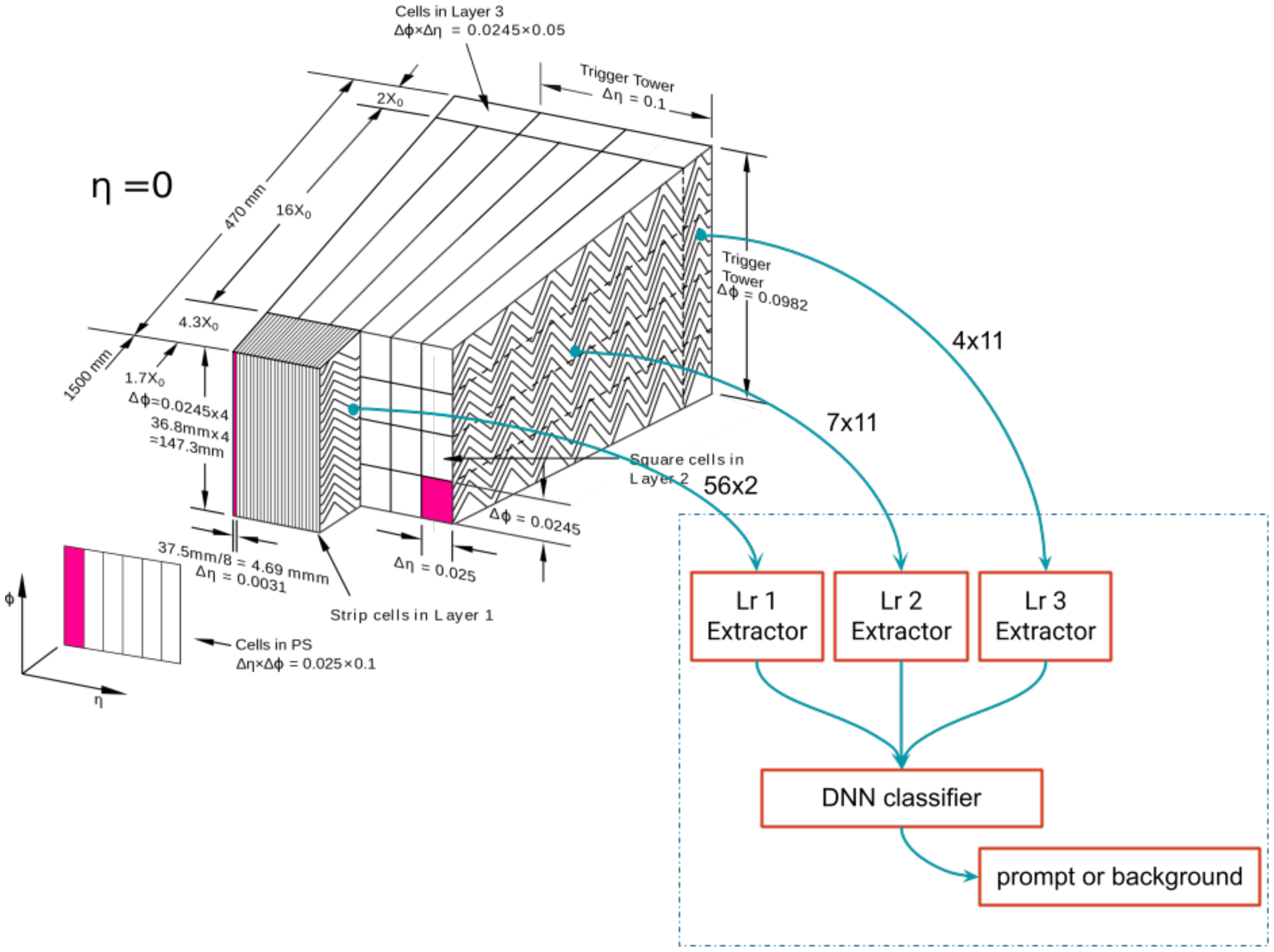
$$R_\eta = \frac{E^{S_2}_{3\times7}}{E^{S_2}_{7\times7}}$$

ηxϕ window

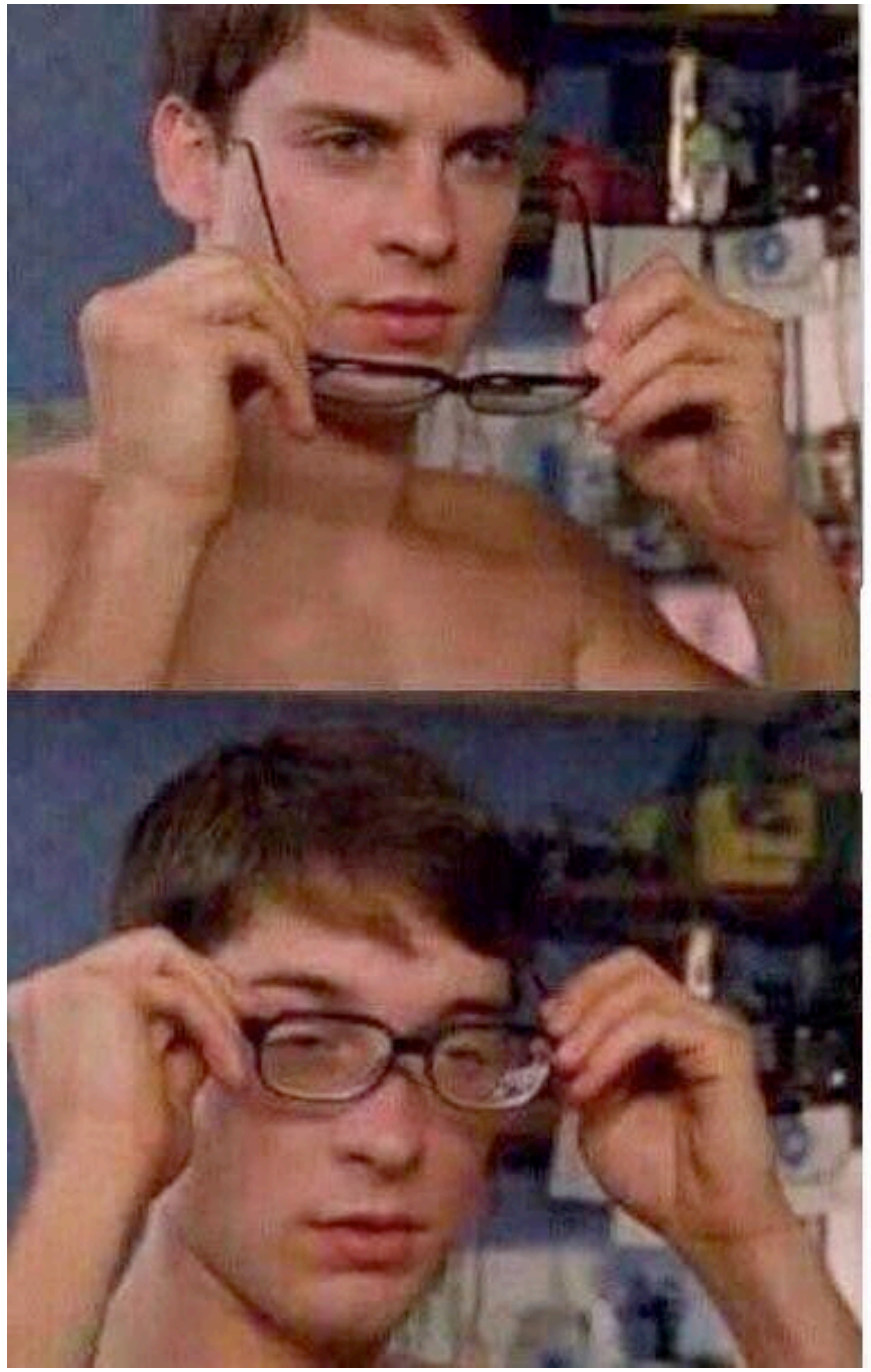
L. Franco - 02/05/2025

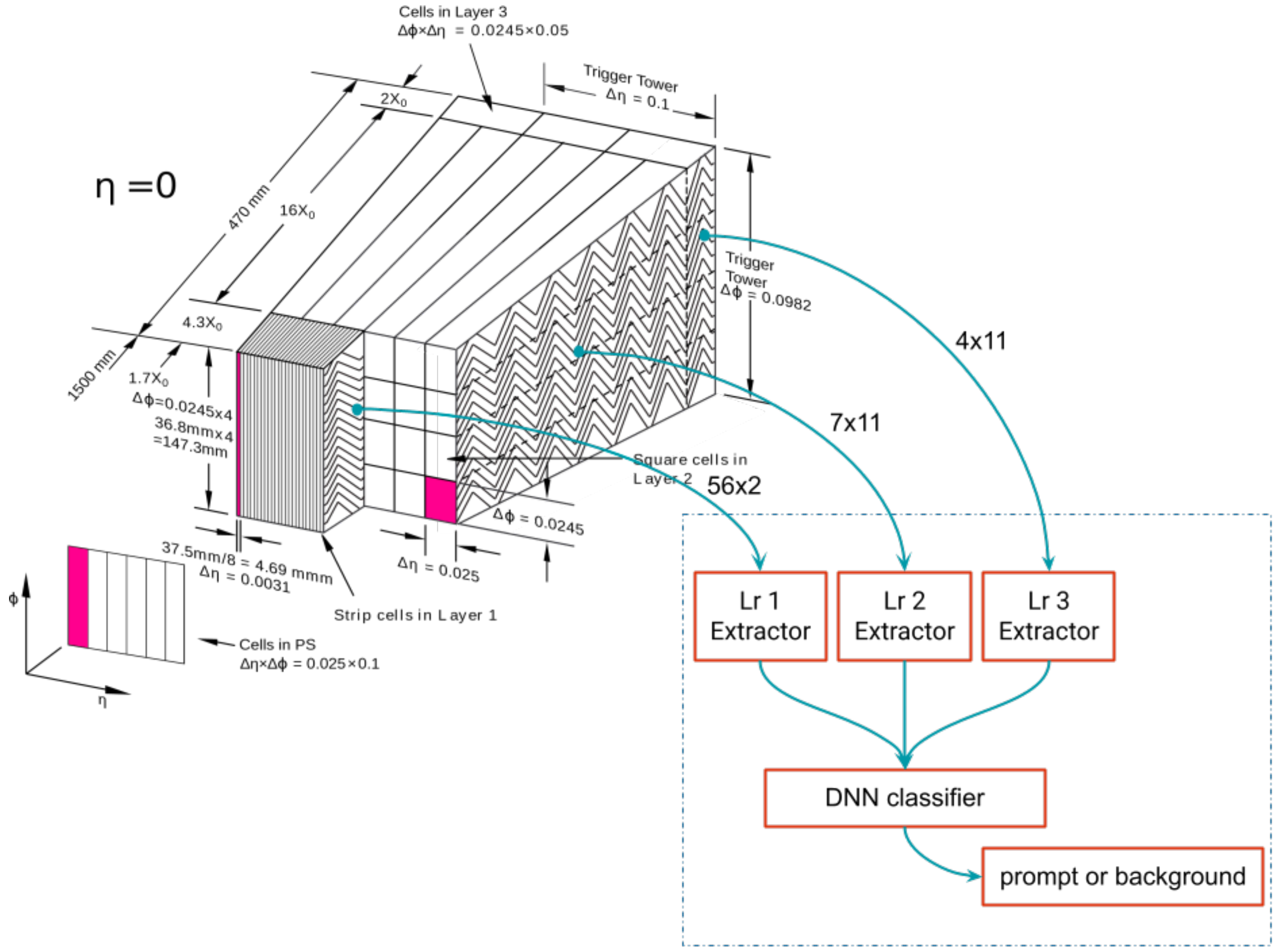
Photon ID in ATLAS tomorrow?

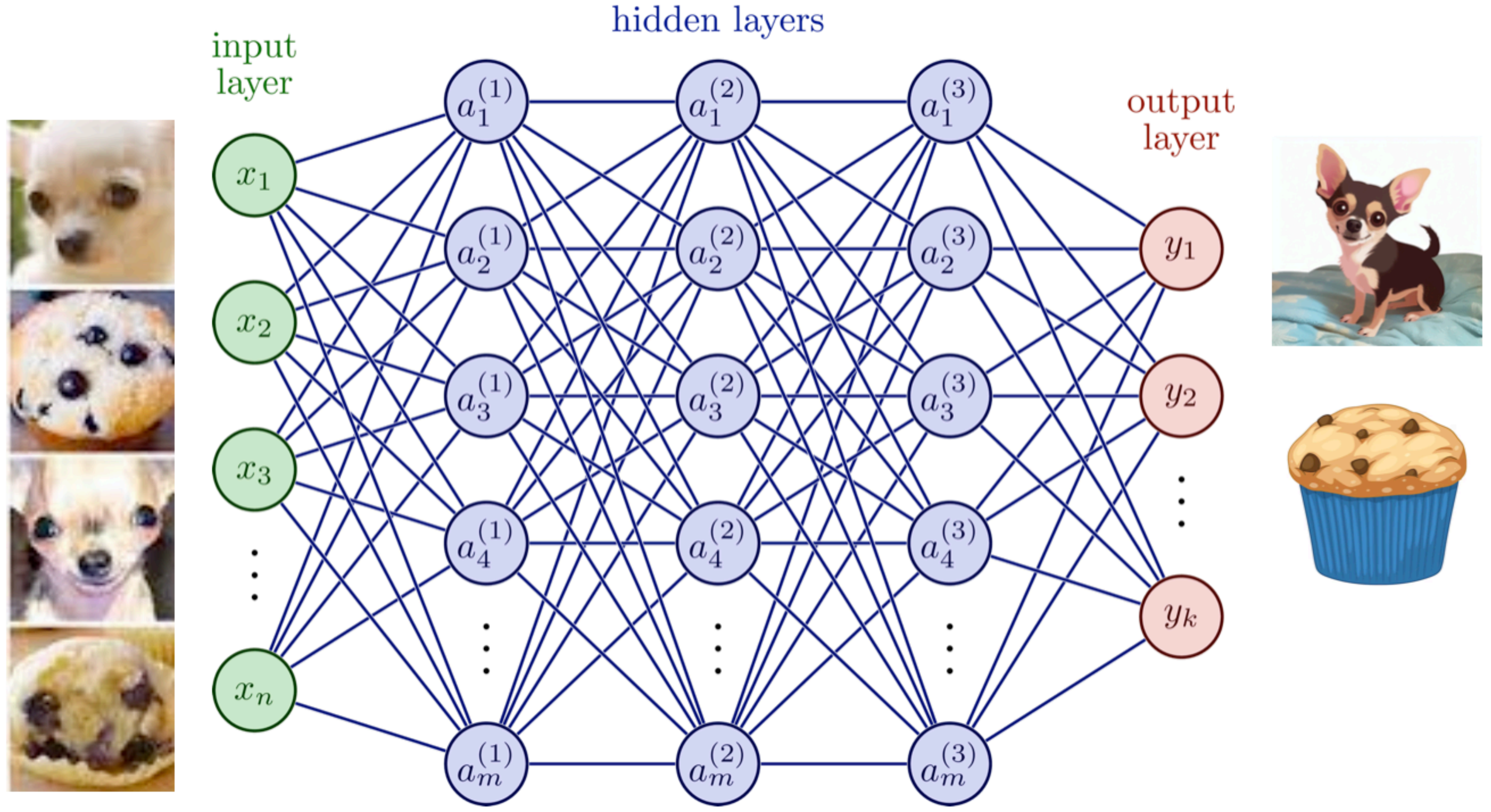
- Exploit Machine Learning
- Drop shower shape (high-level) variables
- Feed 2D (ηxϕ) energy clusters to a **convolutional neural network (CNN)**
- Low-level input: LAr calorimeter cells (3 layers)
- Classification task with prompt photon as signal and non-prompt photons as background



Photon ID in ATLAS tomorrow?





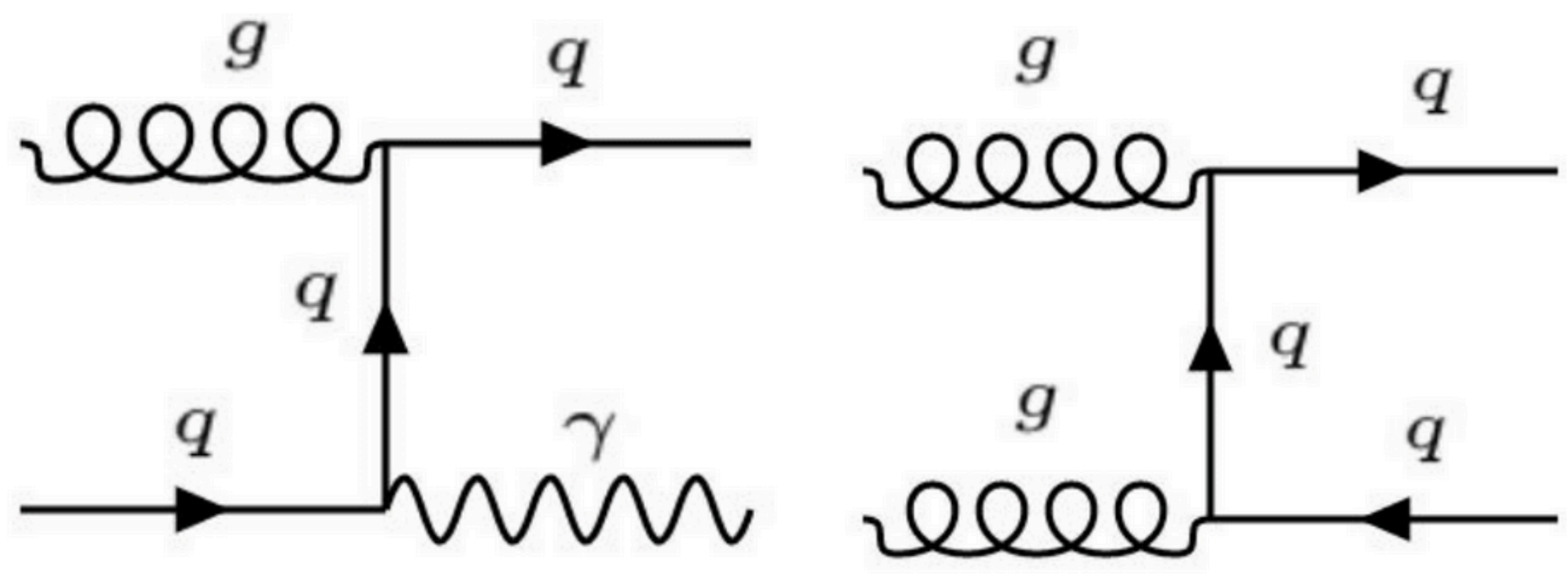


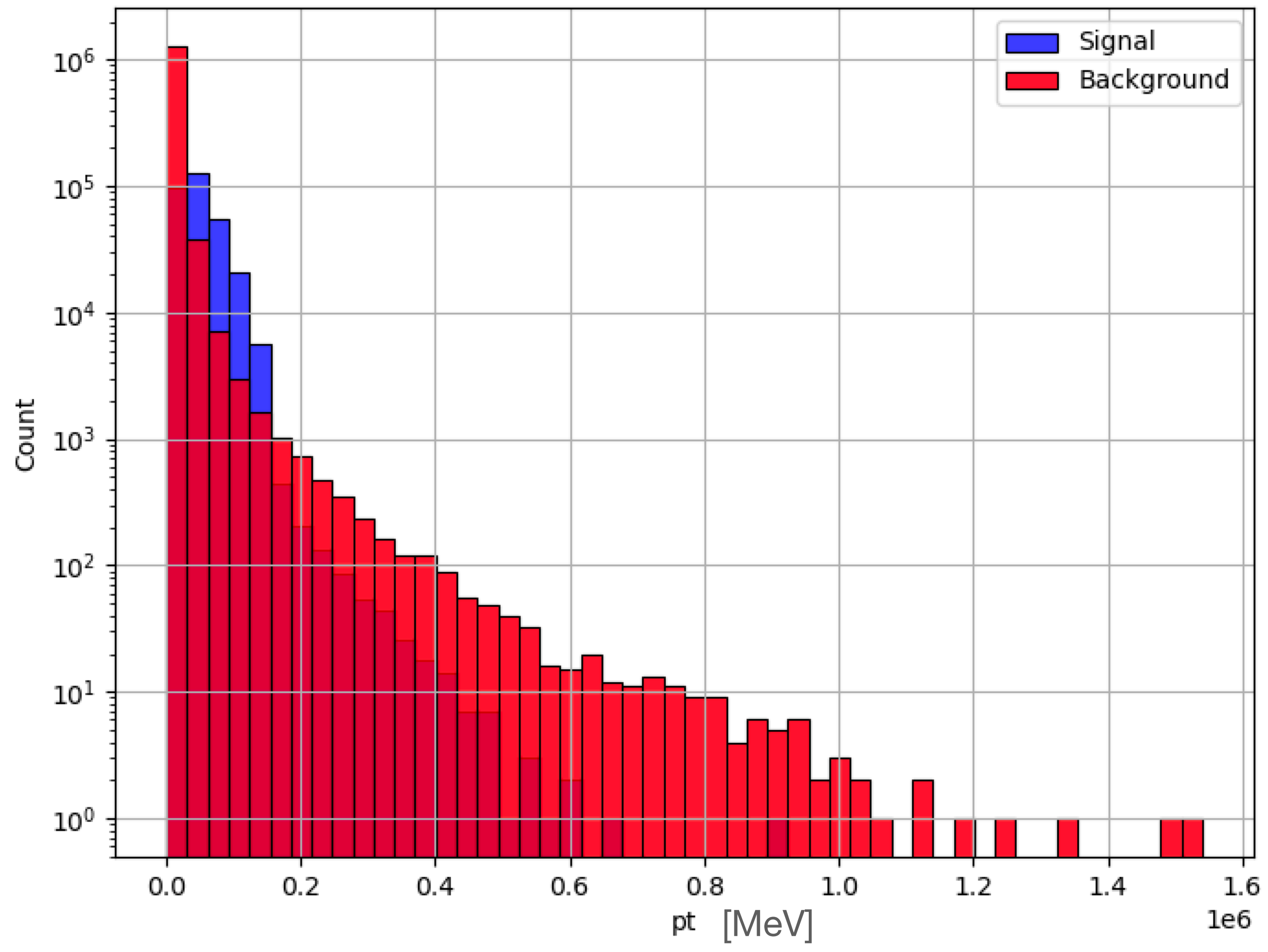
Photon ID becomes an **image recognition** problem with **machine learning**

A revolution

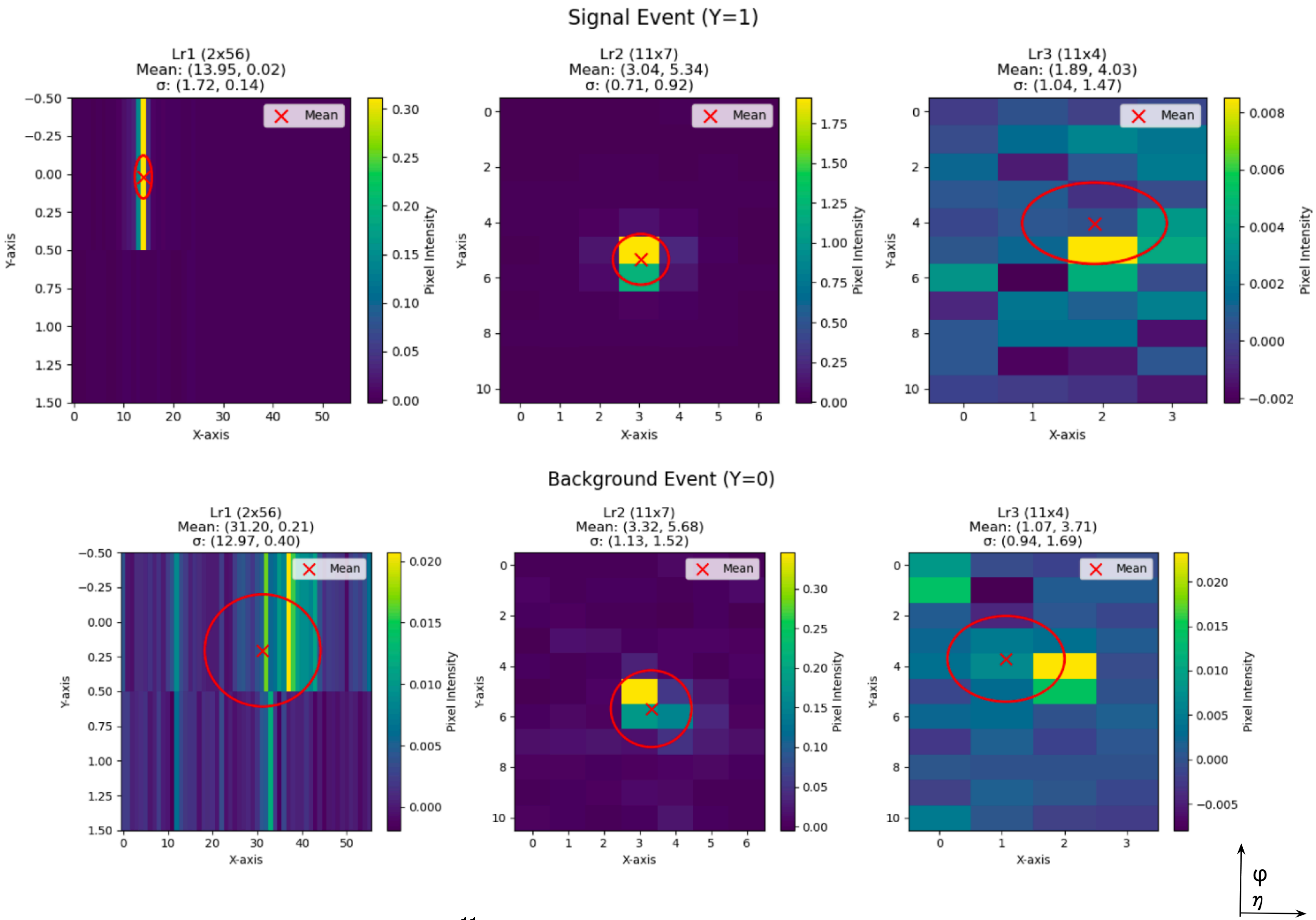
Design of a CNN for Photon ID

- Training with Monte Carlo (MC) simulations of γj and jj processes
- Truth MC label (prompt/non-prompt) used as class label
- Photon pT between ~1 GeV and ~1 TeV
- pT-reweighting in the training to compensate for class imbalance
- 1.5M entries: 80% Training (of which 33% validation), 20% Test



