Heavy-Flavour Jet Tagging with Graph Neural Networks at LHCb

Gabriella Pesticci (Kenyon College)
Dr. Conor Henderson, Dr. Nathan Grieser (University of Cincinnati)

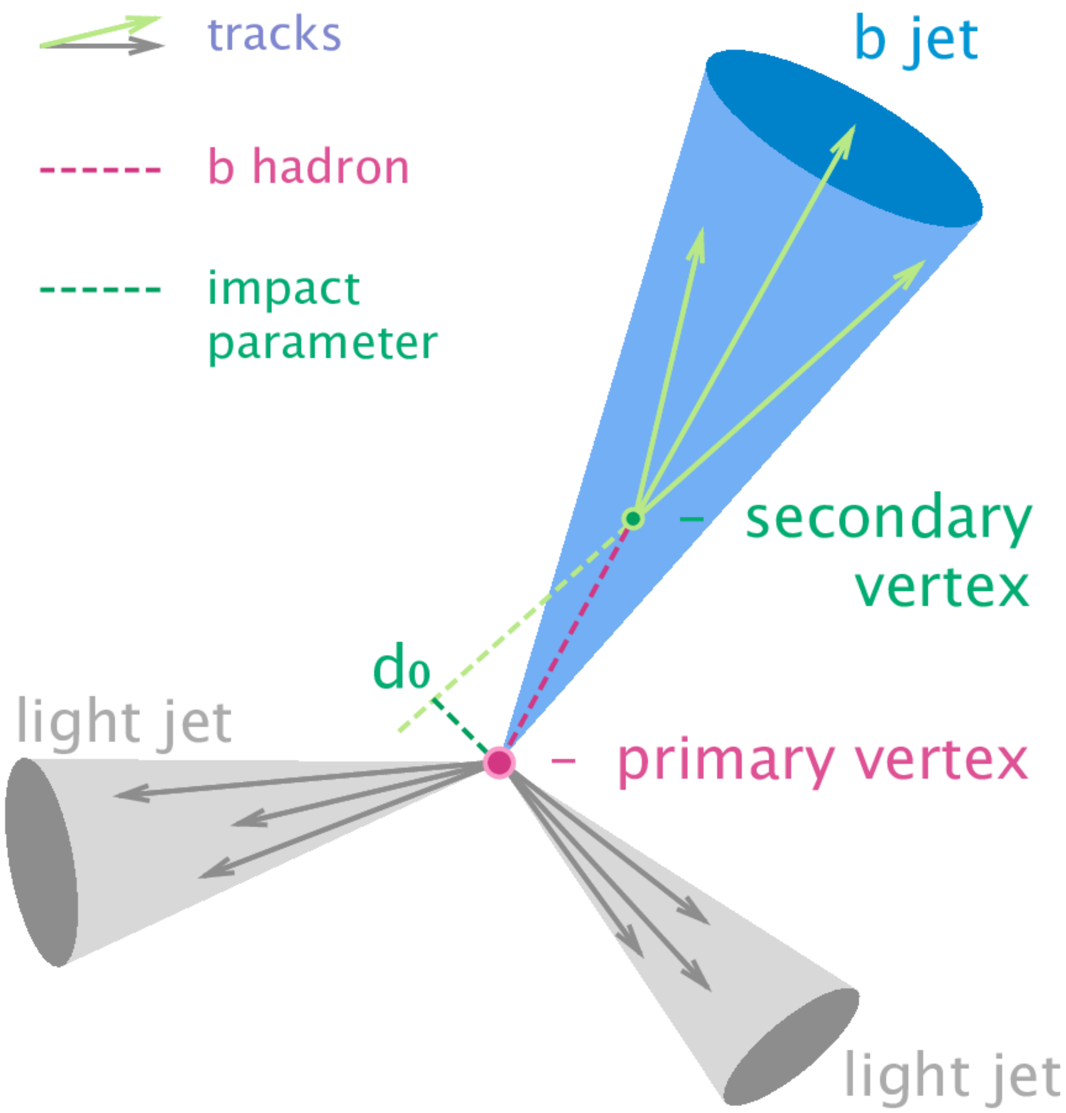
US LUA Annual Meeting 2024

Dec 18, 2024

Hadronization

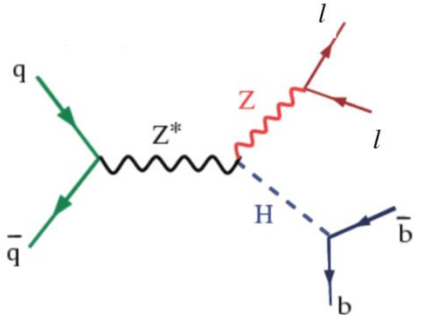
- *pp* collisions at LHC produce quarks and gluons which hadronize due to QCD confinement
- Longer lifetime of heavy B-hadron creates a characteristic secondary vertex