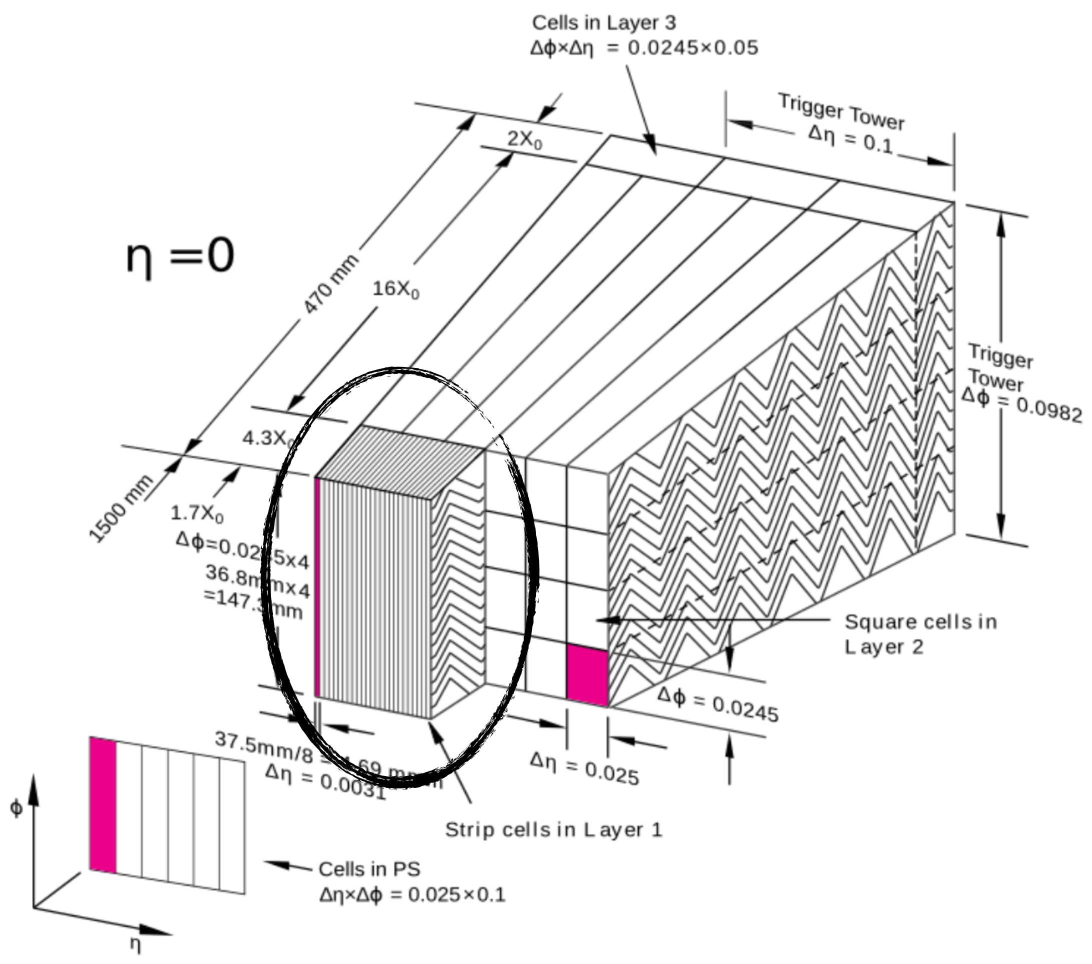


Input features

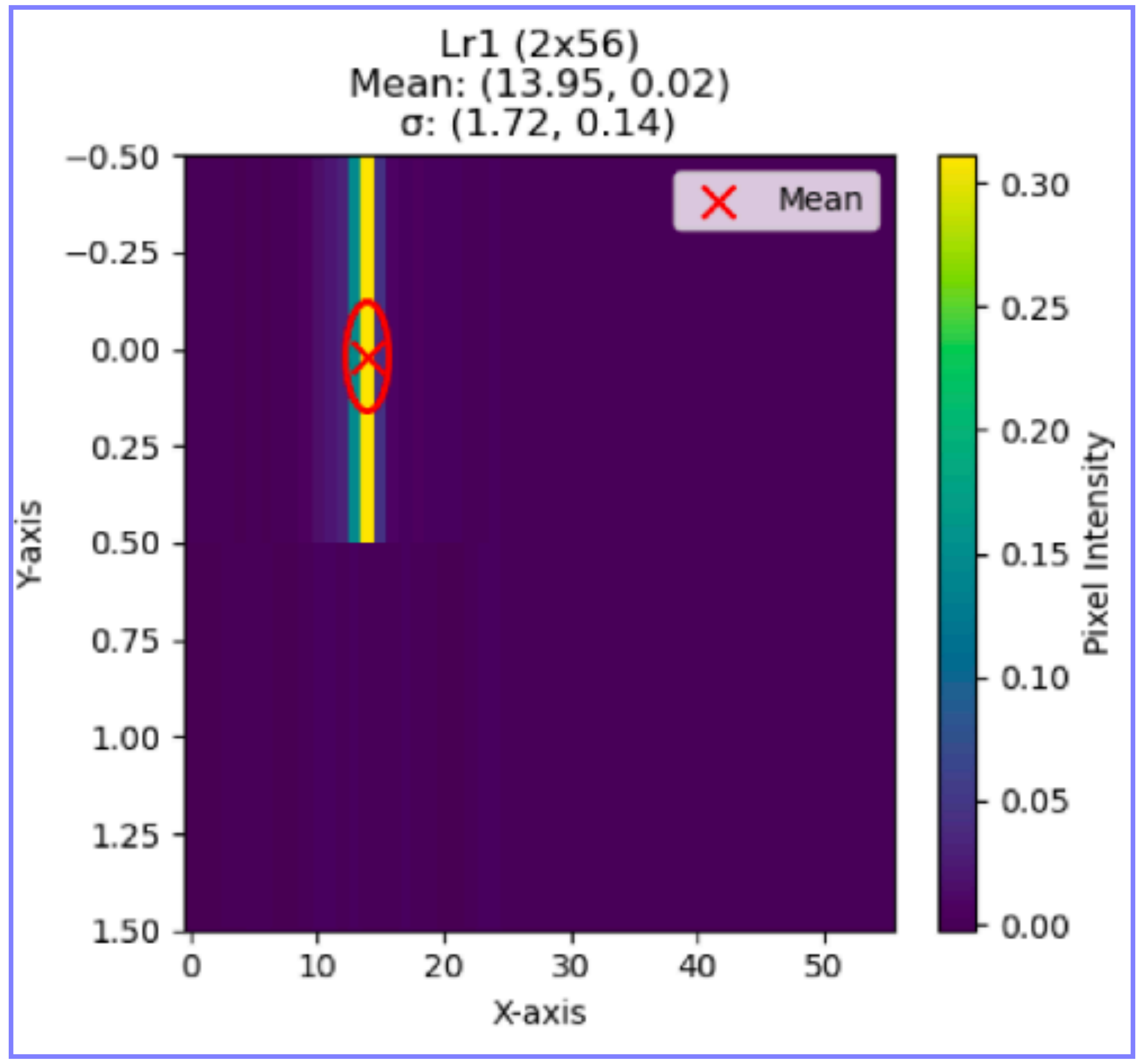


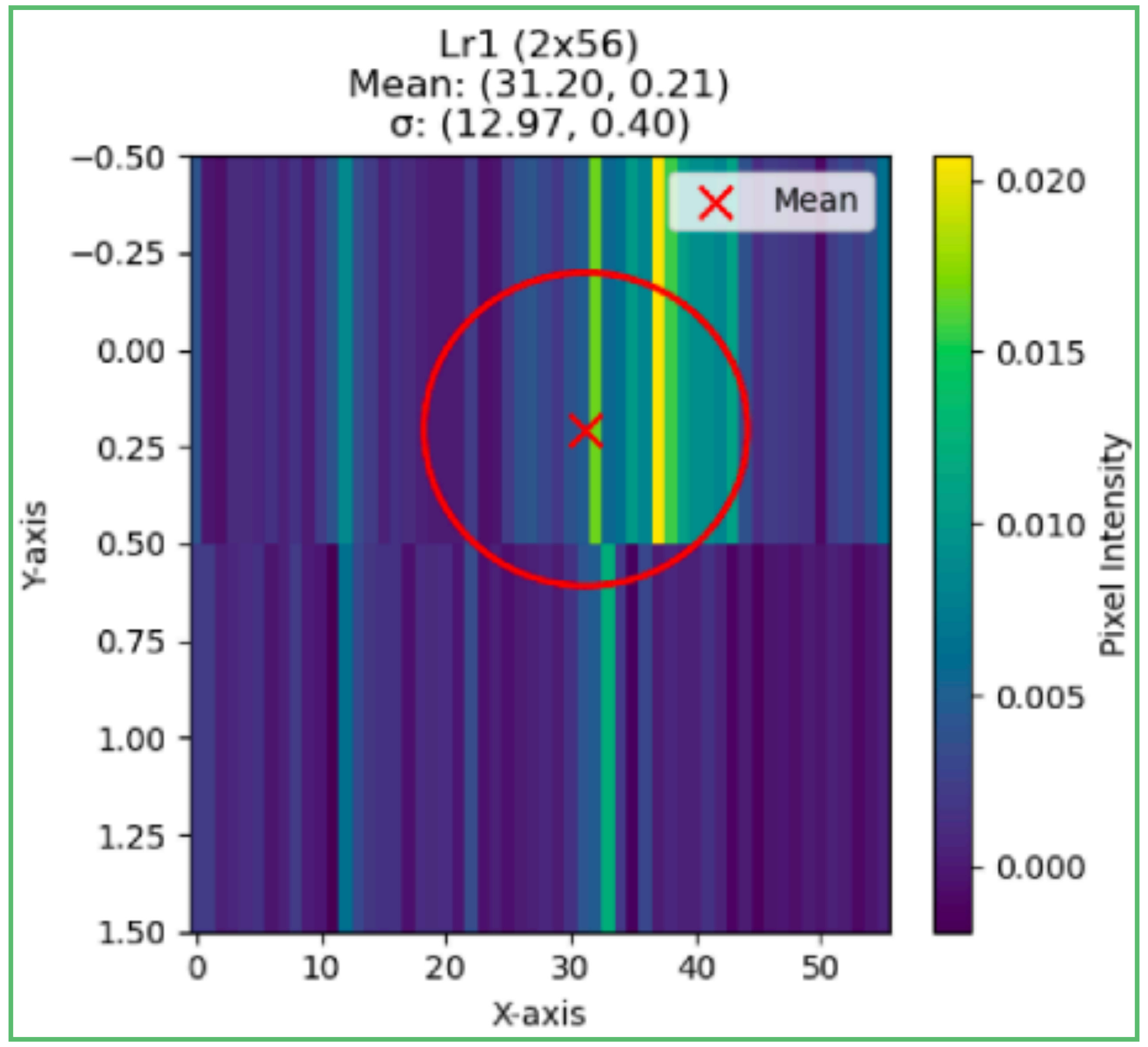
L. Franco - 02/05/2025

Input features

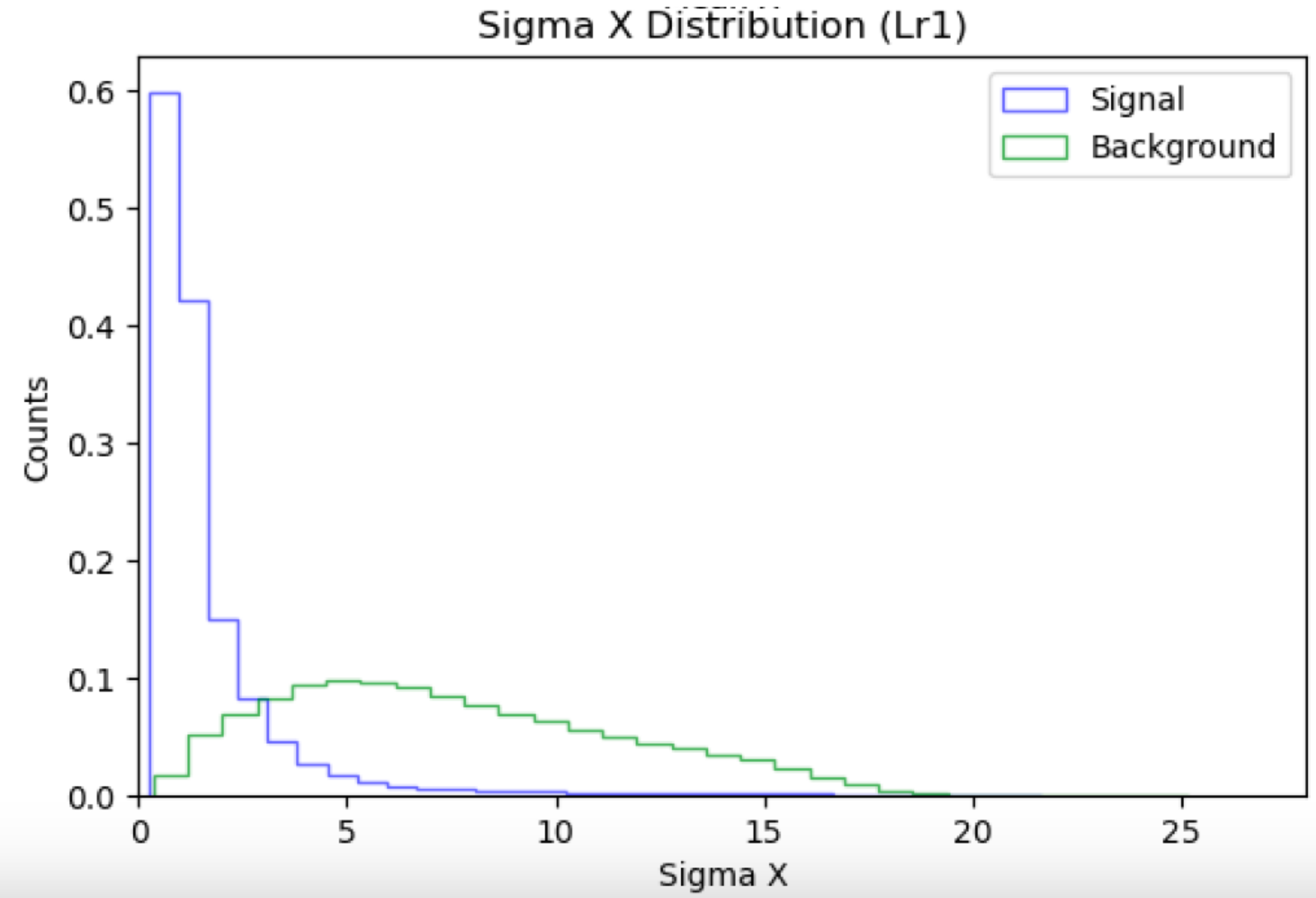


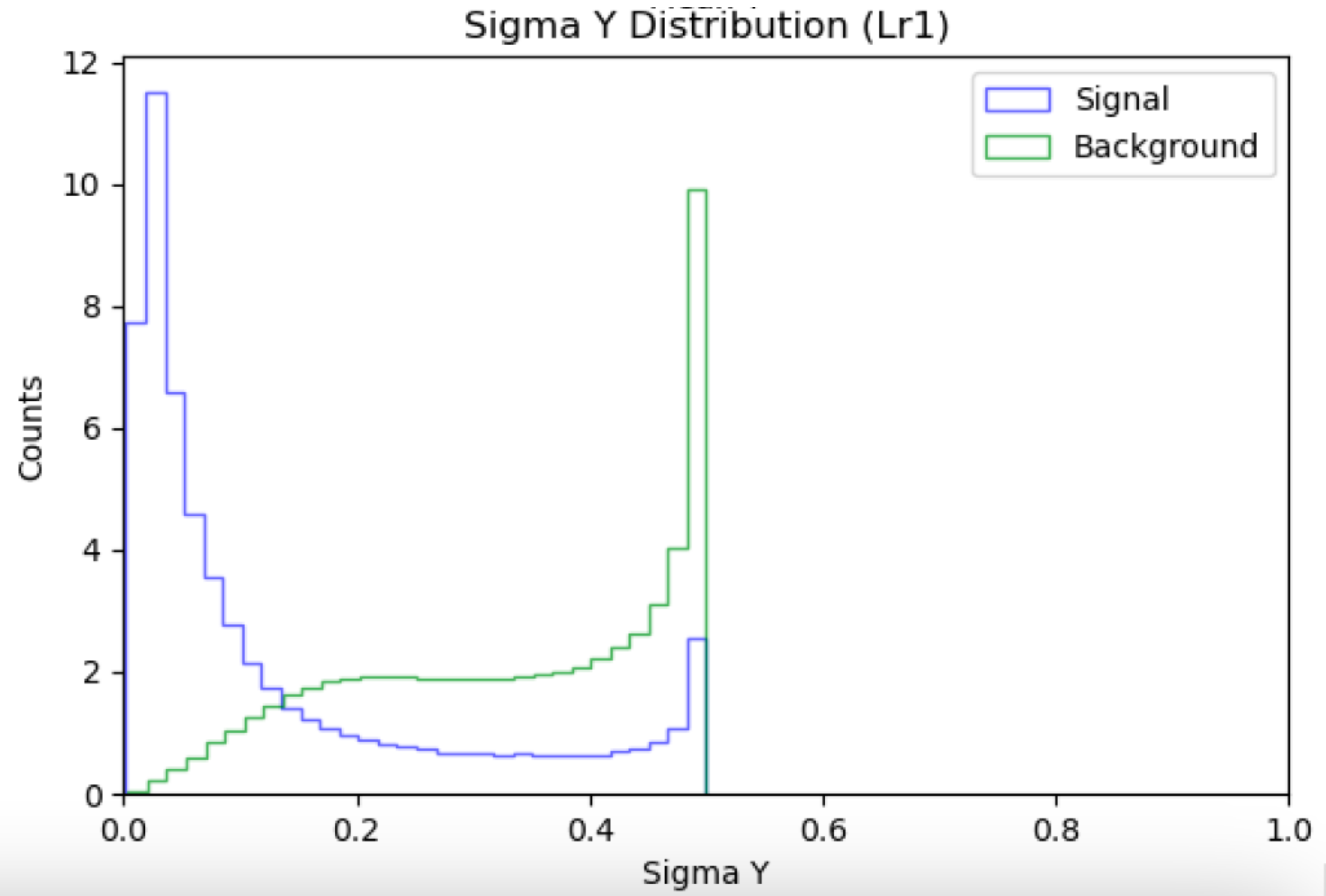
For one entry →





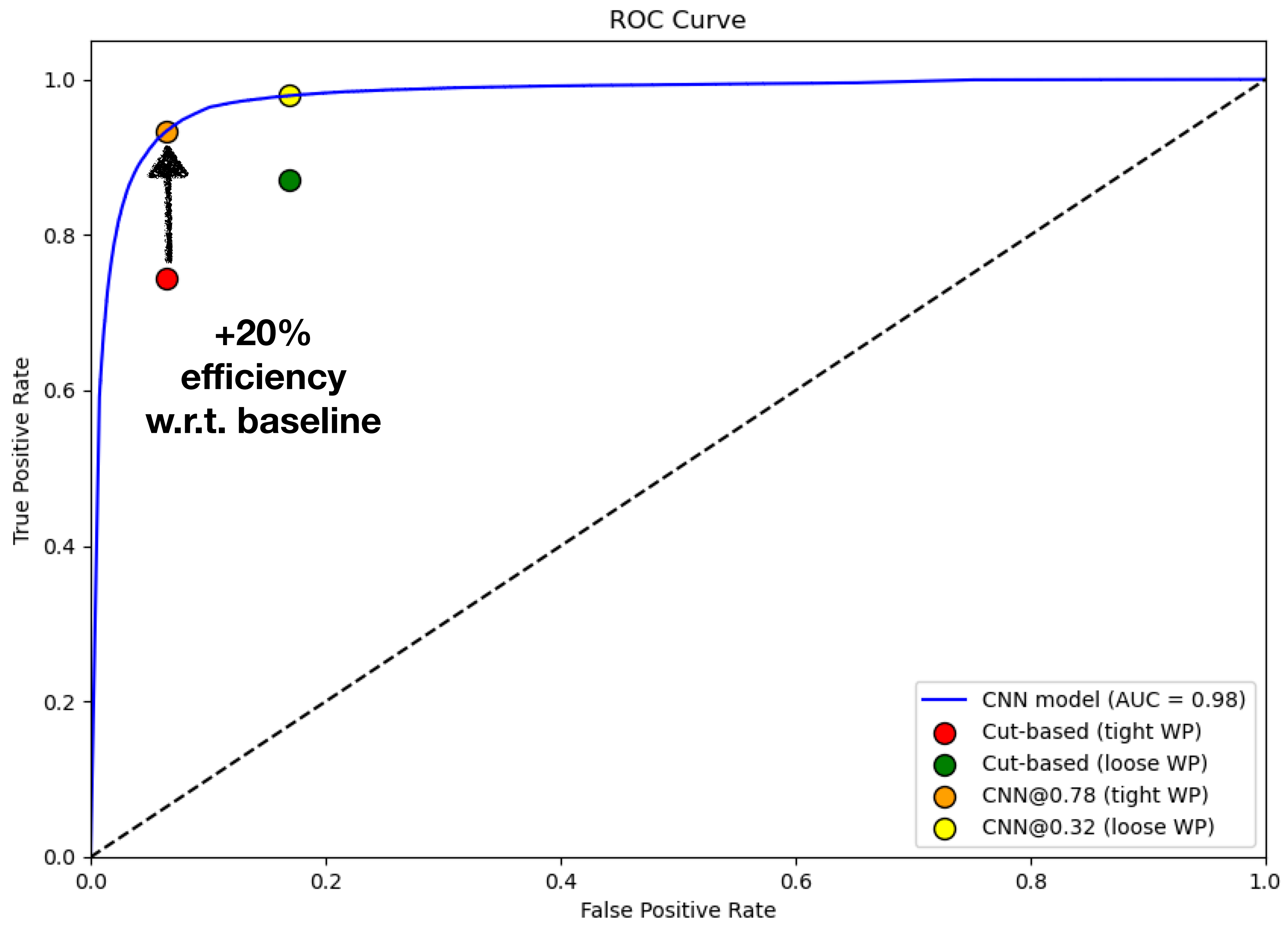
For all entries →



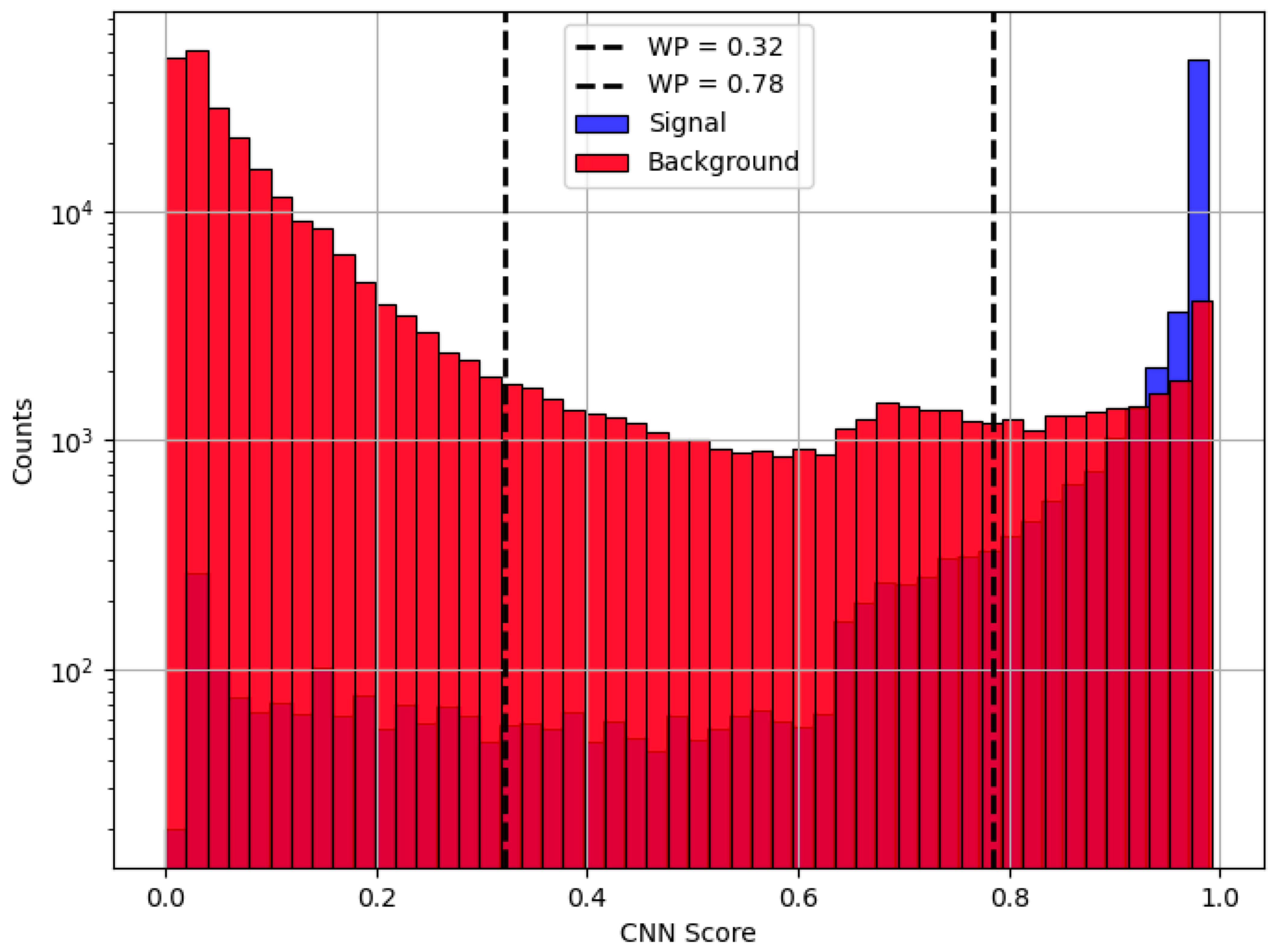


L. Franco - 02/05/2025

CNN performance



DISCLAIMER: for a better comparison, WPs will have to be defined in regions of pT, η, conversion status



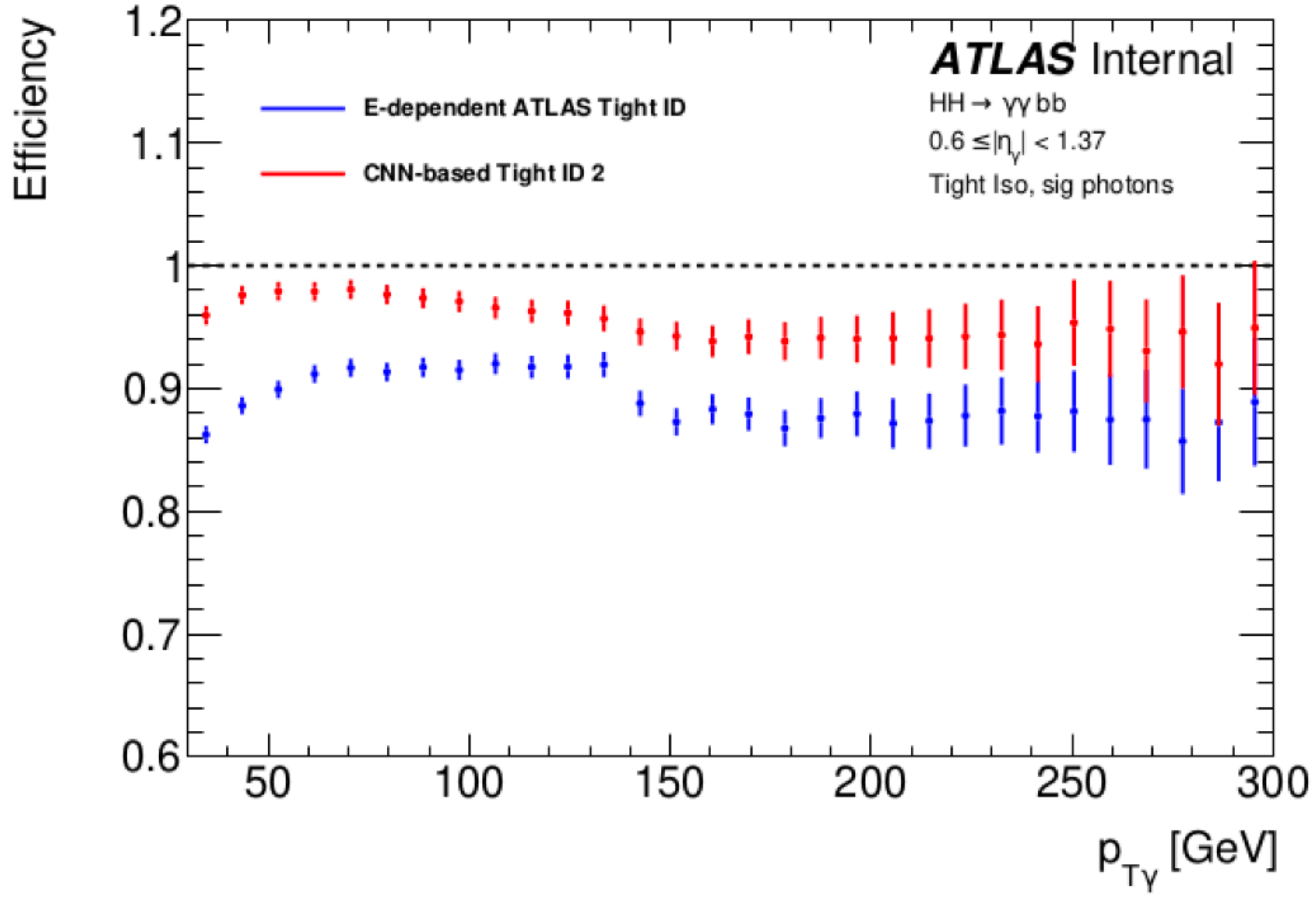
- Two working points (WP) defined on the CNN score:
  - Loose for higher signal efficiency
  - Tight for higher background rejection
- Both match the bkg rej. of cut-based method

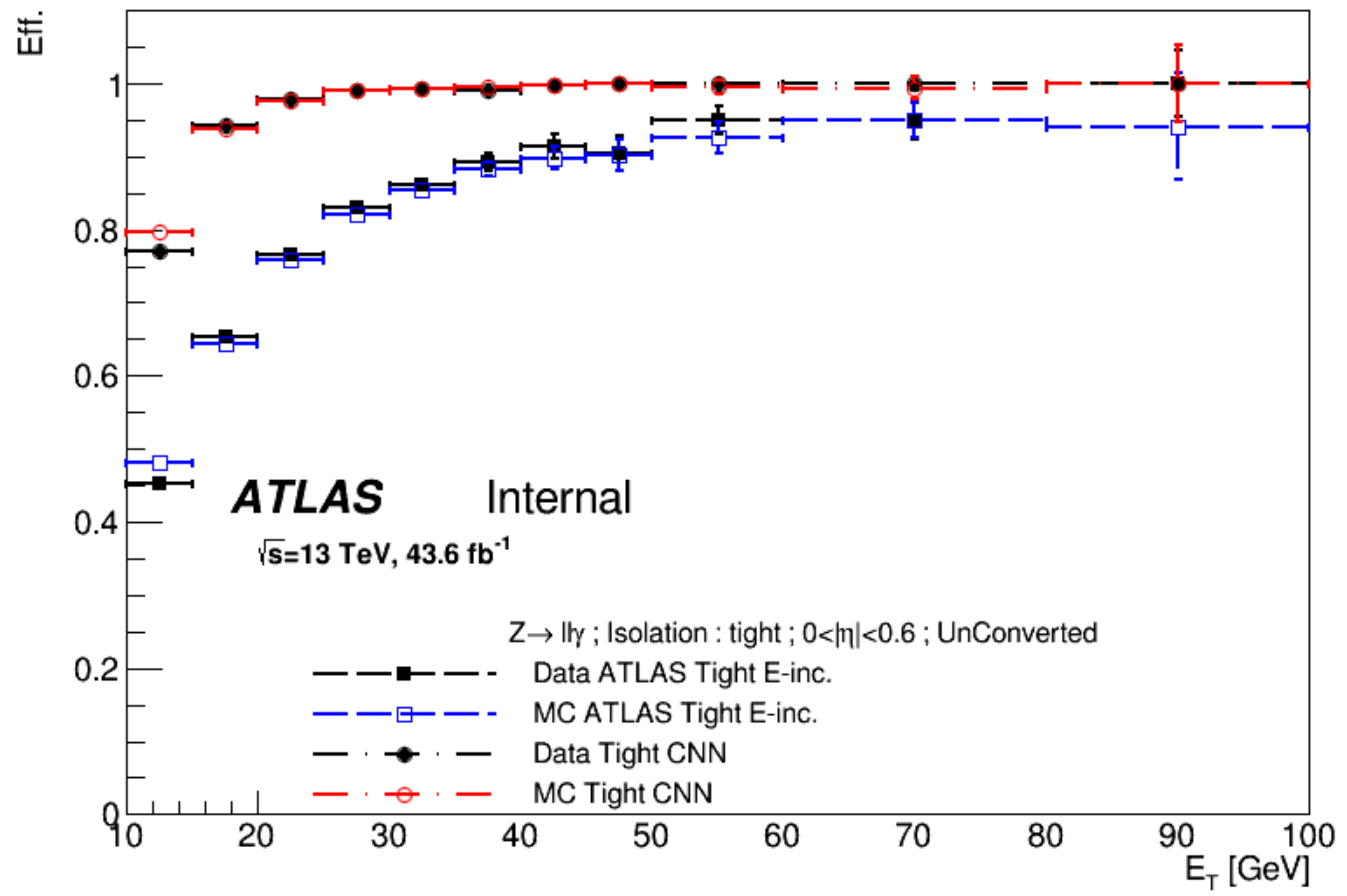
L. Franco - 02/05/2025

Application to physics analyses

This exercise was done with a Rel.21 prototype in 2024

- CNN in HH→bbγγ
  - up to 8% more HH signal in selection
  - final results (HH limit) improved by 5% (GN2 brought ~8%)
- CNN in Z→llγ (Z radiative decays used for calibration)
  - x2 improvement in efficiency at low pT w.r.t. baseline
  - Out-of-sample validation: CNN trained on inclusive photon production and applied to Z→llγ

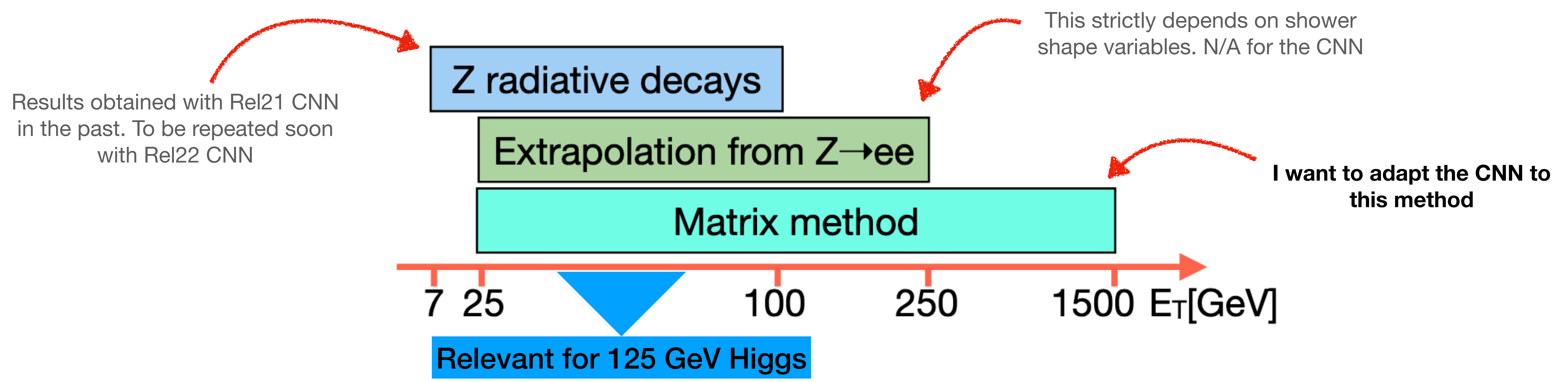




L. Franco - 02/05/2025

The real challenge: data/MC calibration

- The CNN shows promising results when applied to MC
- What about ID efficiencies in data?
- In EGamma there are 3 methods for the measurement of the ID efficiencies in data: all of them were designed for a non-ML-based ID

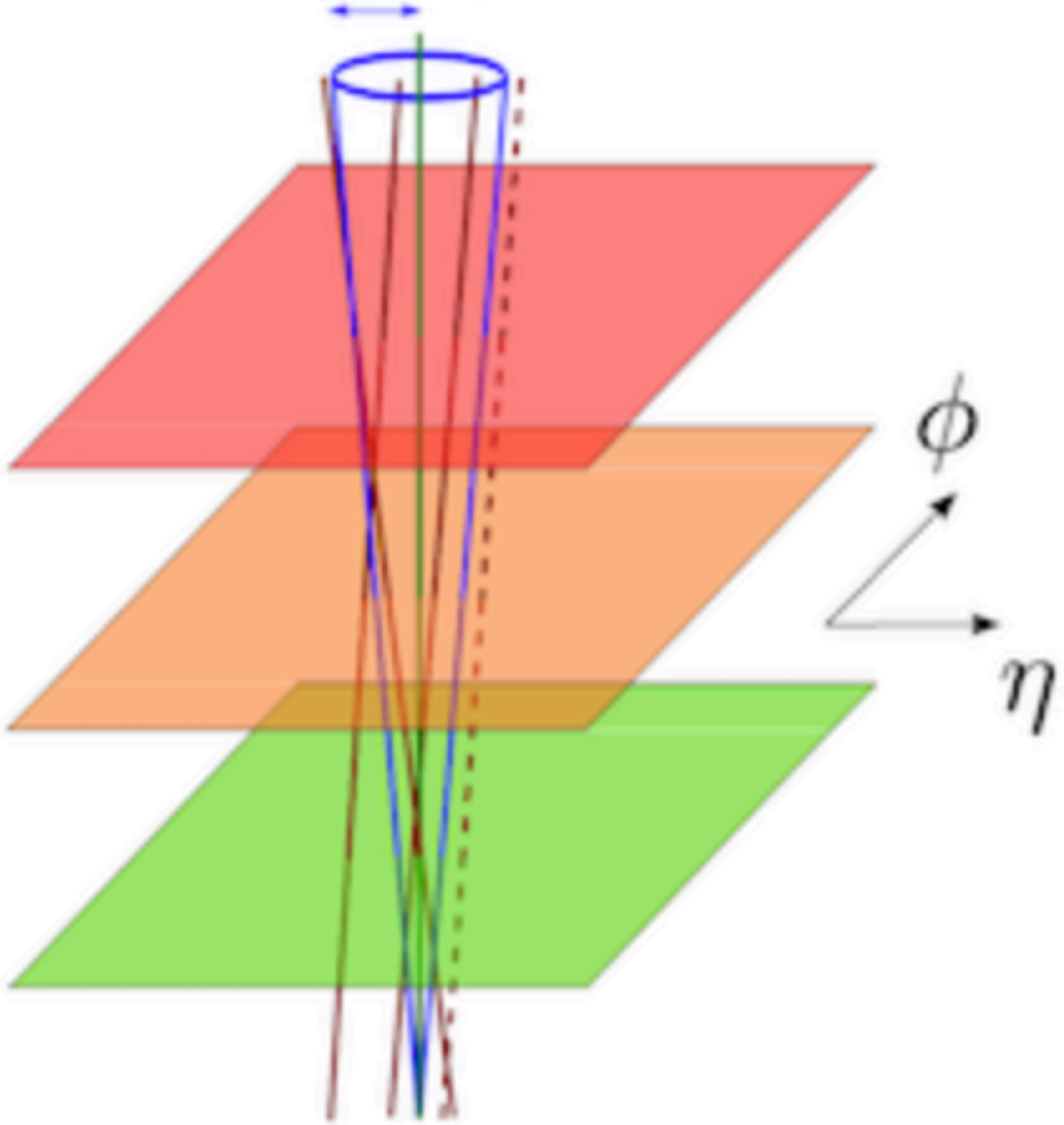


L. Franco - 02/05/2025

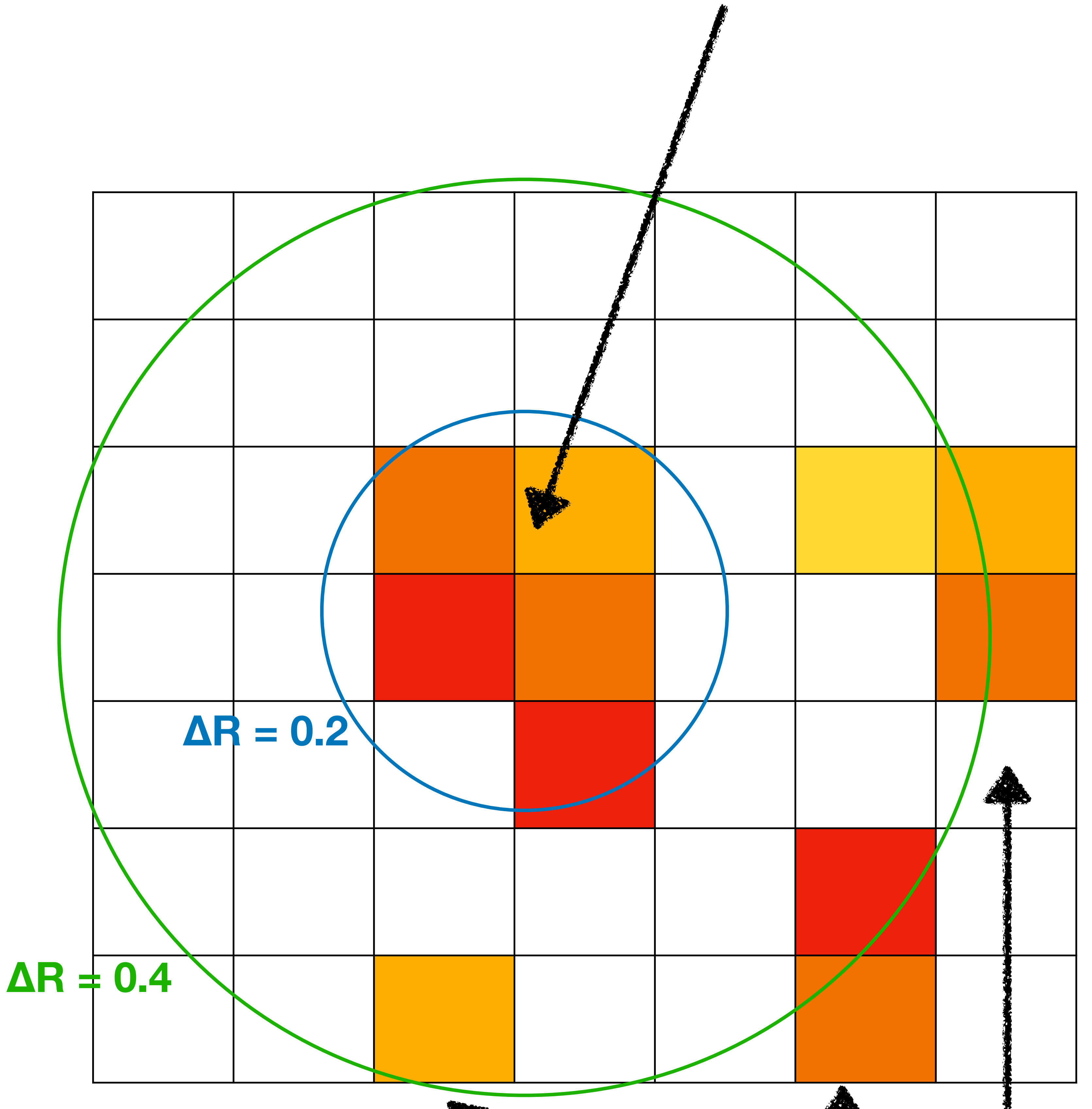
Existing calibration methods rely on the assumption that ID and isolation are *weakly* correlated

Isolation variables

Track-based isolation
scalar sum of pT of all the tracks found inside
the isolation cone of size XX
Ptcone20
Ptcone40



Energy cluster associated to photon candidate



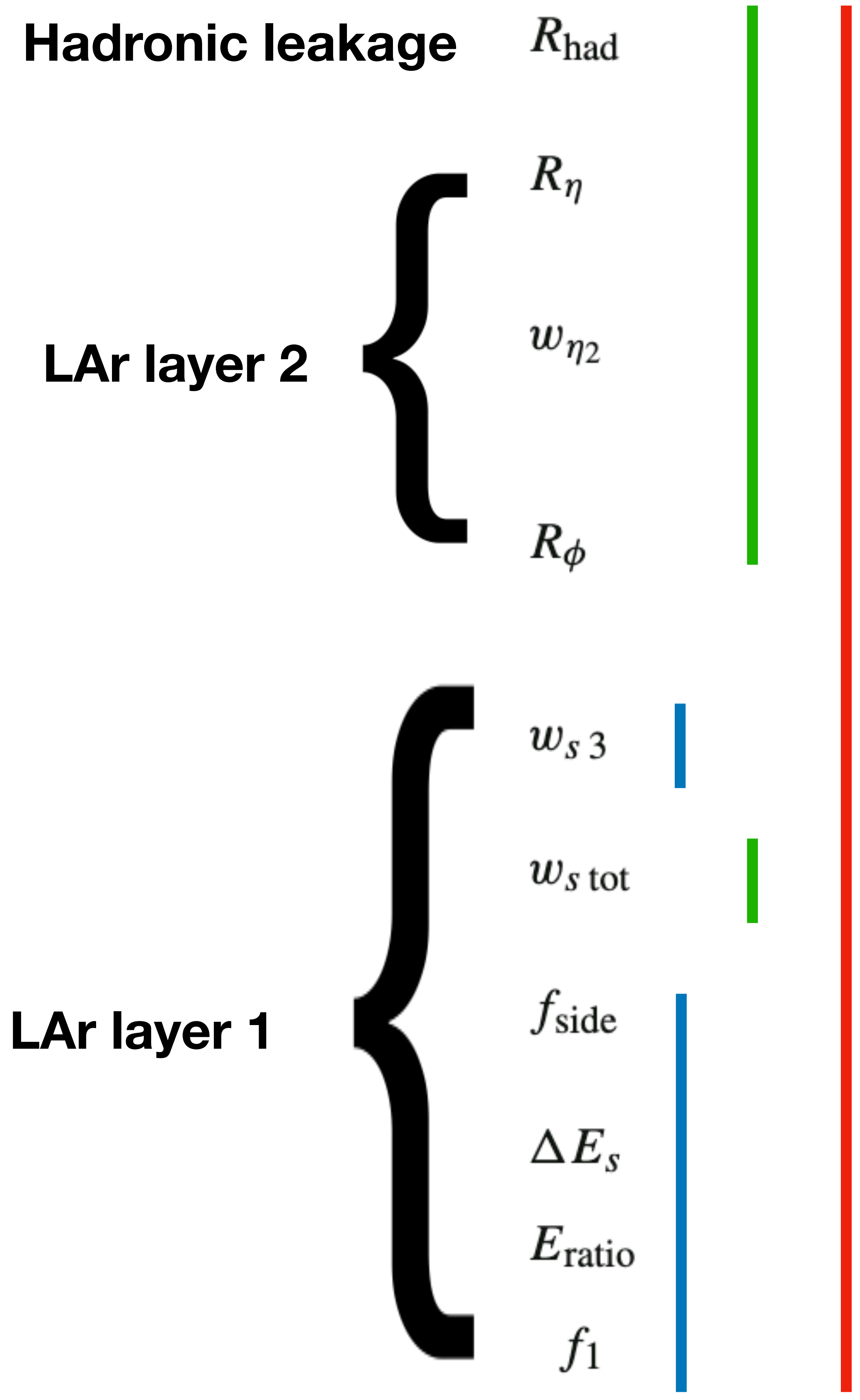
Calo-based isolation
Sum of ET of all clusters in a cone of size
XX, after removig the photon candidate
Etcone20
Etcone40

Other energy clusters

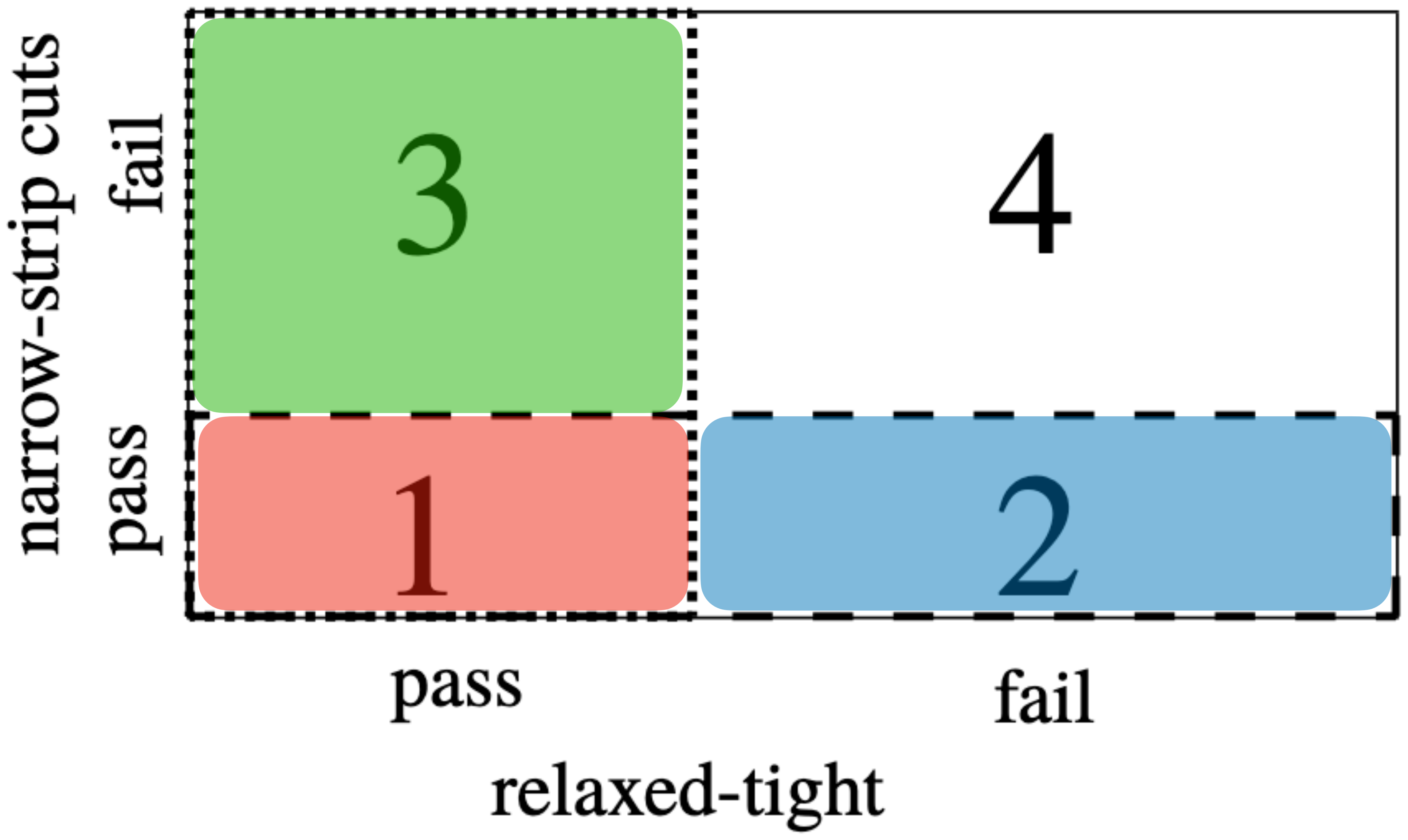
Recommended (loose) WP: used in the bbyy analysis and others…

$$\mathrm{E}_{\mathrm{T}}^{\mathrm{iso}}\Big|_{\Delta R<0.2} < 0.065 \cdot E_{\mathrm{T}} \quad \text{and} \quad \mathrm{p}_{\mathrm{T}}^{\mathrm{iso}}\Big|_{\Delta R<0.2} < 0.05 \cdot E_{\mathrm{T}}$$

Efficiency measurement with Matrix Method



- **Shower shapes** are split into two subsets: **narrow-strip** and **relaxed-tight**
- These are used to define 4 regions
- Track-Isolation used to estimate efficiency in each region

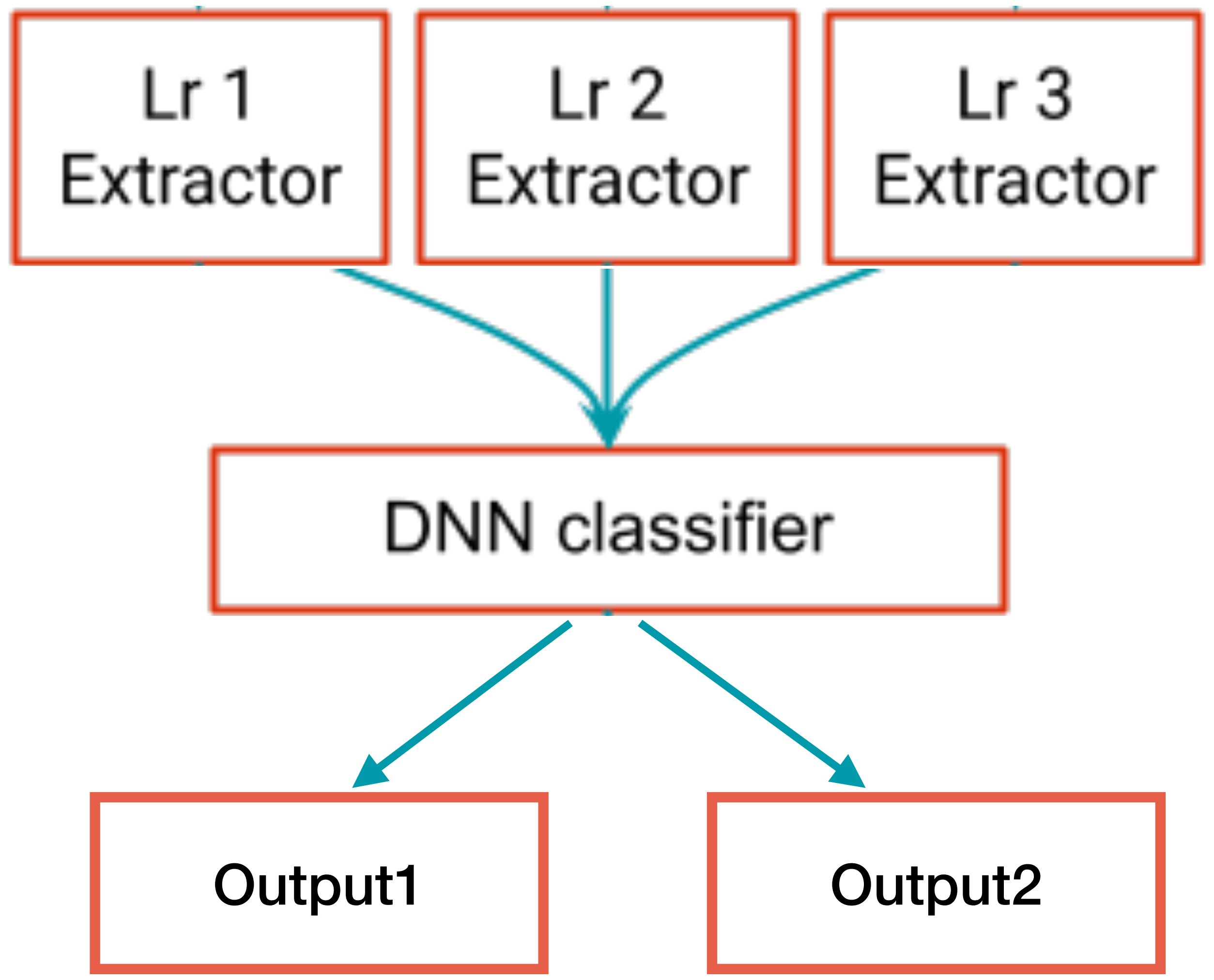


	cut-based	feature
1 must pass	all shower shapes	best performance
2 must pass	narrow-strip variables	weak correlation with Iso
3 must pass	relaxed-tight variables	orthogonal (?) to narrow-strip

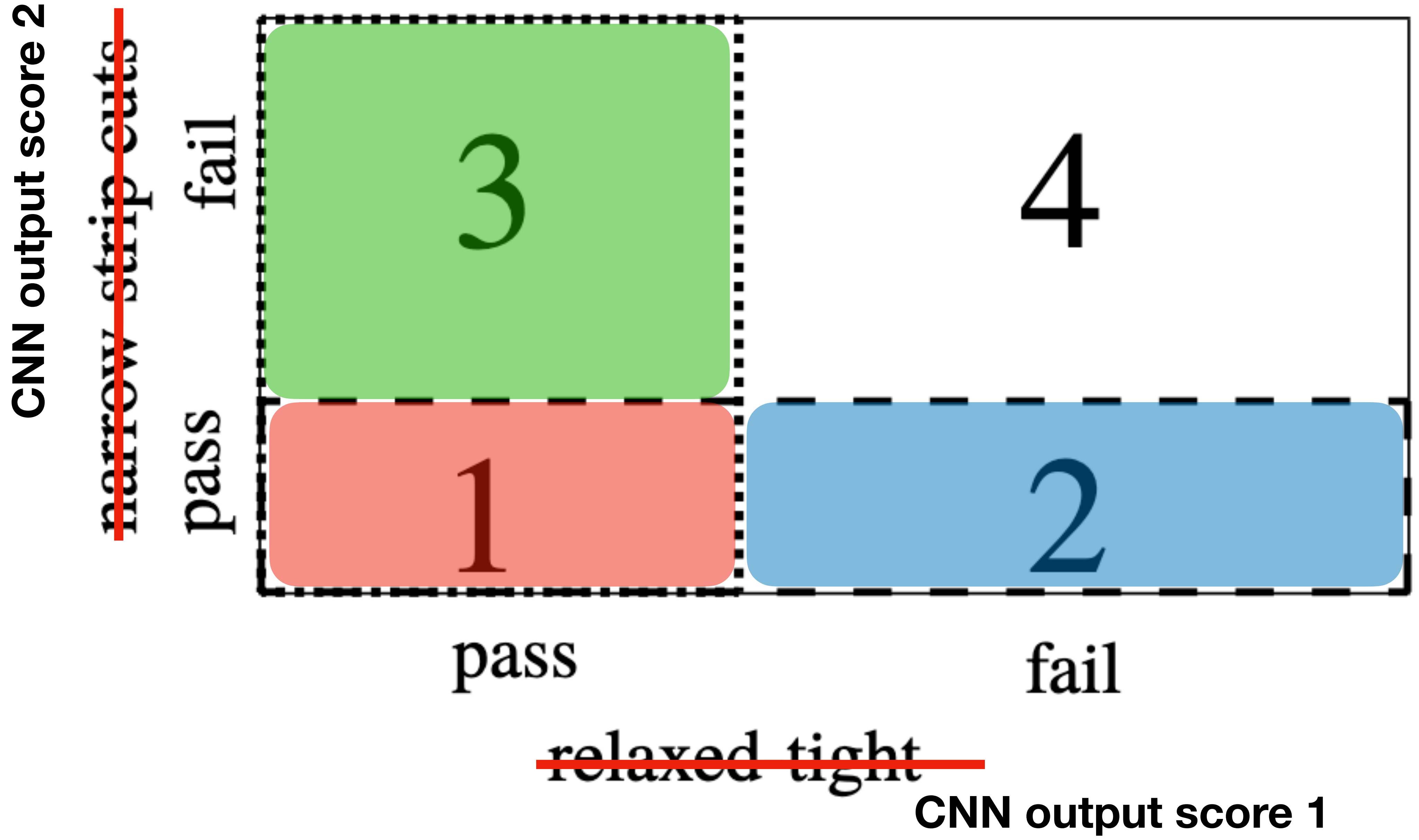
L. Franco - 02/05/2025

Matrix Method with the CNN

- Idea: design a multi-output CNN to “mimic” narrow-strip and relaxed-tight cuts
- Define a separate CNN output which is explicitly decorrelated w.r.t. track-isolation in the loss



$$\mathscr{L} = \sum_{i=1}^{2} \mathrm{BCE}(\mathrm{Out_i}, \mathrm{y}) + \lambda \mathrm{Decorr}(\mathrm{Out_2}, \mathrm{ptcone40})$$

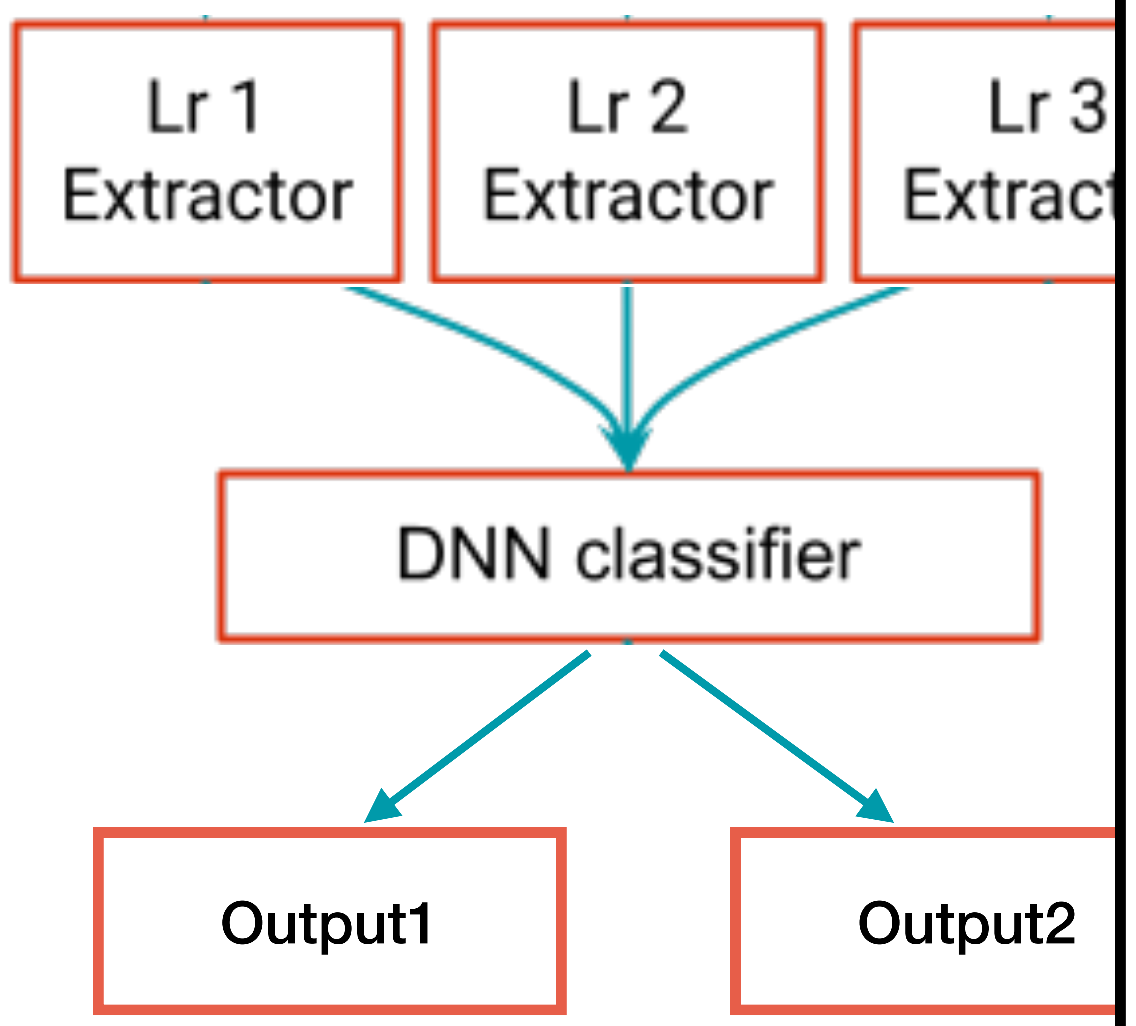


	CNN	feature
1 must pass	OS1 and OS2	best performance
2 must pass	Output score 2	weak correlation with Iso
3 must pass	Output score 1	orthogonal (?) to OS2

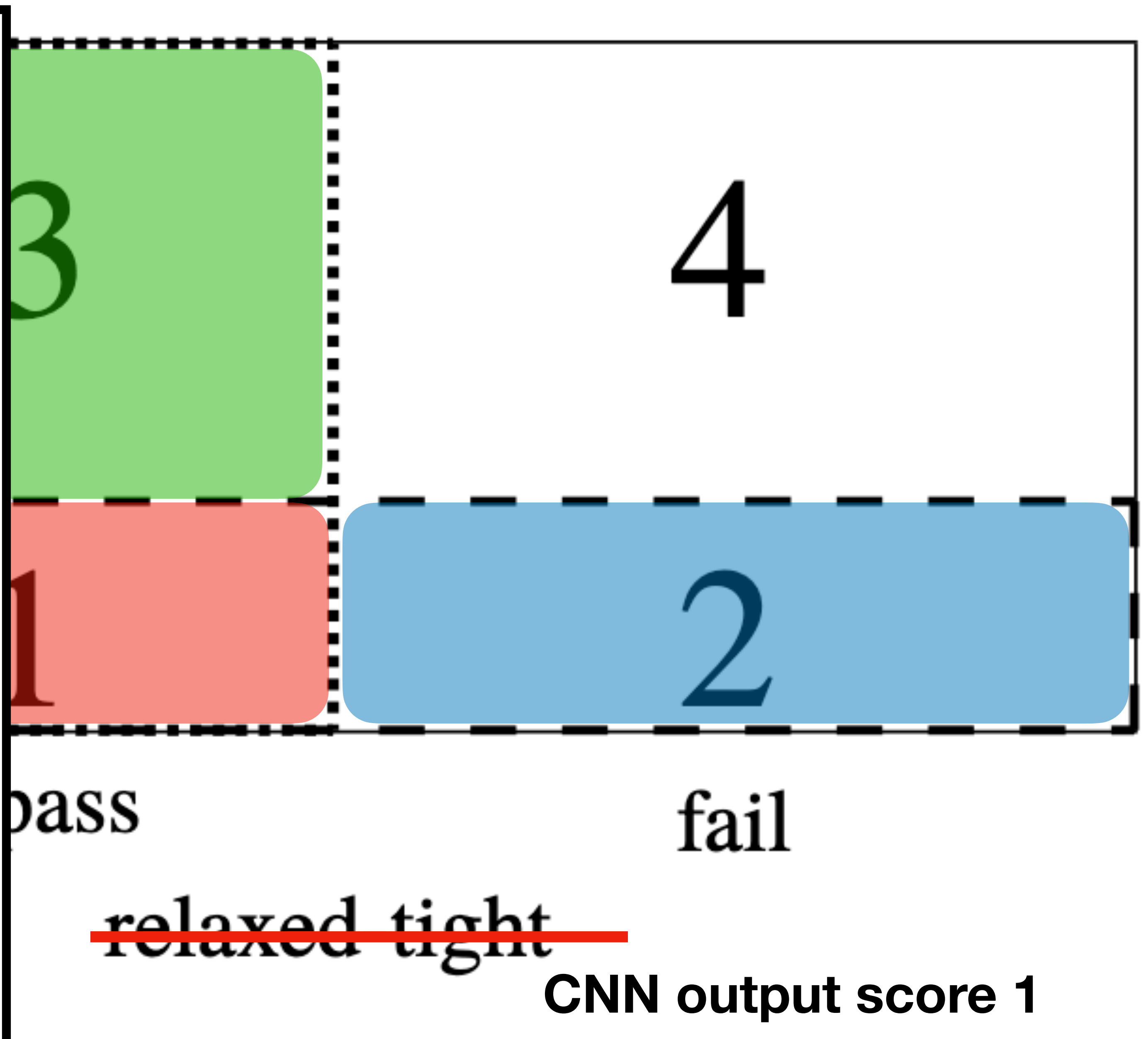
Matrix Method with the CNN

- Idea: design a multi-output C
narrow-strip and relaxed-tigh

- Define a separate CNN outp
decorrelated w.r.t. track-isola



$$\mathscr{L} = \sum_{i=1}^{2} \mathrm{BCE}(\mathrm{Out_i}, \mathrm{y}) + \lambda \mathrm{Decorr}(\mathrm{Out_2}, \mathrm{ptcone40})$$



	CNN	feature
	OS1 and OS2	best performance
	Output score 2	weak correlation with Iso
3 must pass	Output score 1	orthogonal (?) to OS2

L. Franco - 02/05/2025

Conclusions

- My mission for the next ~2 years (hopefully in time for full Run3 EGamma recommendations): **revolution of photon ID in ATLAS**
- Progress is slow (CNN requires dedicated MC sample production with EM cells) but we are getting there!
- Potential gain for a wide range of physics analyses, mainly searches with photons
- Is it a low-hanging fruit? Not really, data/MC calibration is difficult
- Multi-dimensional project: ML developments, design of calibration method, interplay with isolation, application to analyses
- I will present the status at the next ATLAS EGamma workshop in a few weeks

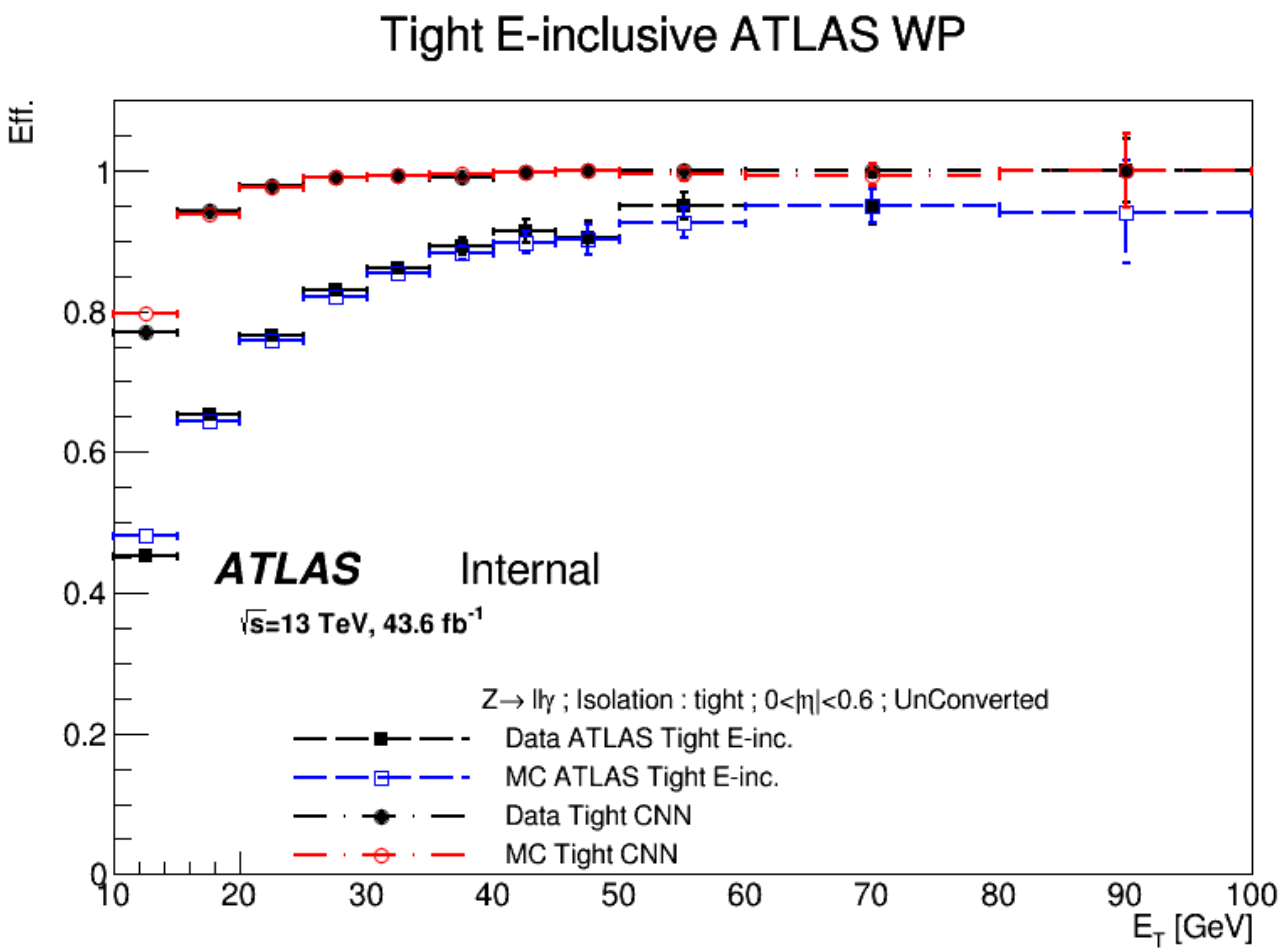
Back-up

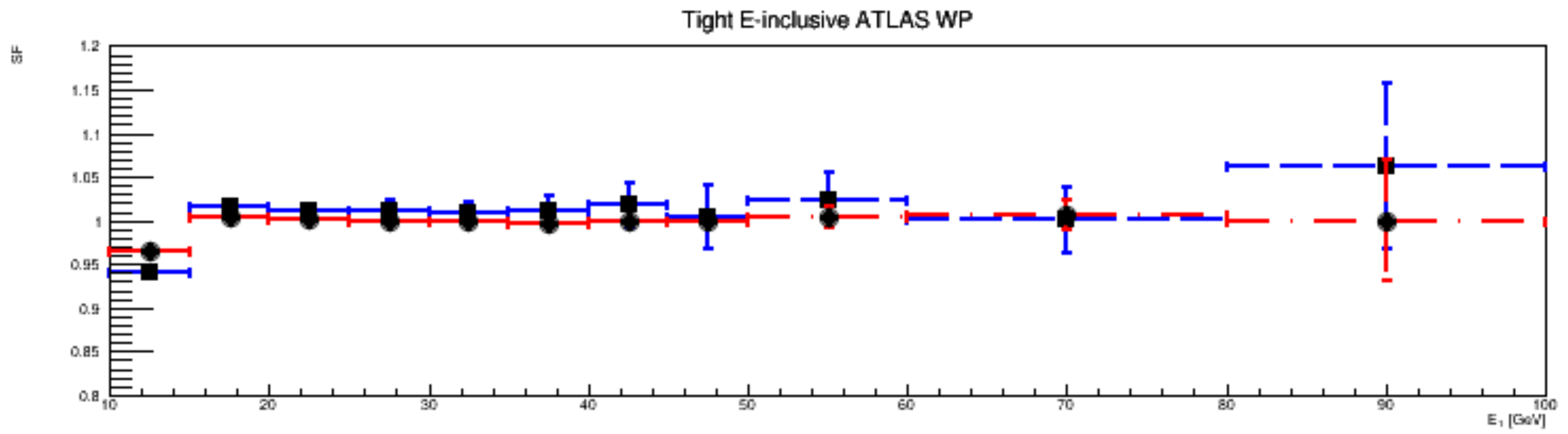
Table 1 Discriminating variables used for *loose* and *tight* photon identification

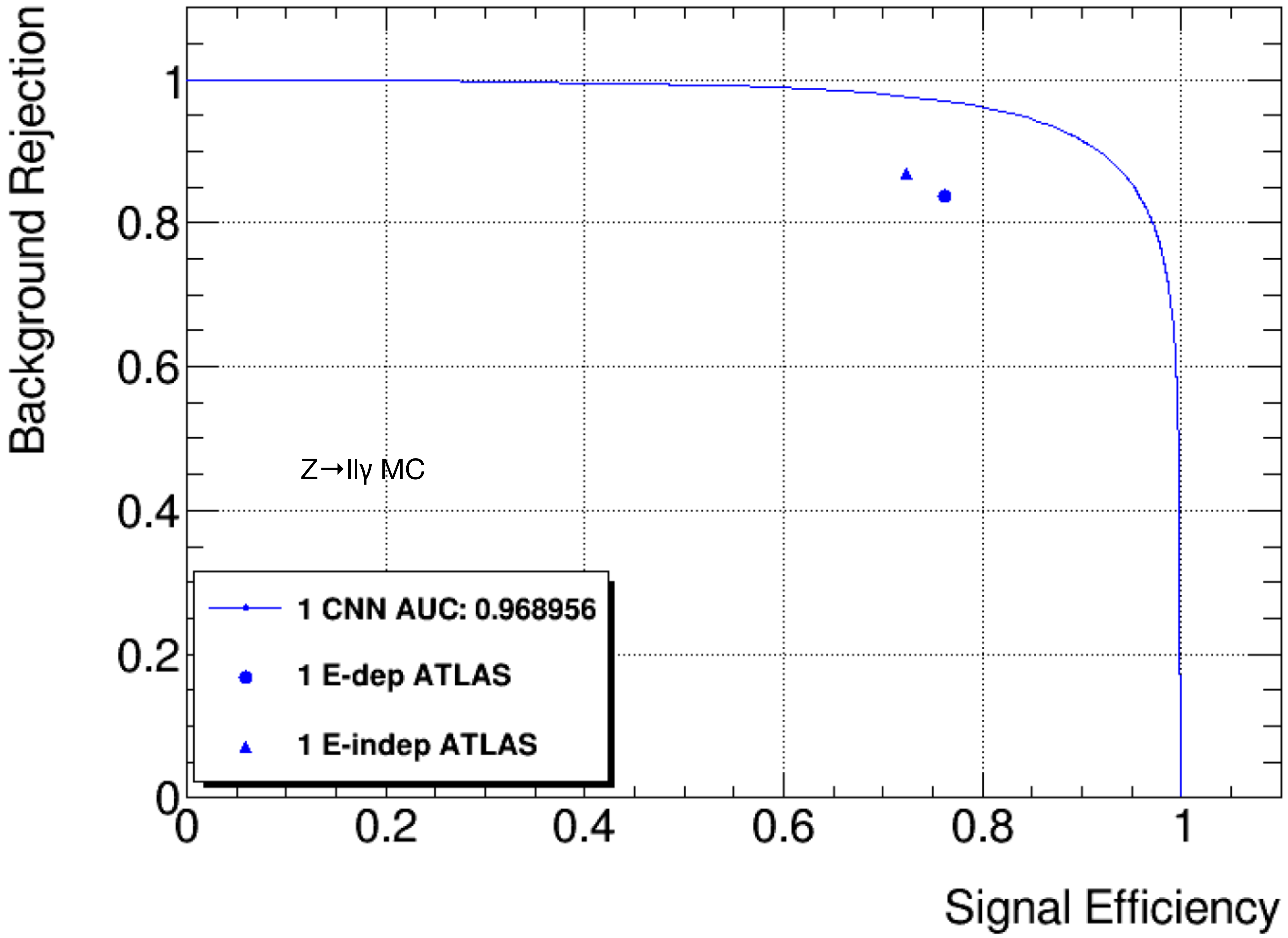
Category	Description	Name	*loose*	*tight*
Acceptance	$\lvert\eta\rvert < 2.37$, with $1.37 \leq \lvert\eta\rvert < 1.52$ excluded	–	✓	✓
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\lvert\eta\rvert < 0.8$ or $\lvert\eta\rvert > 1.52$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \lvert\eta\rvert < 1.37$)	R_had	✓	✓
EM middle layer	Ratio of the energy in 3×7 $\eta \times \phi$ cells over the energy in 7×7 cells centered around the photon cluster position	R_η	✓	✓
	Lateral shower width, $\sqrt{(\Sigma E_i \eta_i^2)/(\Sigma E_i) - ((\Sigma E_i \eta_i)/(\Sigma E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$w_{\eta 2}$	✓	✓
	Ratio of the energy in 3×3 $\eta \times \phi$ cells over the energy of 3×7 cells centered around the photon cluster position	R_ϕ		✓
EM strip layer	Lateral shower width, $\sqrt{(\Sigma E_i (i - i_\mathrm{max})^2)/(\Sigma E_i)}$, where i runs over all strips in a window of 3×2 $\eta \times \phi$ strips, and i_max is the index of the highest-energy strip calculated from three strips around the strip with maximum energy deposit	$w_{s\,3}$		✓
	Total lateral shower width $\sqrt{(\Sigma E_i (i - i_\mathrm{max})^2)/(\Sigma E_i)}$, where i runs over all strips in a window of 20×2 $\eta \times \phi$ strips, and i_max is the index of the highest-energy strip measured in the strip layer	$w_{s\,\mathrm{tot}}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	f_side		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE_s		✓
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in the secondary maximum in the cluster to the sum of these energies	E_ratio		✓
	Ratio of the energy in the first layer to the to the total energy of the EM cluster	f_1		✓

L. Franco - 02/05/2025

CNN with radiative Z

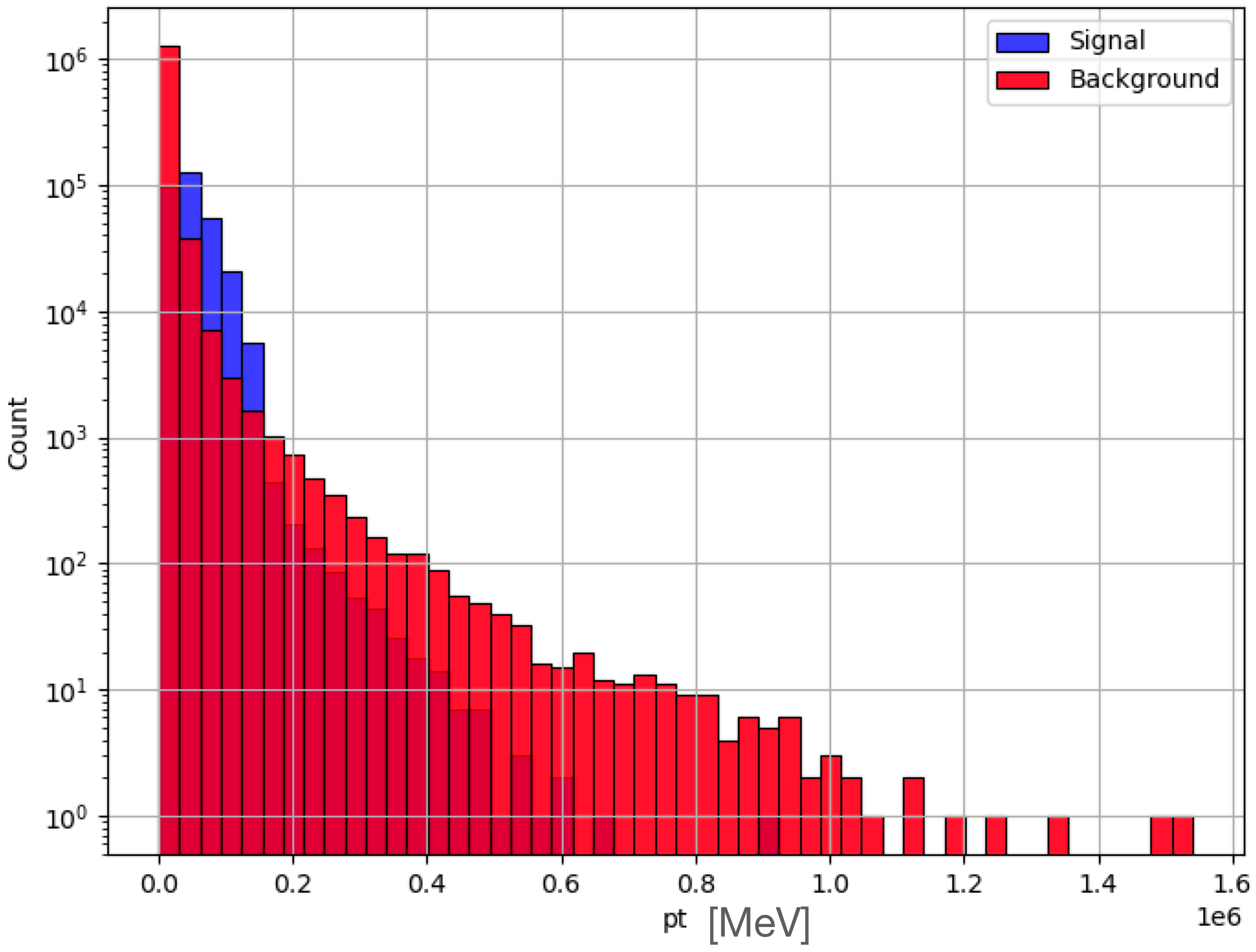


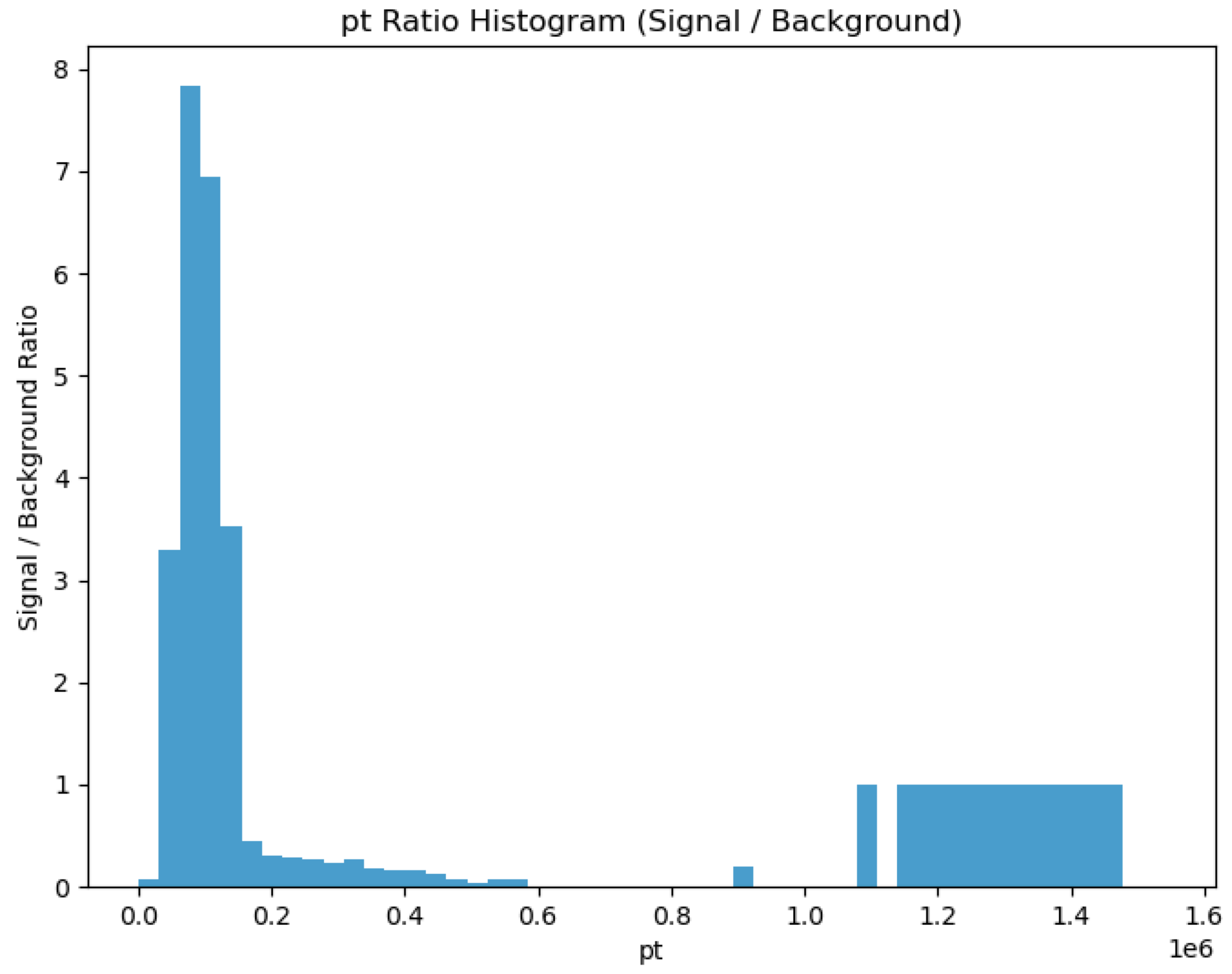


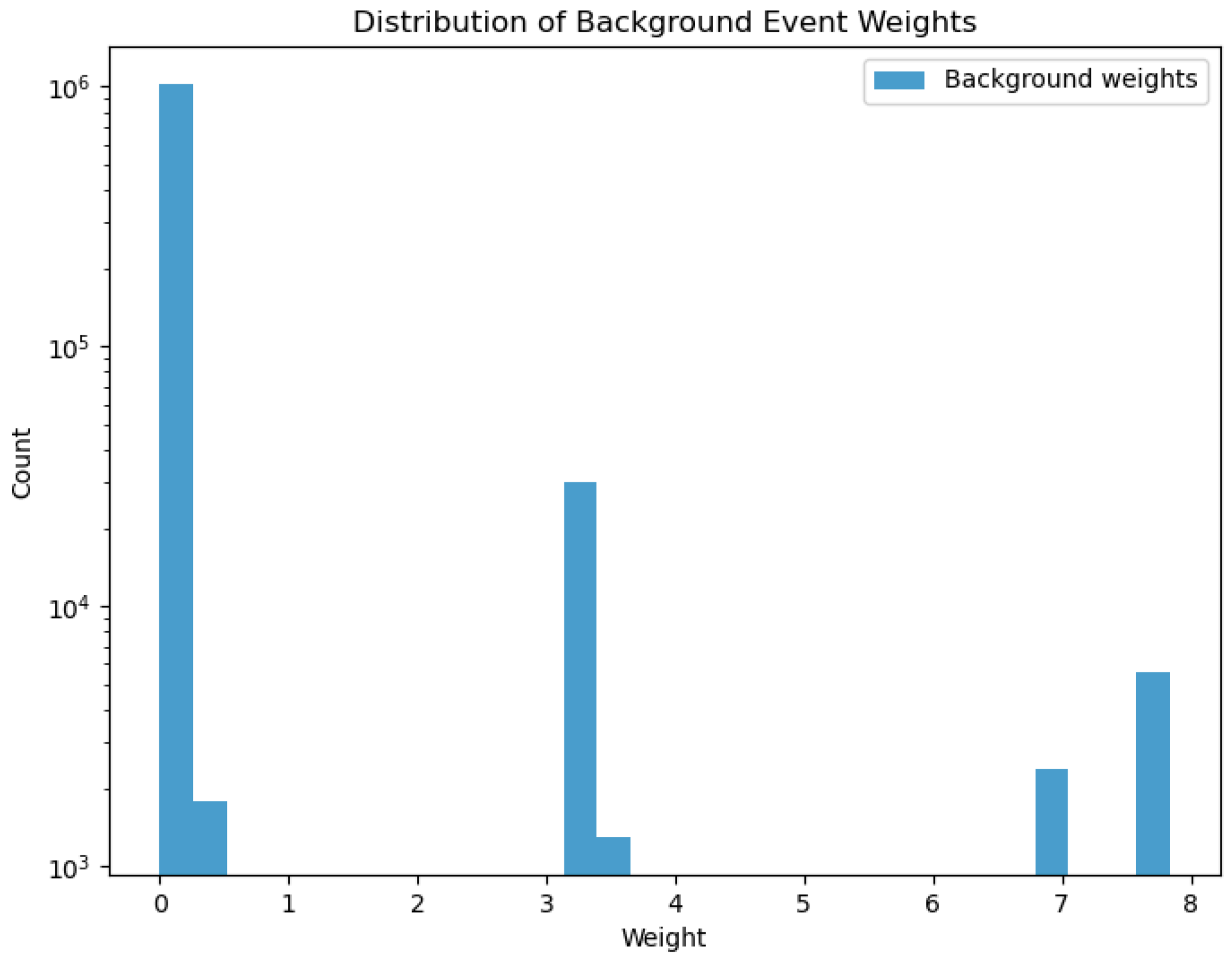


L. Franco - 02/05/2025

pT reweighting



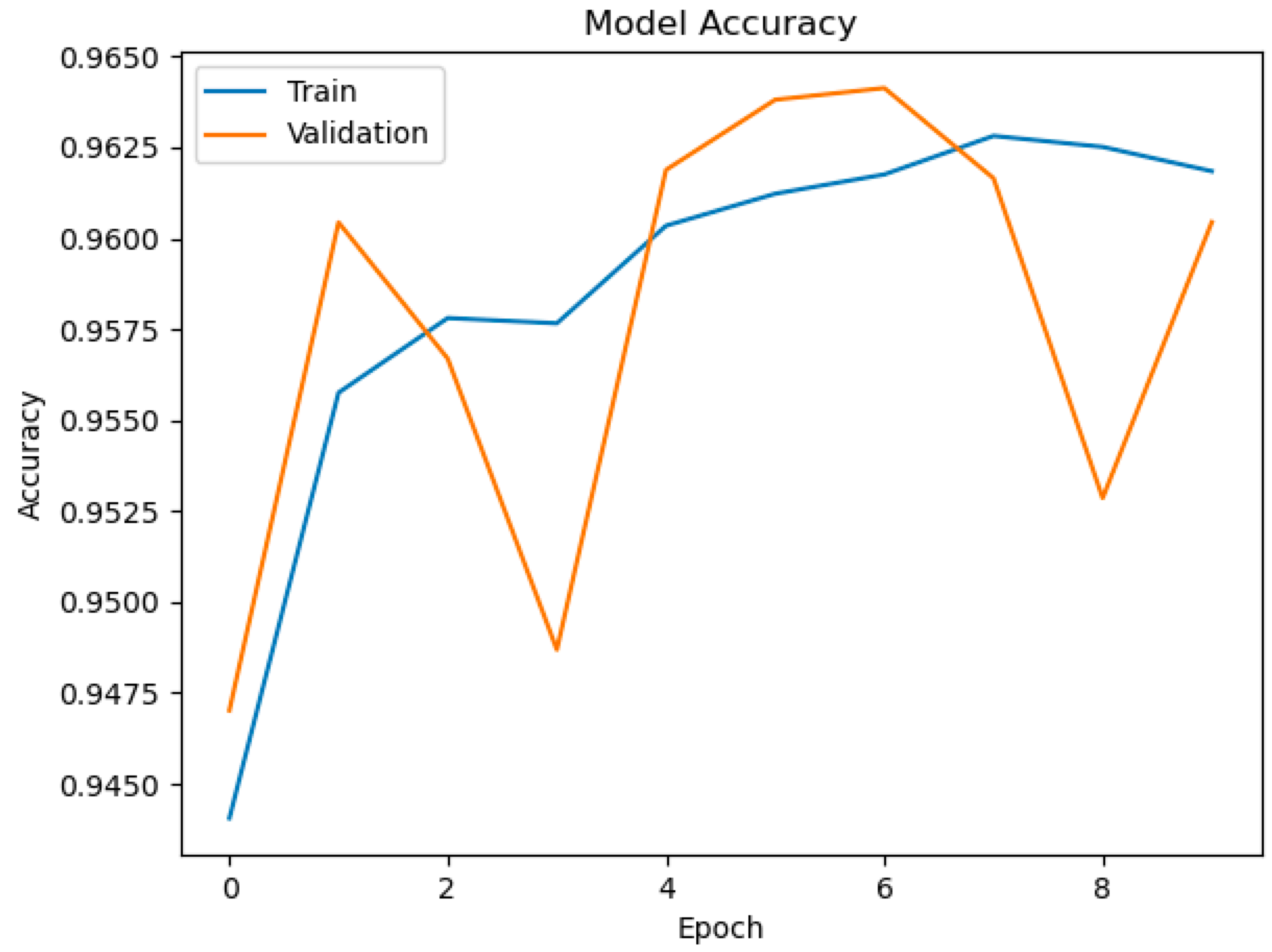


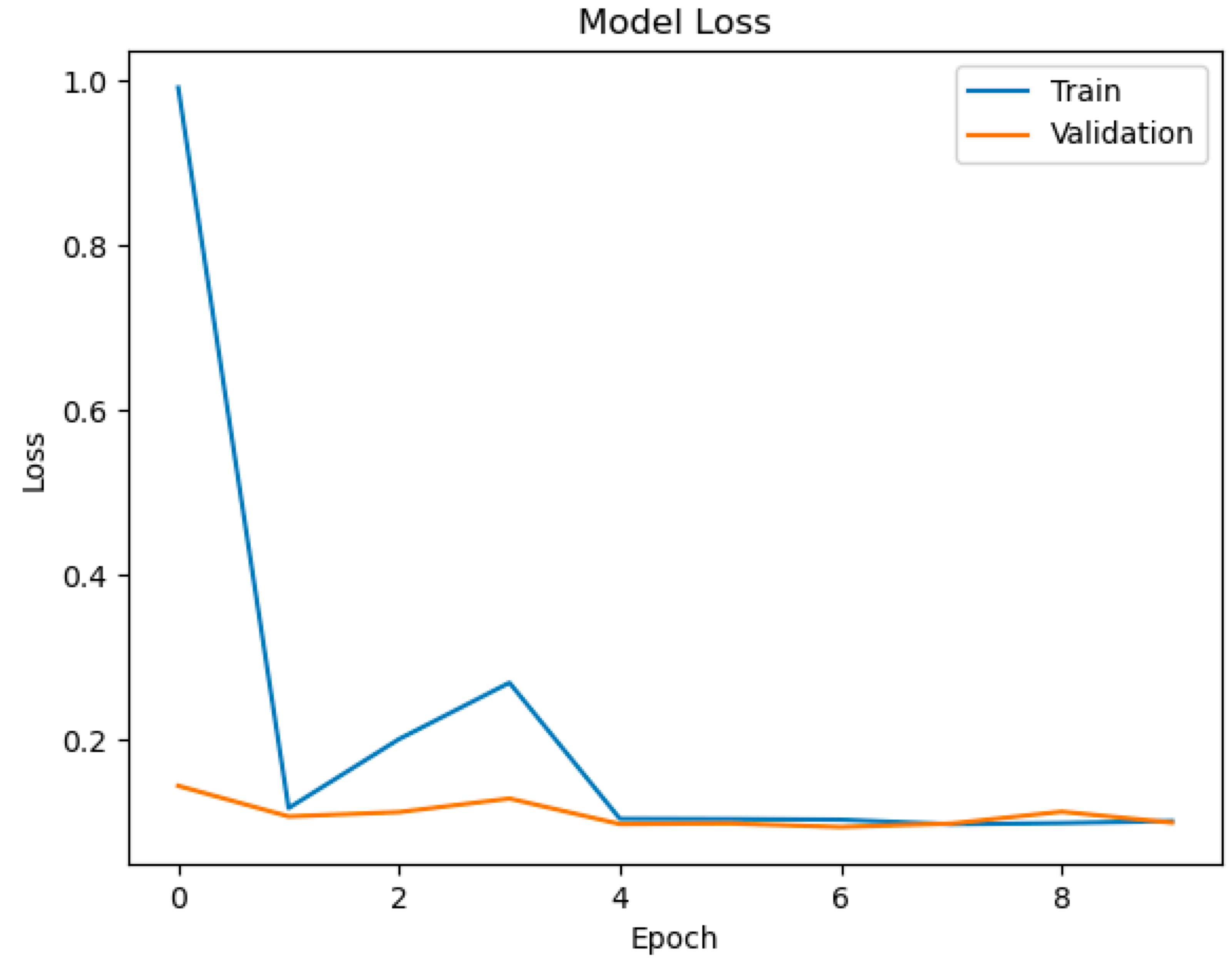


- compute pt ratio of sig/bkg
- reweight each bkg event with the ratio (sig events are weighted with 1)
- provide weights to the CNN during training to over/undercompensate for lack/abundance of signal events in given regions

Training

- Loss function: binary_crossentropy
- Optimizer: Adam(learning_rate=0.001)
- batch_size=128
- Total params: 1248513 (4.76 MB)
- epochs=10 (~2-3 mins each on a 16-core CPU, i.e. on lxplus)



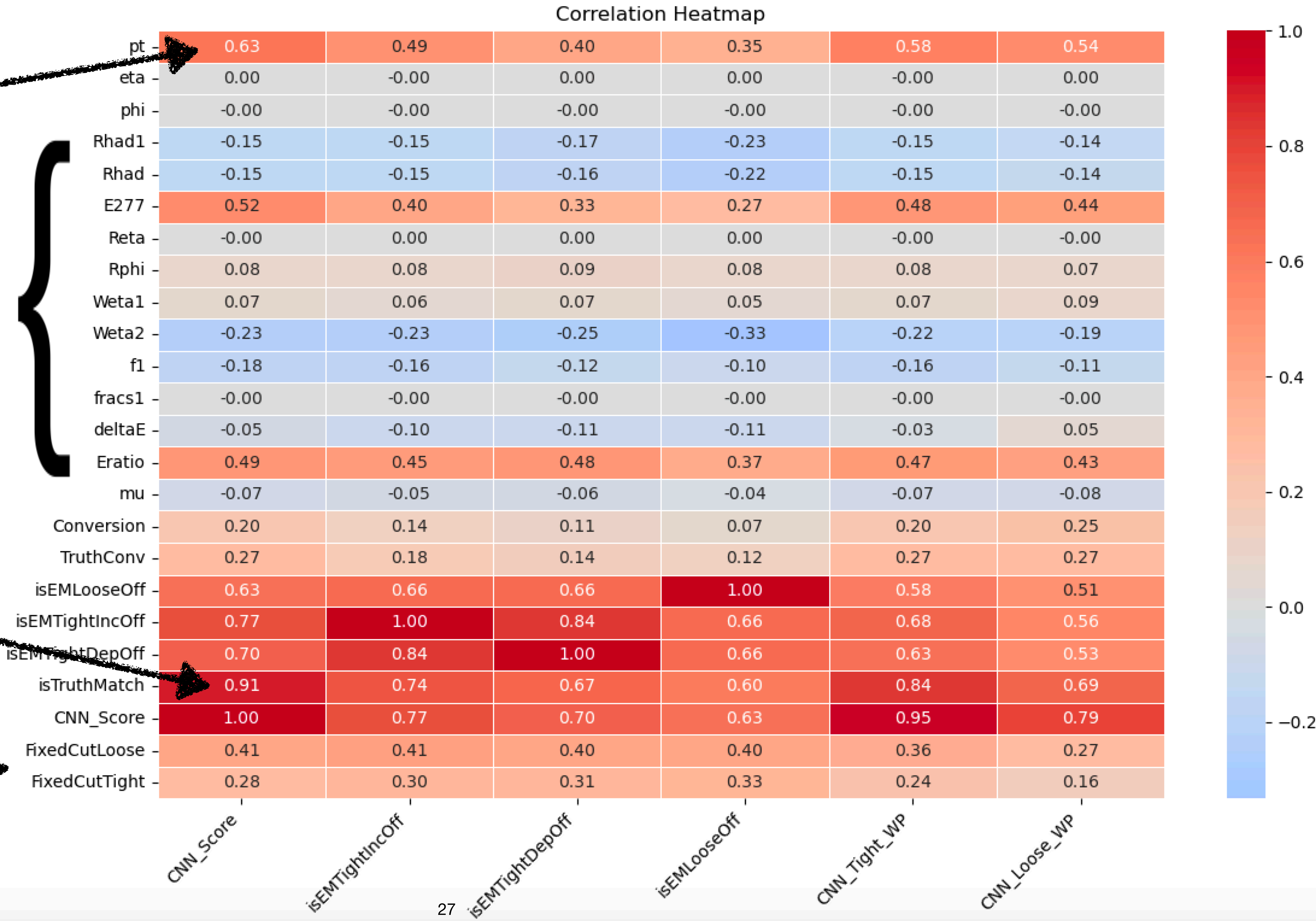


Does the CNN just learn the pt and then use it to separate signal and background?

Shower shapes for reference (they are not used by the CNN)

(Too?) good performance

Same level (or lower) of ID-Iso correlation wrt cut-based WPs



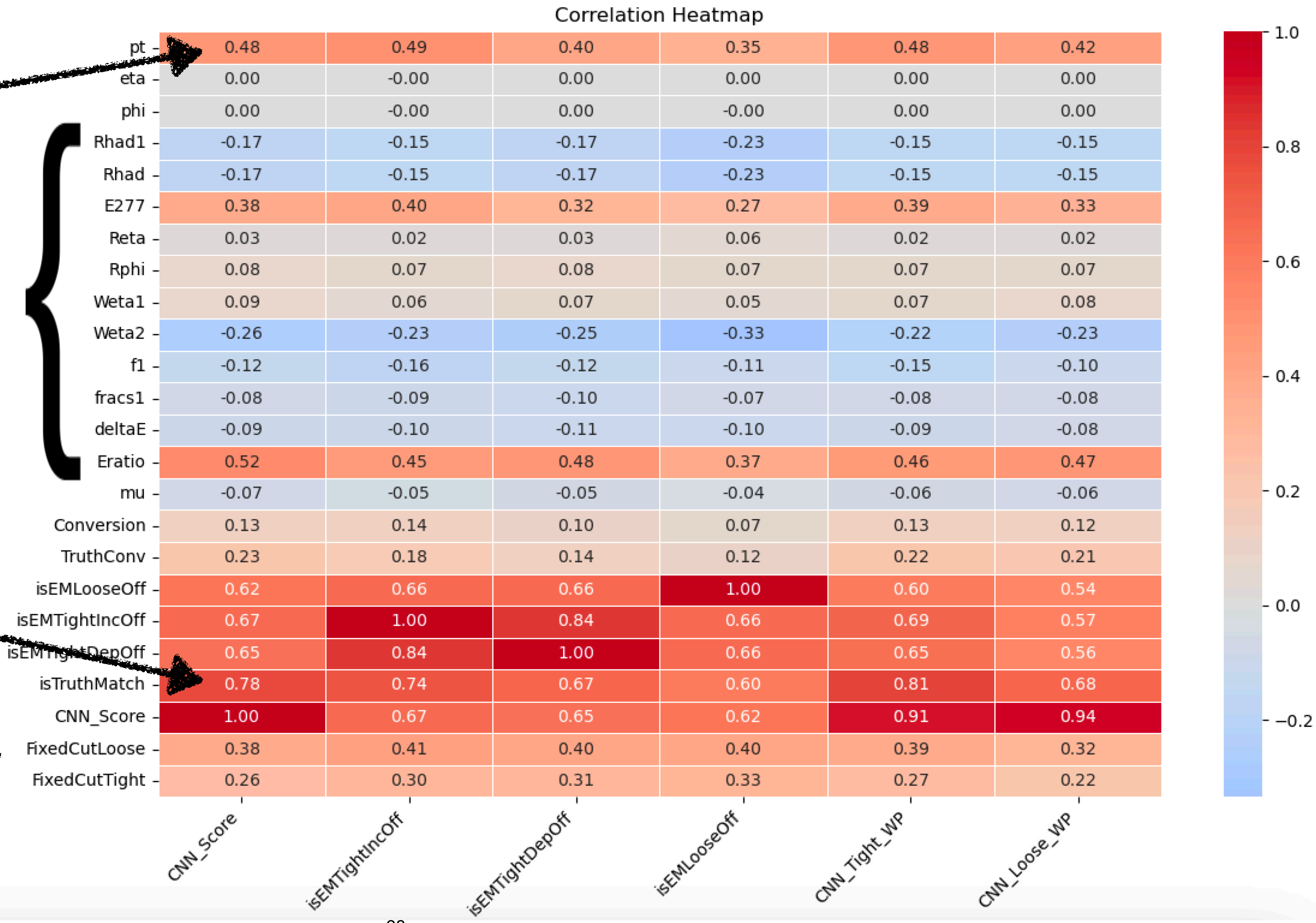
After pT reweighting

Correlation with pt largely reduced…

Shower shapes for reference (they are not used by the CNN)

,,, at the cost of performance

Same level (or lower) of ID-Iso correlation wrt cut-based WPs



Approach #1: split per layer

- Idea: design a multi-output CNN to "mimic" narrow-strip and relaxed-tight cuts

- Narrow-strip variables are computed using energy in Lr1: define a separate CNN output based only on Lr1 input

From Matrix Method INT note (link)

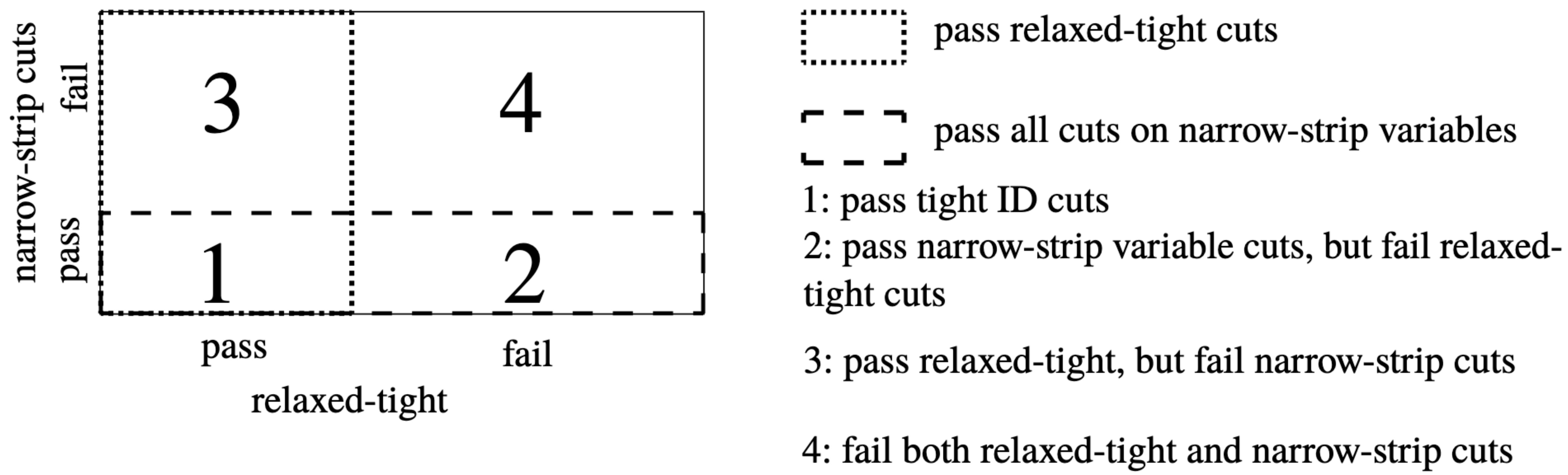
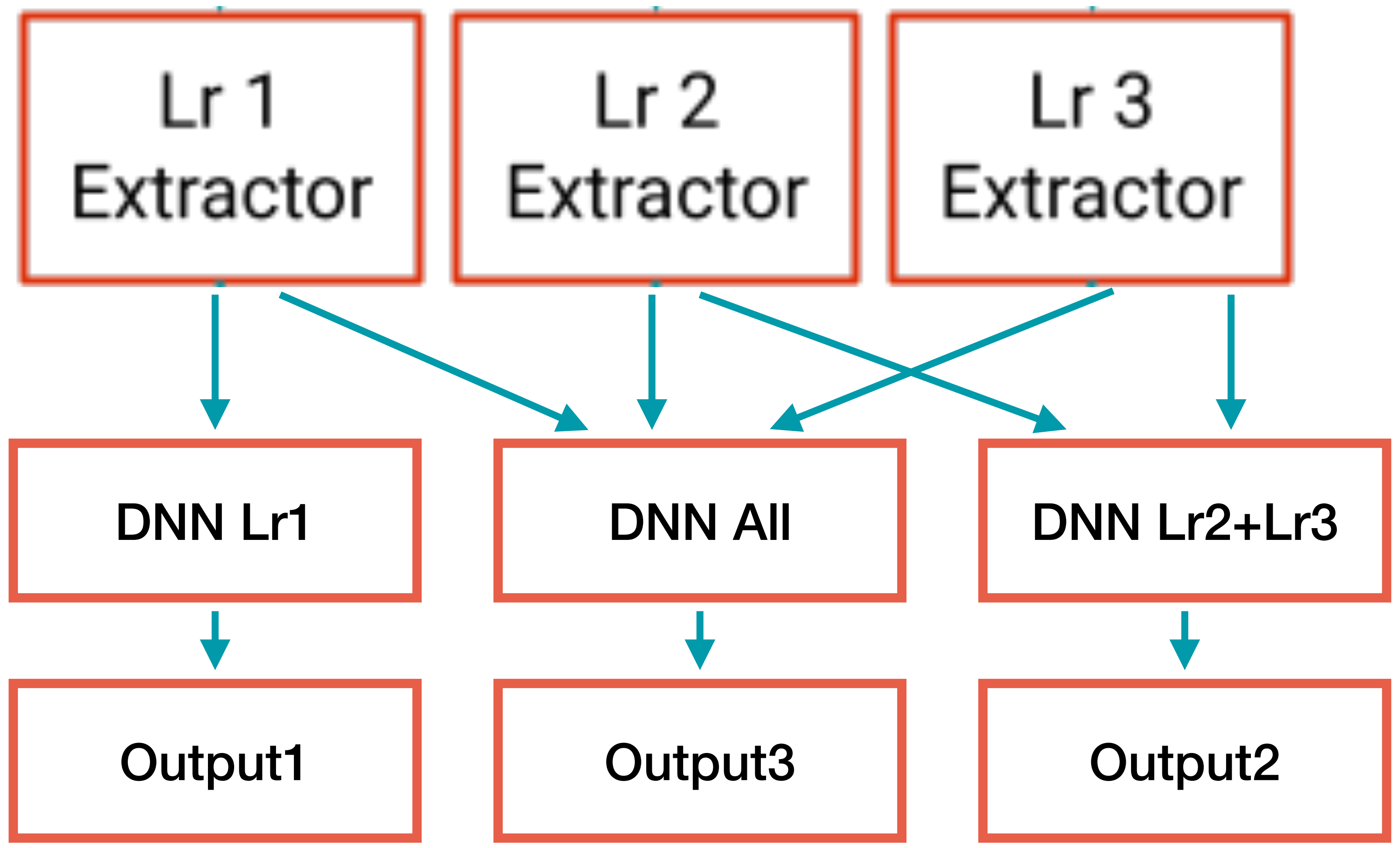
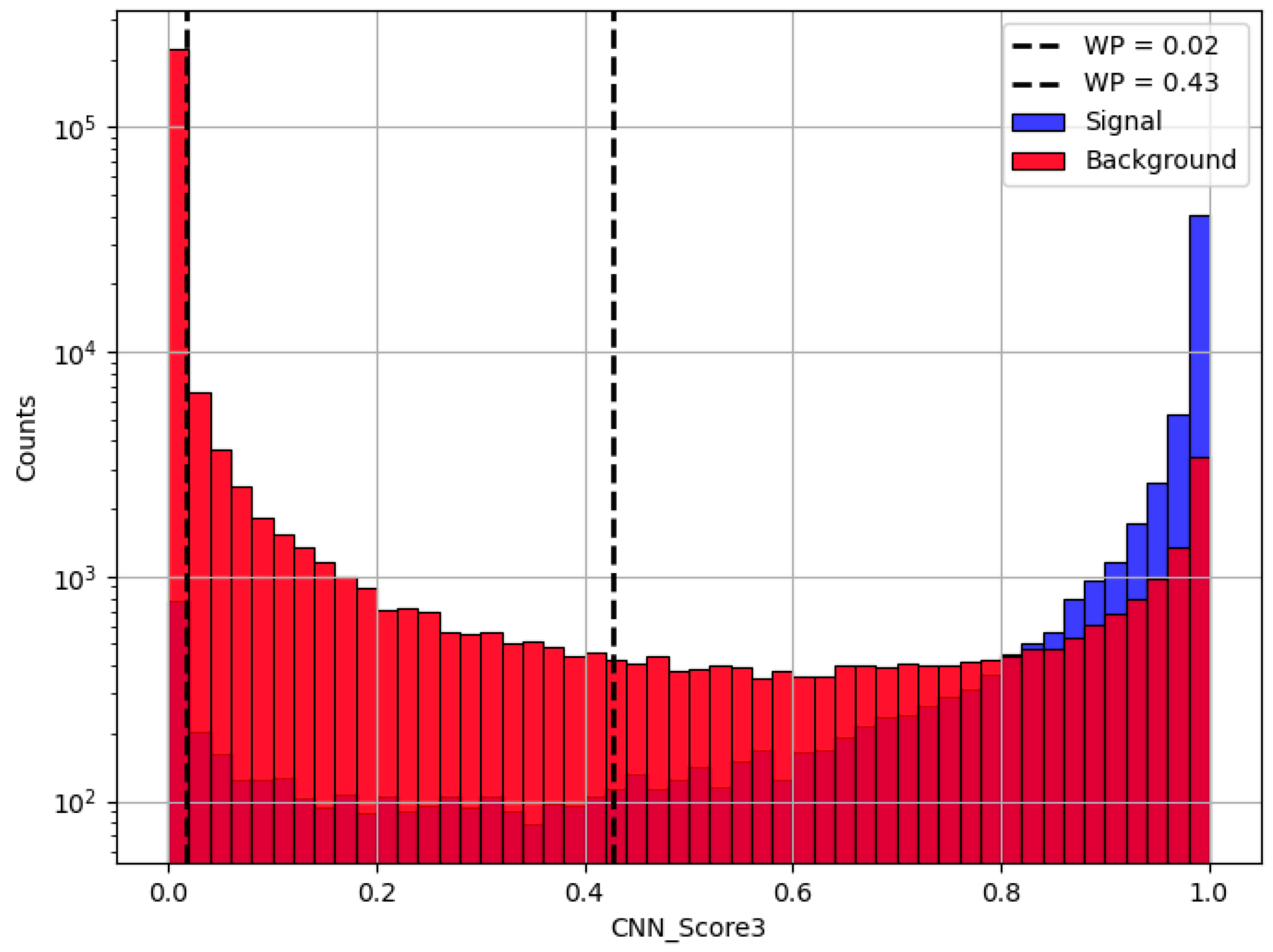


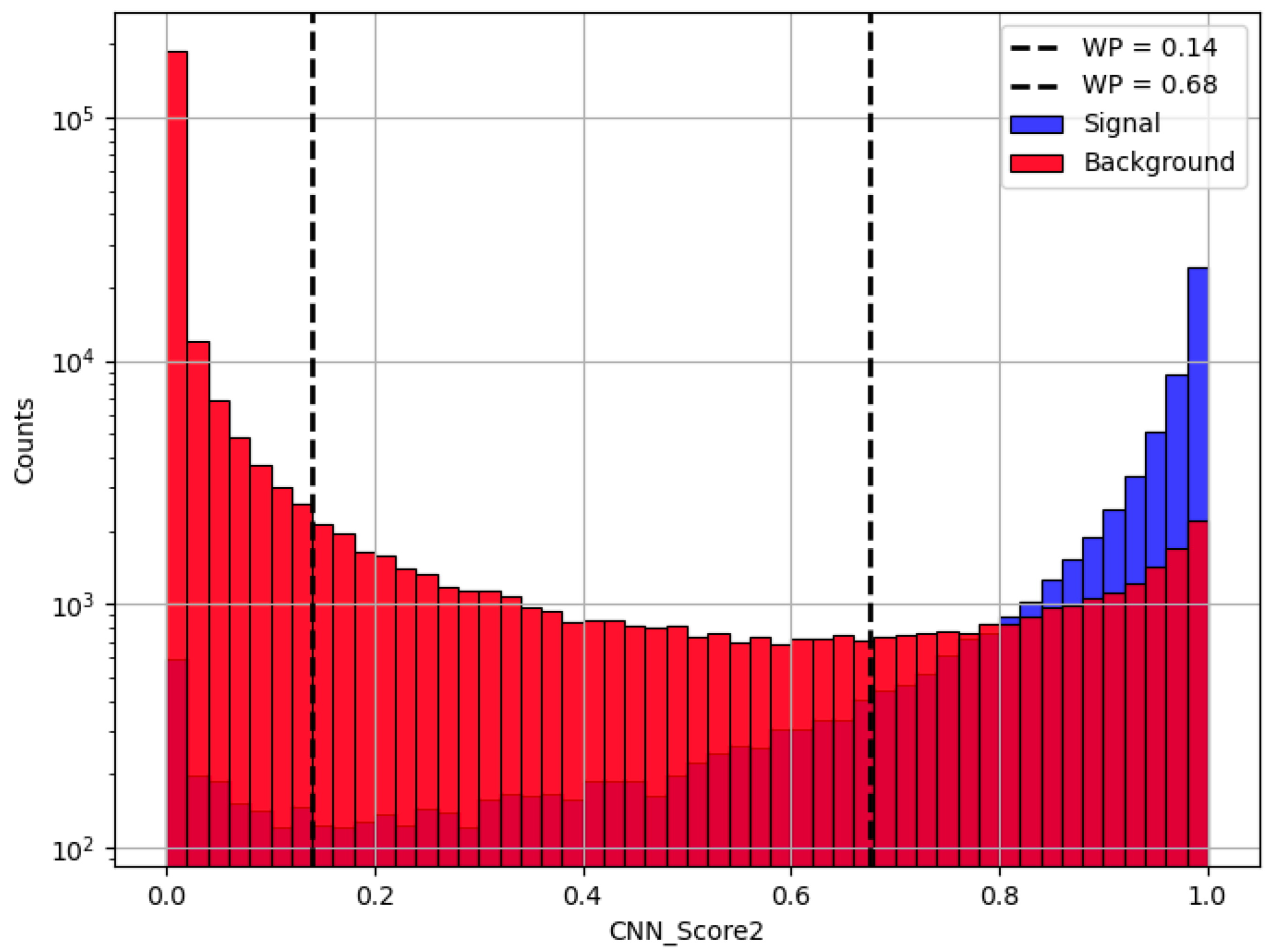
Figure 3: An illustration of the photon classification.

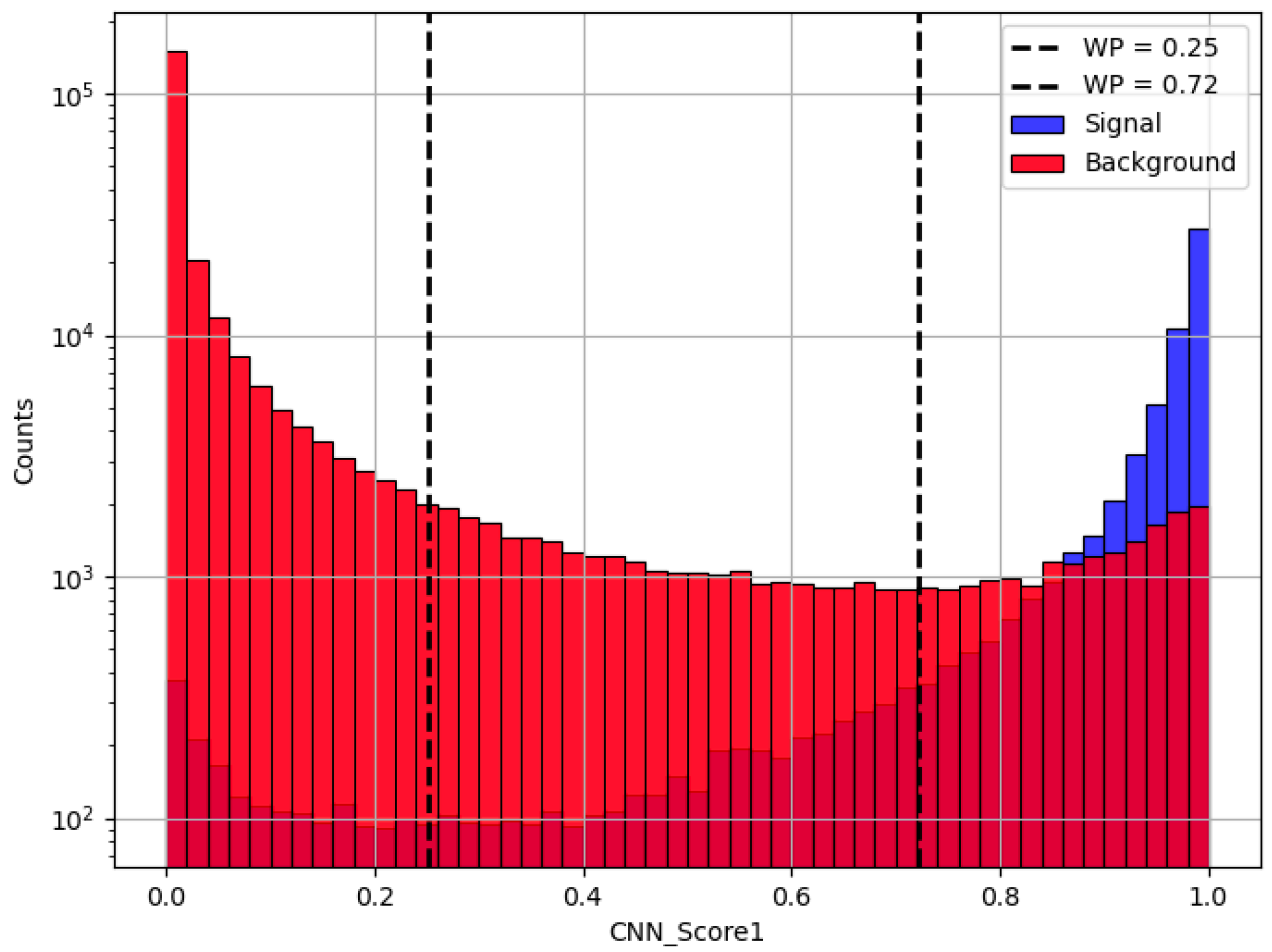


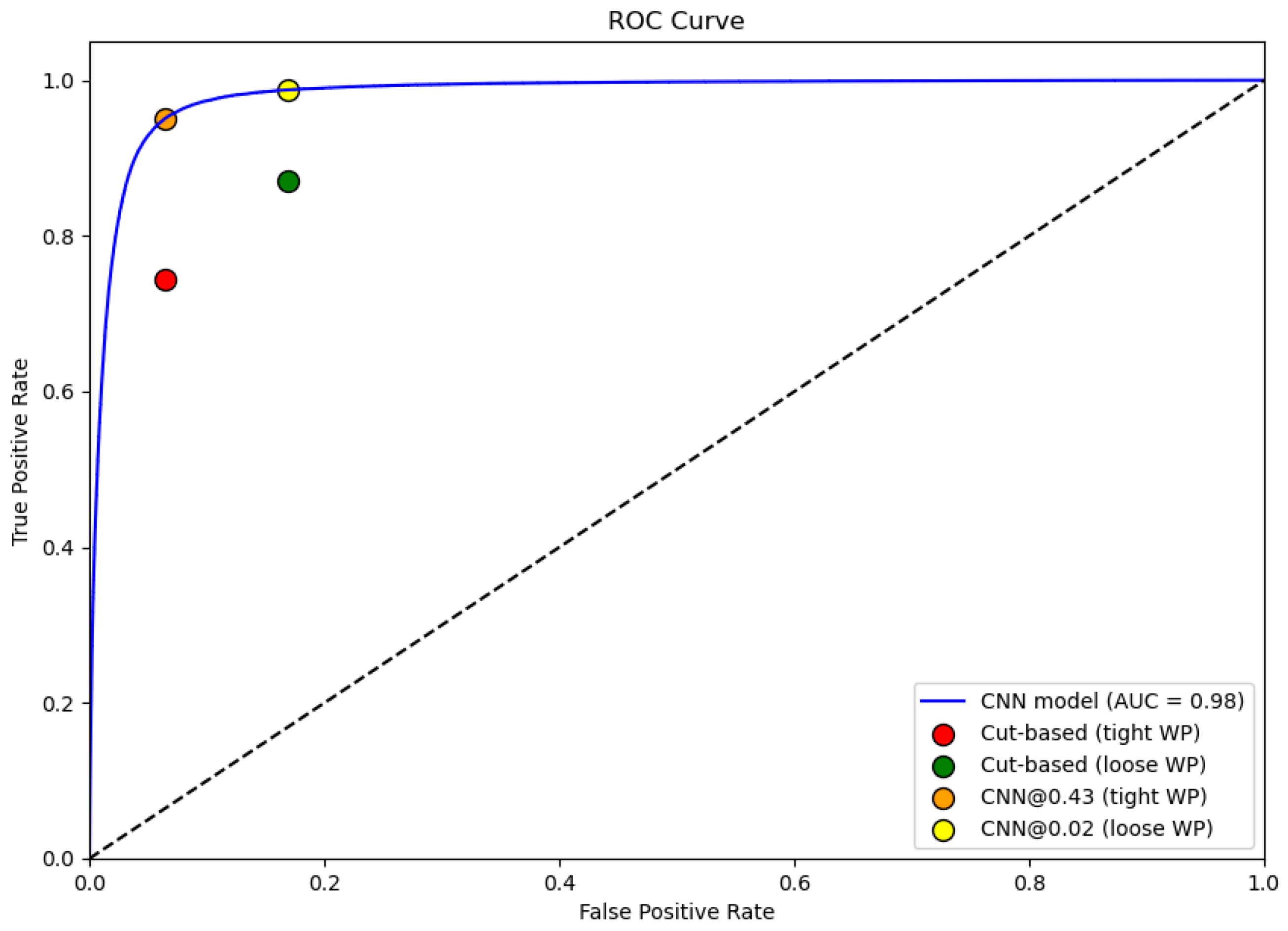
	cut-based	feature	CNN
1	all shower shapes	best performance	final combined output
2	narrow-strip variables	weak correlation with Iso	only Lr1 input
3	relaxed-tight variables	orthogonal (?) to narrow-strip	rest: Lr2, Lr3, HCal (?)

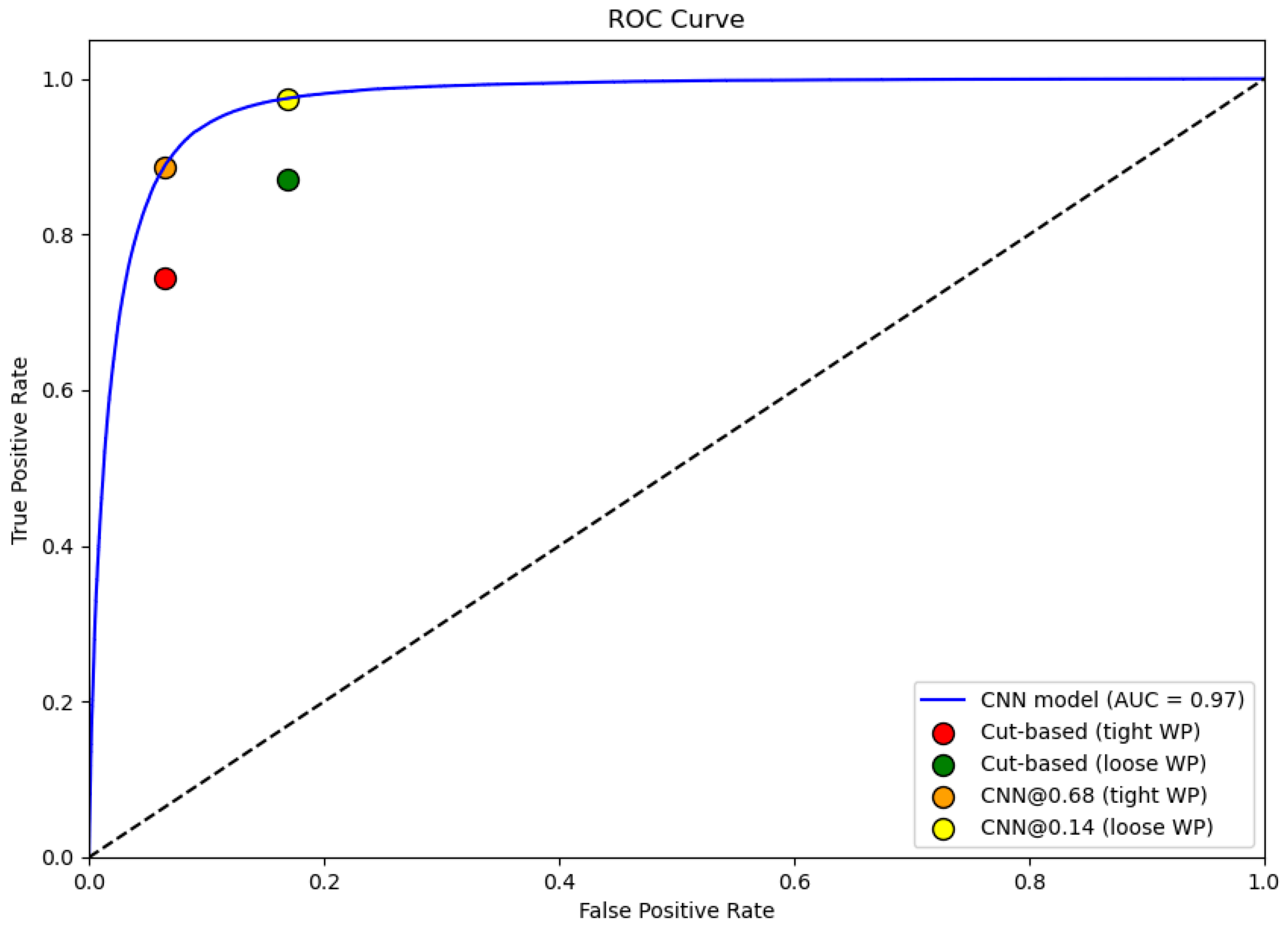
Approach #1: CNN outputs

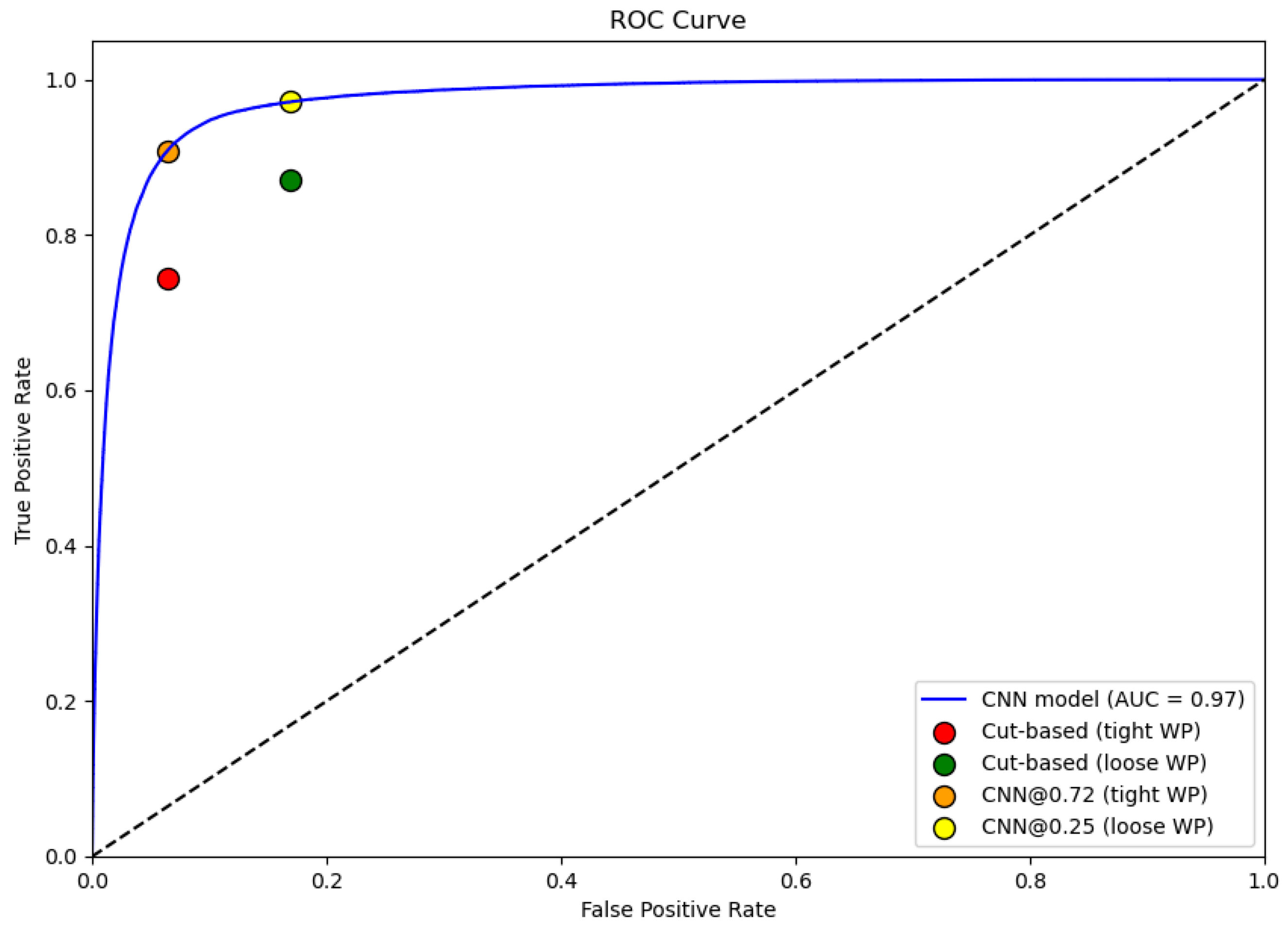




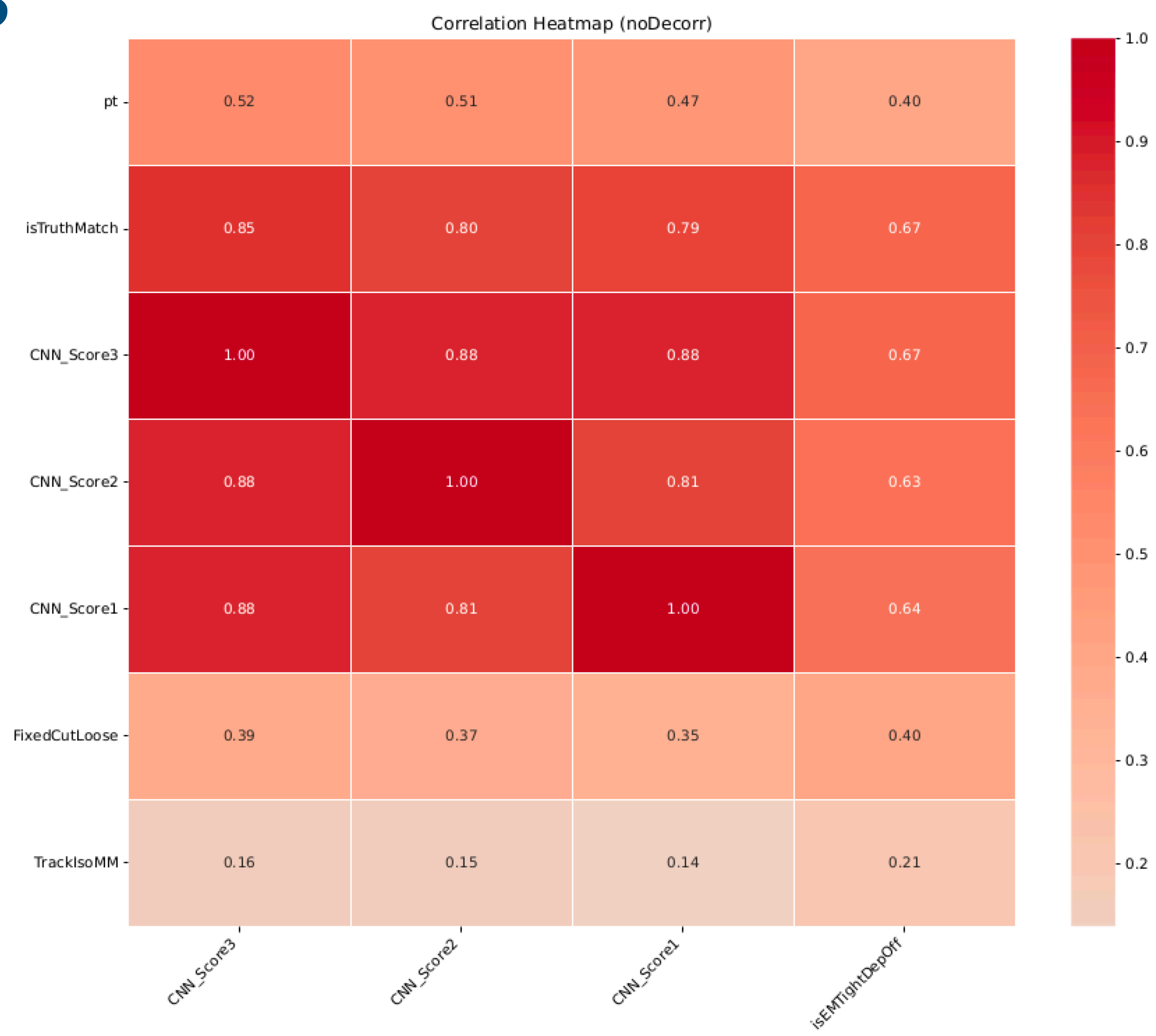








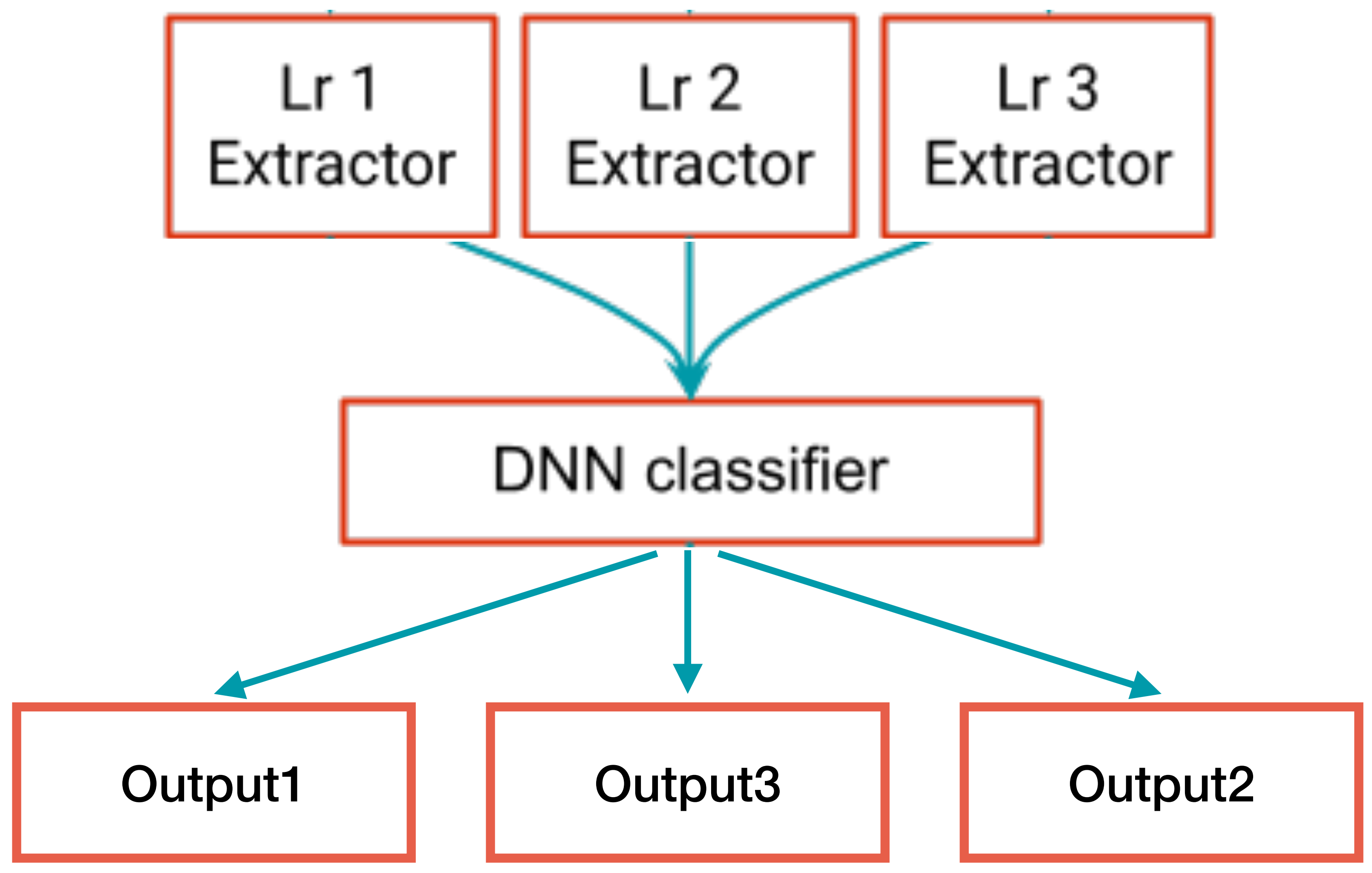
Approach #1: results



Approach #2: decorrelation in the loss

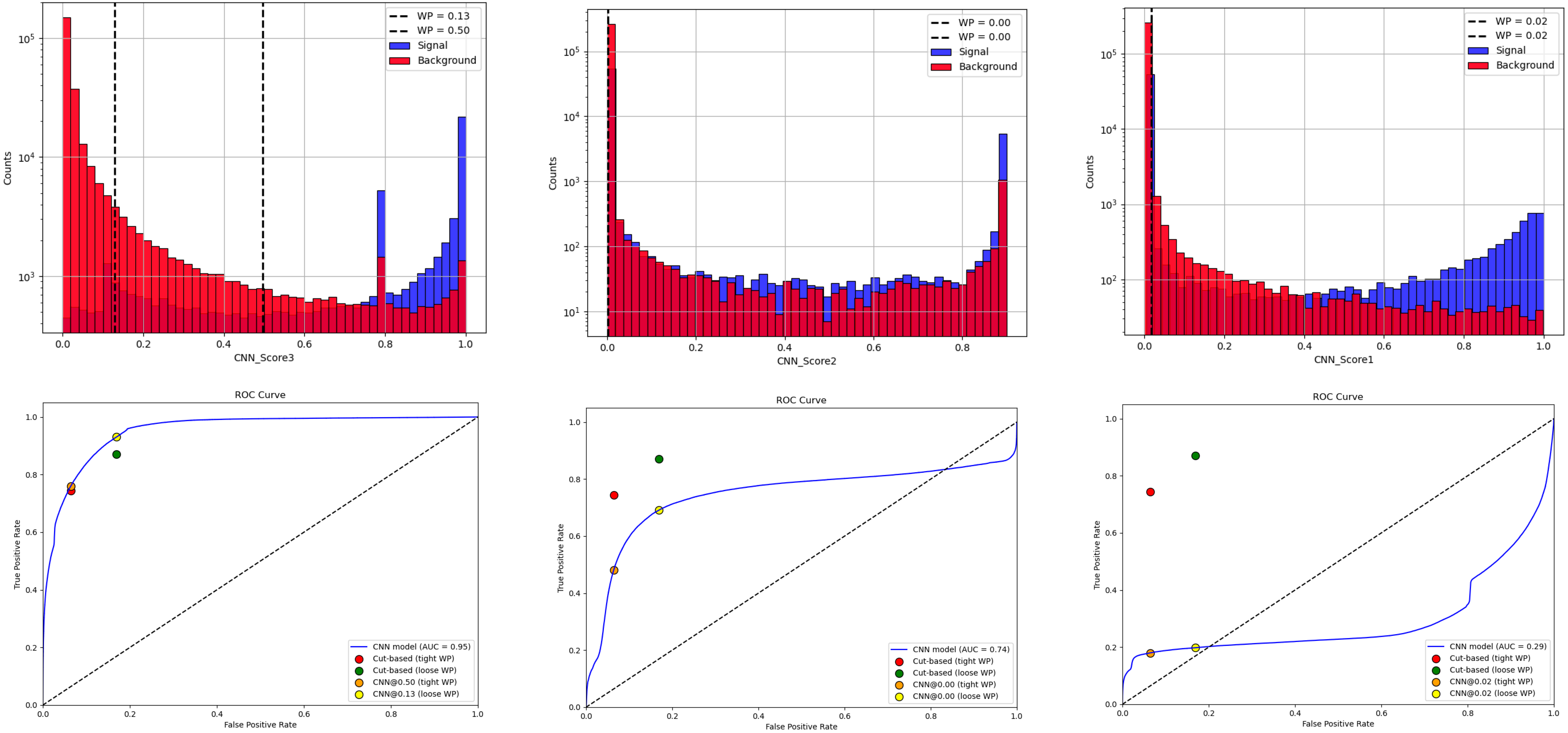
- Idea: design a multi-output CNN to "mimic" narrow-strip and relaxed-tight cuts
- Define a separate CNN output which is explicitly decorrelated w.r.t. track-isolation in the loss

	cut-based	feature	CNN
1	all shower shapes	best performance	no penalization in loss
2	narrow-strip variables	weak correlation with Iso	regularization term w.r.t. TrackIsoMM
3	relaxed-tight variables	orthogonal (?) to narrow-strip	regularization term w.r.t. output1

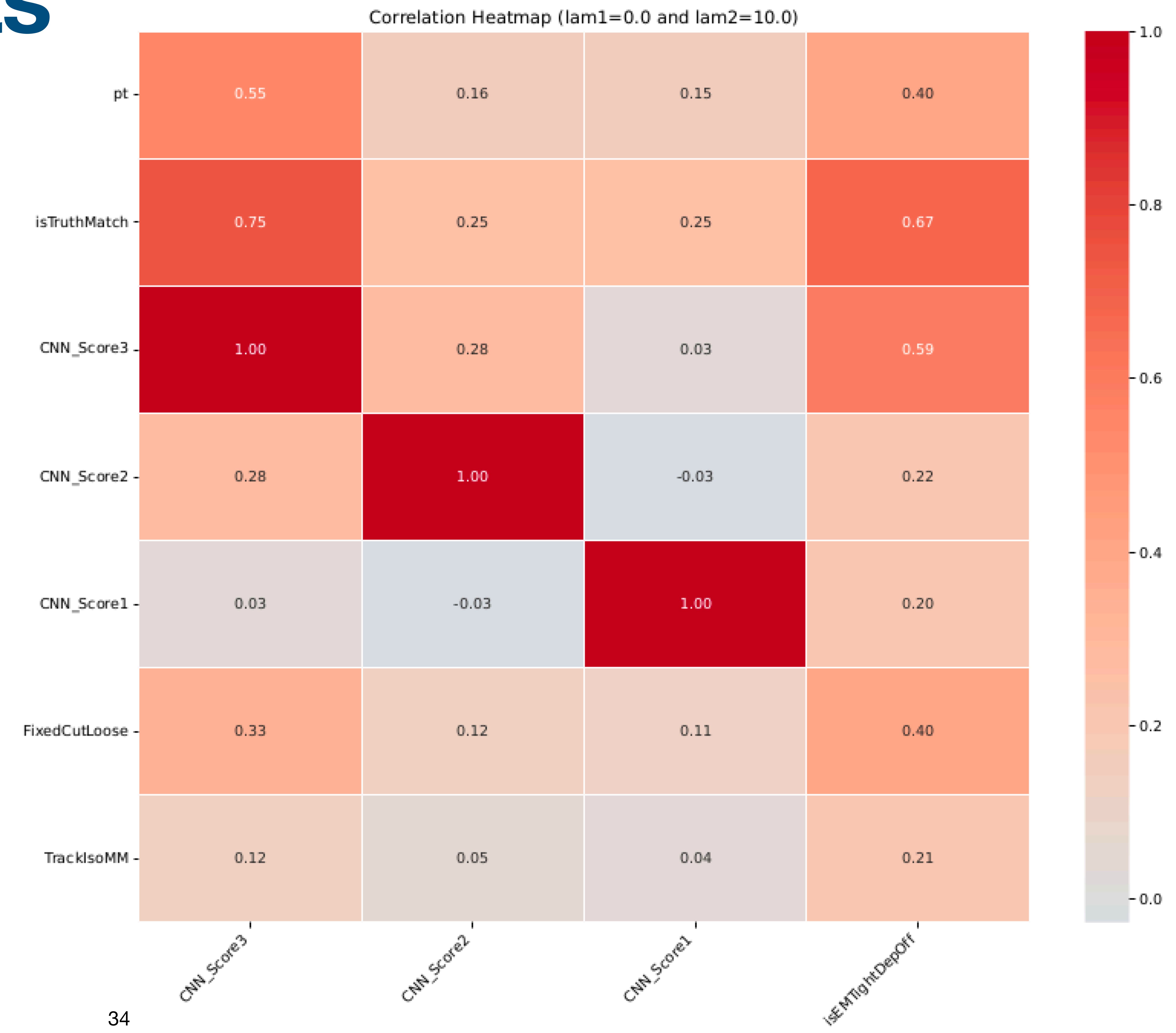


$$\mathscr{L} = \sum_{i=1}^{3} \mathrm{BCE}(\mathrm{Out_i}, \mathrm{y}) + \lambda_1 \mathrm{DisCo}(\mathrm{Out}_1, \mathrm{ptcone40}) + \lambda_2 \mathrm{DisCo}(\mathrm{Out}_2, \mathrm{Out}_1)$$

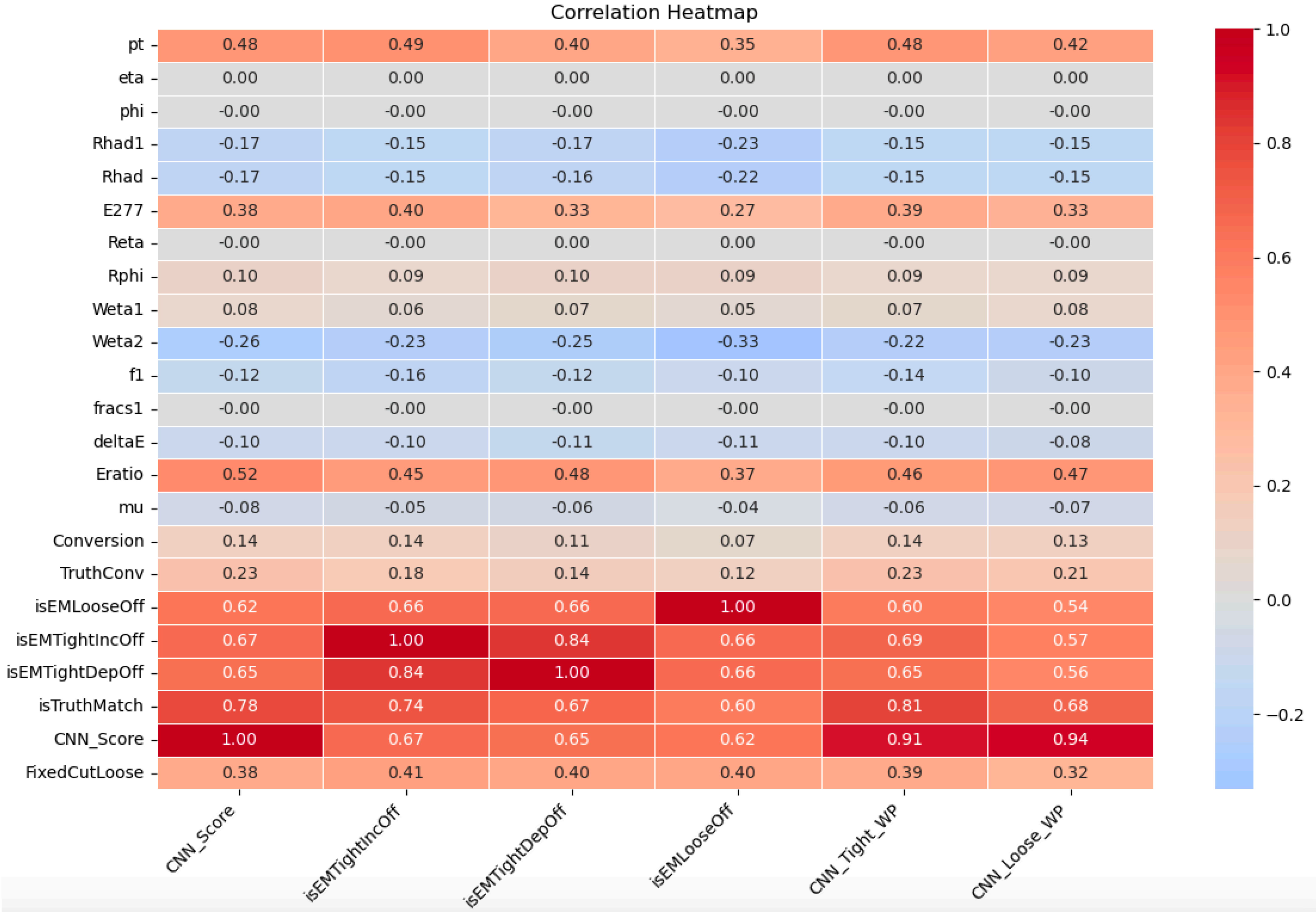
Approach #2: CNN outputs

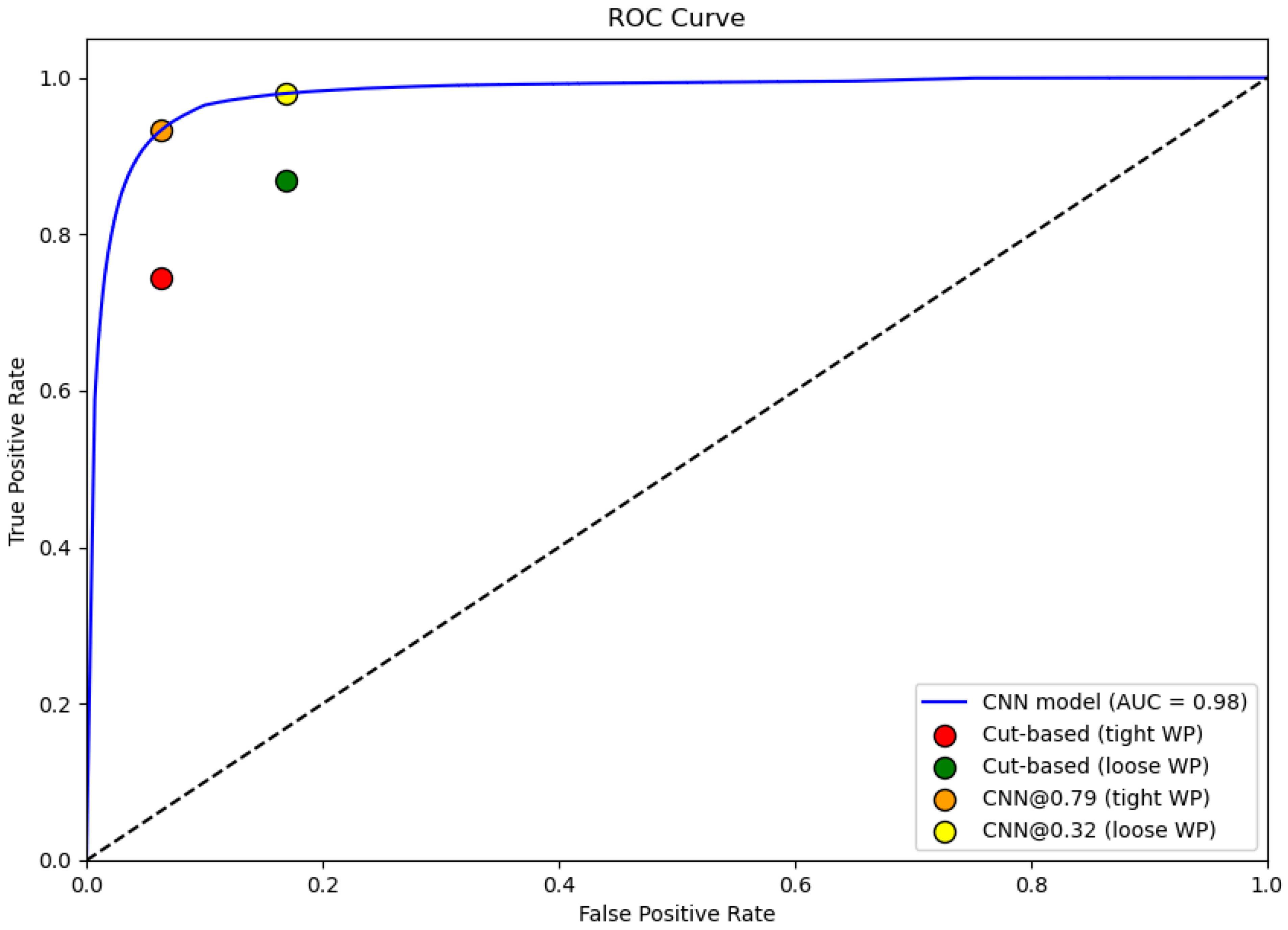


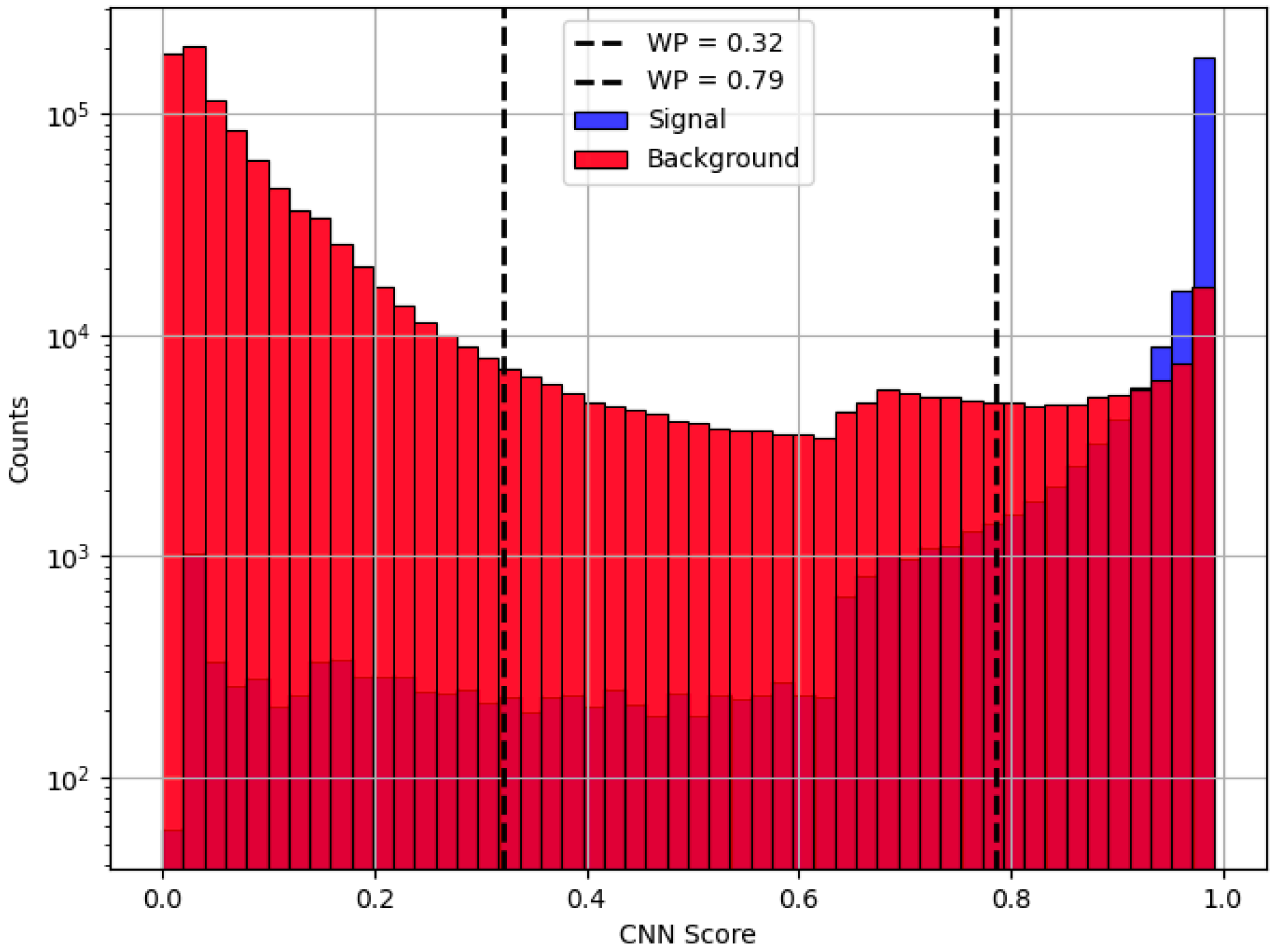
Approach #2: results



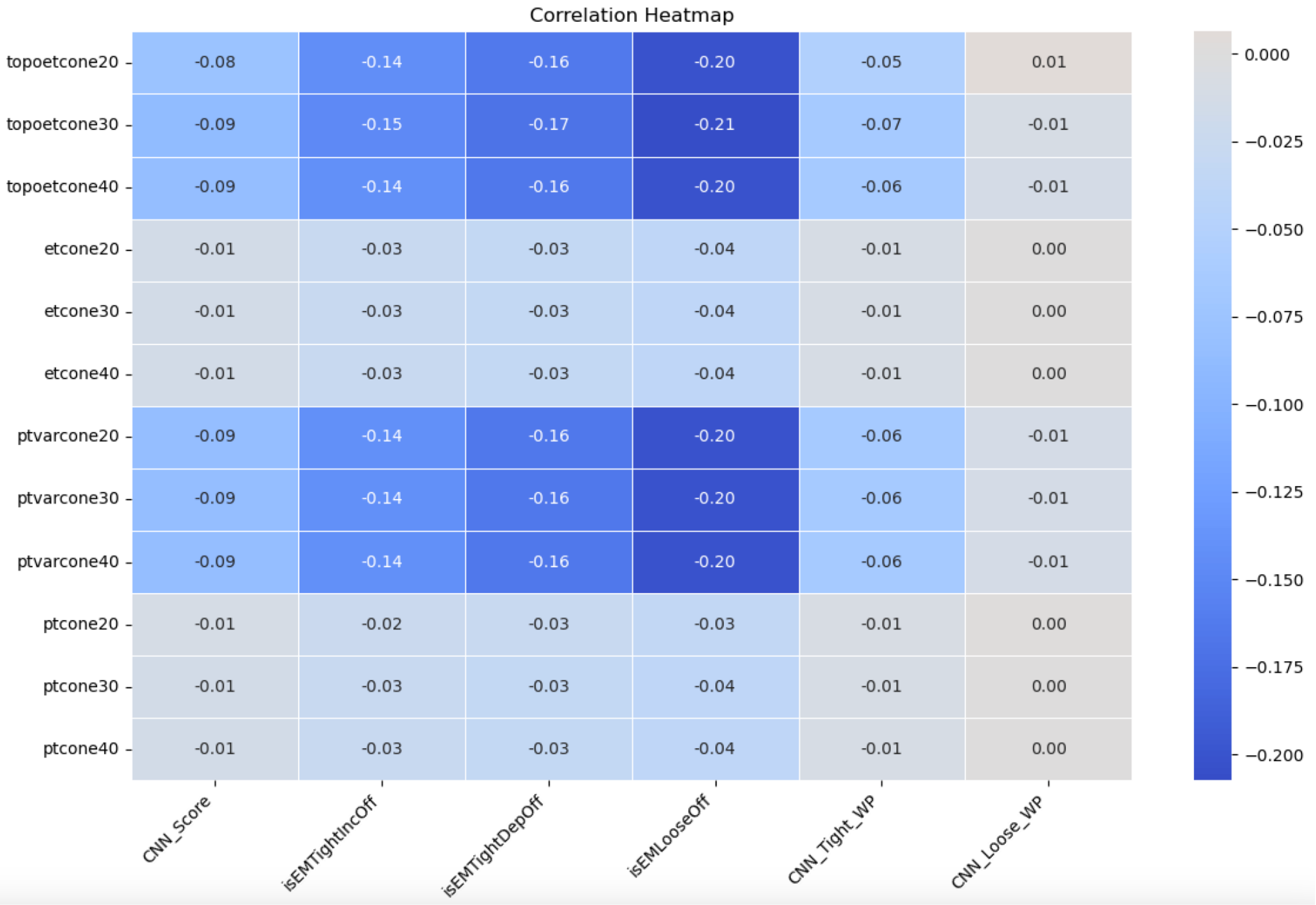
Train set


Correlation Heatmap
pt 0.48 0.49 0.40 0.35 0.48 0.42
eta 0.00 0.00 0.00 0.00 0.00 0.00
phi -0.00 -0.00 -0.00 -0.00 -0.00 -0.00
Rhad1 -0.17 -0.15 -0.17 -0.23 -0.15 -0.15
Rhad -0.17 -0.15 -0.16 -0.22 -0.15 -0.15
E277 0.38 0.40 0.33 0.27 0.39 0.33
Reta -0.00 -0.00 0.00 0.00 -0.00 -0.00
Rphi 0.10 0.09 0.10 0.09 0.09 0.09
Weta1 0.08 0.06 0.07 0.05 0.07 0.08
Weta2 -0.26 -0.23 -0.25 -0.33 -0.22 -0.23
f1 -0.12 -0.16 -0.12 -0.10 -0.14 -0.10
fracs1 -0.00 -0.00 -0.00 -0.00 -0.00 -0.00
deltaE -0.10 -0.10 -0.11 -0.11 -0.10 -0.08
Eratio 0.52 0.45 0.48 0.37 0.46 0.47
mu -0.08 -0.05 -0.06 -0.04 -0.06 -0.07
Conversion 0.14 0.14 0.11 0.07 0.14 0.13
TruthConv 0.23 0.18 0.14 0.12 0.23 0.21
isEMLooseOff 0.62 0.66 0.66 1.00 0.60 0.54
isEMTightIncOff 0.67 1.00 0.84 0.66 0.69 0.57
isEMTightDepOff 0.65 0.84 1.00 0.66 0.65 0.56
isTruthMatch 0.78 0.74 0.67 0.60 0.81 0.68
CNN_Score 1.00 0.67 0.65 0.62 0.91 0.94
FixedCutLoose 0.38 0.41 0.40 0.40 0.39 0.32
CNN_Score
isEMTightIncOff
isEMTightDepOff
isEMLooseOff
CNN_Tight_WP
CNN_Loose_WP
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

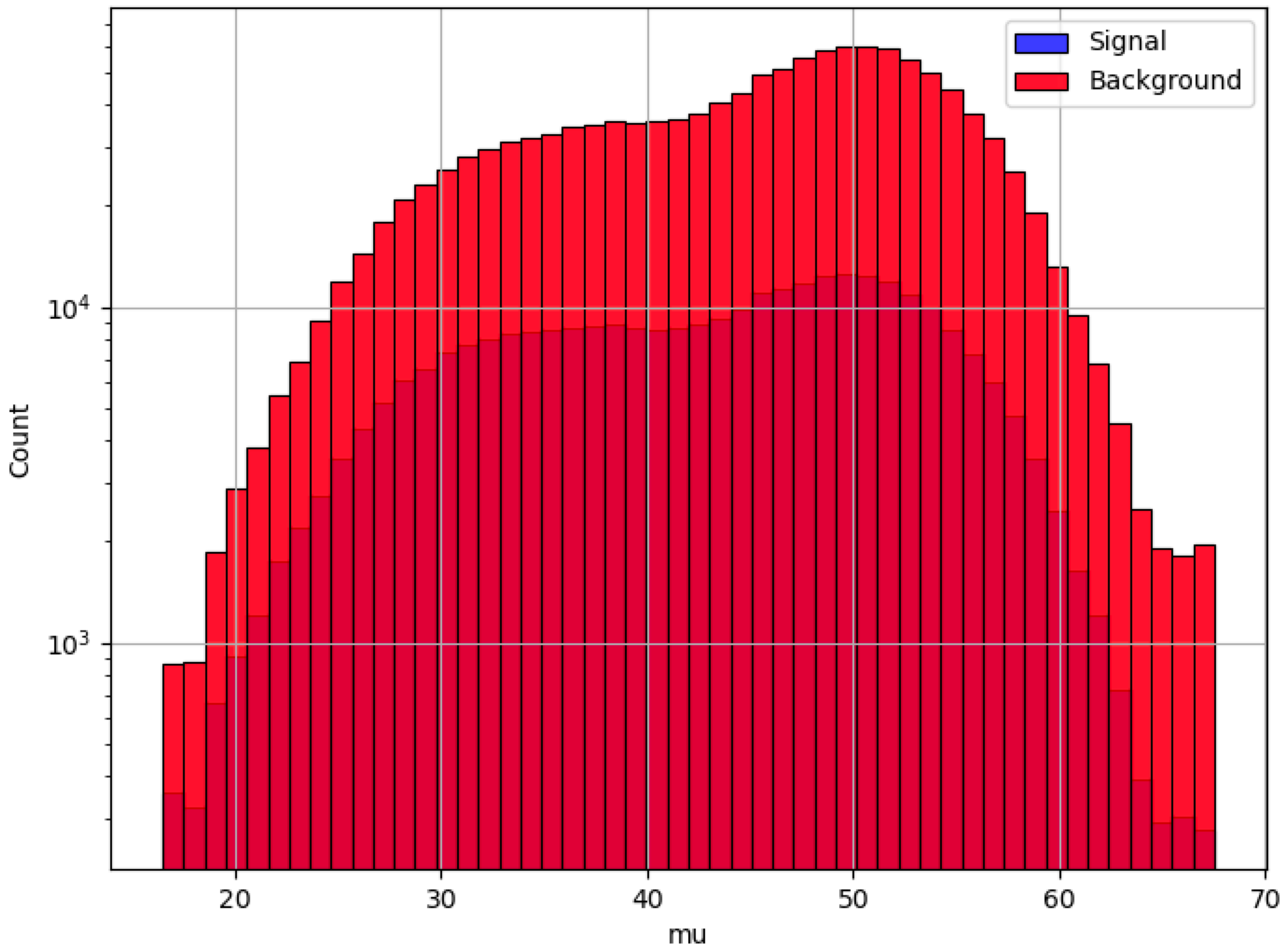

ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
CNN model (AUC = 0.98)
Cut-based (tight WP)
Cut-based (loose WP)
CNN@0.79 (tight WP)
CNN@0.32 (loose WP)


WP = 0.32
WP = 0.79
Signal
Background
Counts
CNN Score
10^5
10^4
10^3
10^2
0.0
0.2
0.4
0.6
0.8
1.0

Correlation wrt Isolation variables



Pile-up



Model definition

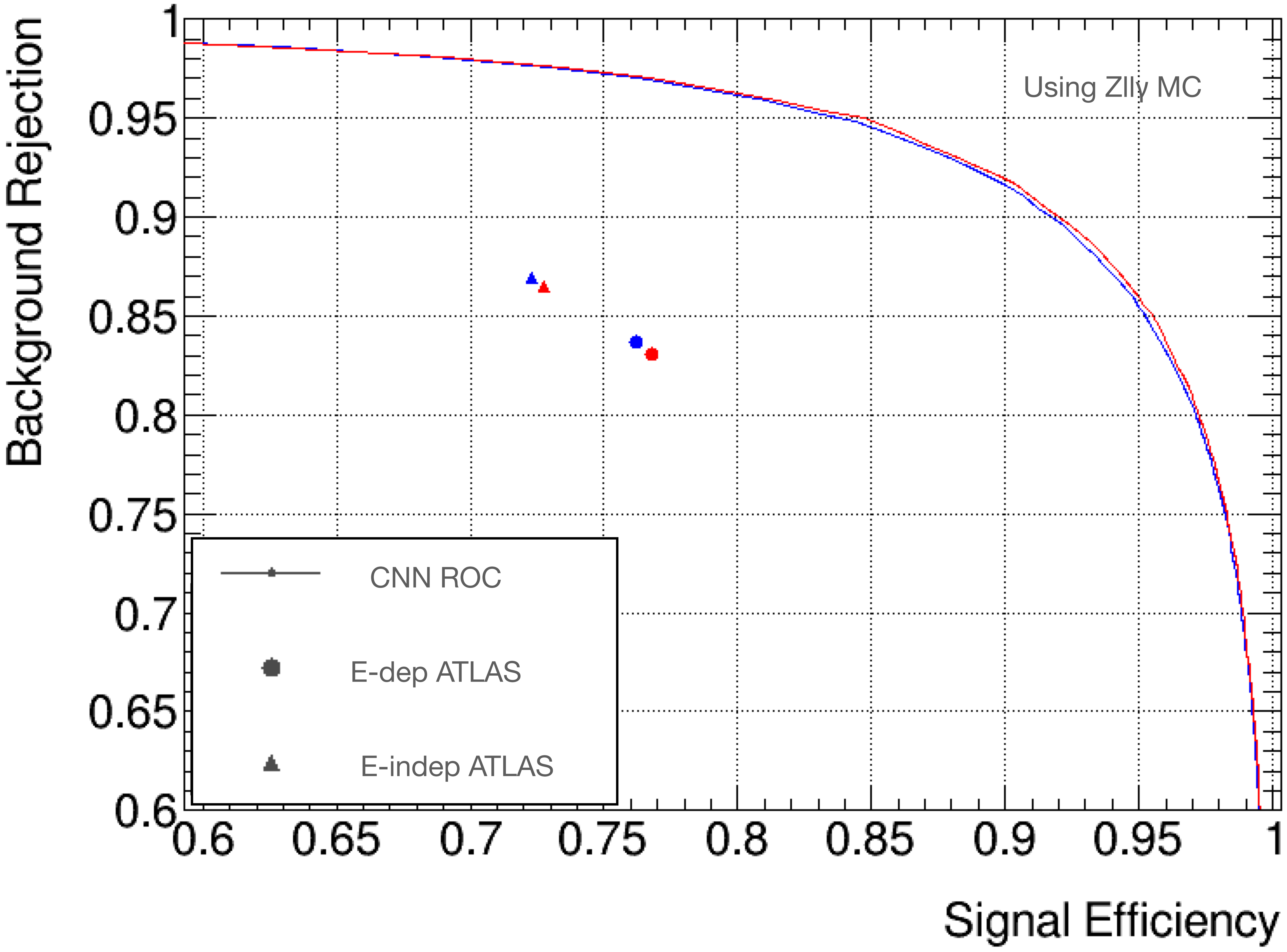
Model: "model"

Layer (type)	Output Shape	Param #	Connected to
Input1 (InputLayer)	[(None, 2, 56, 1)]	0	[]
Input2 (InputLayer)	[(None, 11, 7, 1)]	0	[]
Input3 (InputLayer)	[(None, 11, 4, 1)]	0	[]
Cov11 (Conv2D)	(None, 1, 55, 128)	640	['Input1[0][0]']
Cov21 (Conv2D)	(None, 10, 6, 128)	640	['Input2[0][0]']
Cov31 (Conv2D)	(None, 10, 3, 128)	640	['Input3[0][0]']
MaxP11 (MaxPooling2D)	(None, 1, 27, 128)	0	['Cov11[0][0]']
MaxP21 (MaxPooling2D)	(None, 5, 6, 128)	0	['Cov21[0][0]']
MaxP31 (MaxPooling2D)	(None, 5, 3, 128)	0	['Cov31[0][0]']
Cov12 (Conv2D)	(None, 1, 26, 128)	32896	['MaxP11[0][0]']
Cov22 (Conv2D)	(None, 4, 6, 128)	32896	['MaxP21[0][0]']
Cov32 (Conv2D)	(None, 4, 3, 128)	32896	['MaxP31[0][0]']
MaxP12 (MaxPooling2D)	(None, 1, 13, 128)	0	['Cov12[0][0]']
MaxP22 (MaxPooling2D)	(None, 2, 6, 128)	0	['Cov22[0][0]']
MaxP32 (MaxPooling2D)	(None, 2, 3, 128)	0	['Cov32[0][0]']
Flat1 (Flatten)	(None, 1664)	0	['MaxP12[0][0]']
Flat2 (Flatten)	(None, 1536)	0	['MaxP22[0][0]']
Flat3 (Flatten)	(None, 768)	0	['MaxP32[0][0]']
Concatenate (Concatenate)	(None, 3968)	0	['Flat1[0][0]', 'Flat2[0][0]', 'Flat3[0][0]']
D1 (Dense)	(None, 256)	1016064	['Concatenate[0][0]']
gaussian_noise (GaussianNoise)	(None, 256)	0	['D1[0][0]']
D2 (Dense)	(None, 256)	65792	['gaussian_noise[0][0]']
gaussian_noise_1 (GaussianNoise)	(None, 256)	0	['D2[0][0]']
D3 (Dense)	(None, 256)	65792	['gaussian_noise_1[0][0]']
Output (Dense)	(None, 1)	257	['D3[0][0]']

Total params: 1248513 (4.76 MB)
Trainable params: 1248513 (4.76 MB)
Non-trainable params: 0 (0.00 Byte)

Iso-ID correlation (1)

- CNN trained with Tight Isolation
- Applied to
  - **Loose Iso**
  - **Tight Iso**
- No change in performance of the algorithm



Iso-ID correlation (2)

WP name	Definition
FixedCutLoose	topoetcone20 < 0.065 p_T && ptcone20/p_T < 0.05
FixedCutTight	topoetcone40 < 0.022 p_T + 2.45 [GeV] && ptcone20/p_T < 0.05
TightCaloOnly	topoetcone40 < 0.022 p_T + 2.45 [GeV]

- CNN trained with Tight Isolation
- Applied to
  - **Topoetcone20/pt<-0.1**
  - **Topoetcone20/pt>=-0.1**
- CNN has same trend as ATLAS tight WPs: more isolated worse performance, because you remove info around the center of the cluster