Goal: Classify jet flavours (*b*, *c*, *q*)

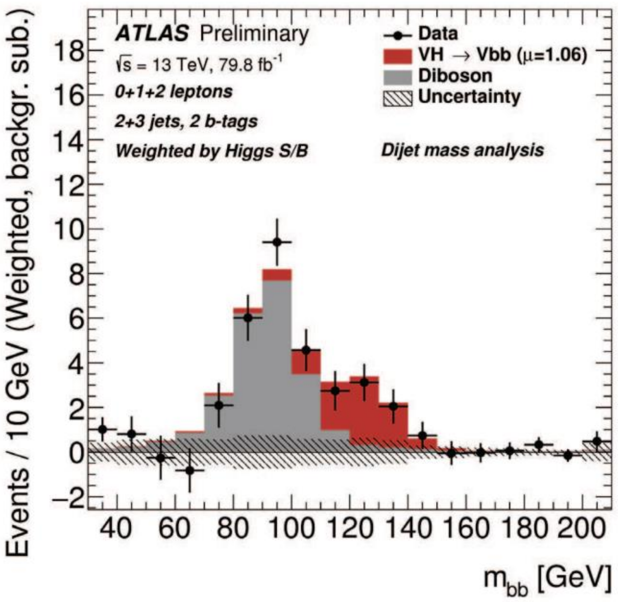


Selection

$H \rightarrow b\bar{b}$

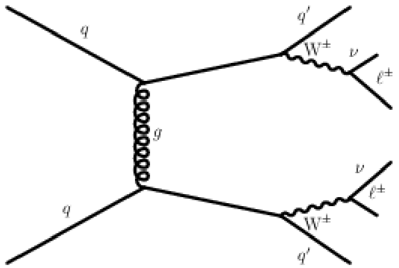


ATLAS Higgs Hunting



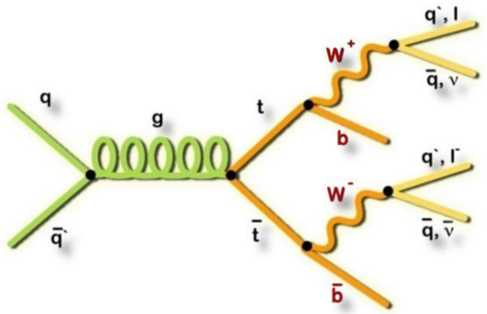
ATLAS Higgs Hunting

Veto

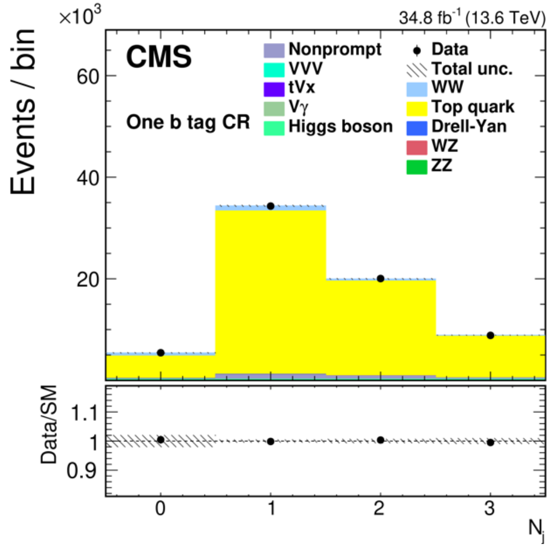


CMS-SMP-19-012

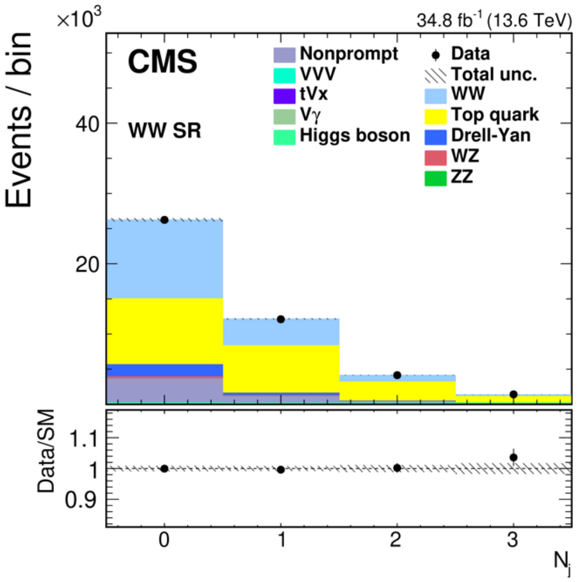
WW analysis



ATLAS b-tagging



$N_{b-jet} = 1$

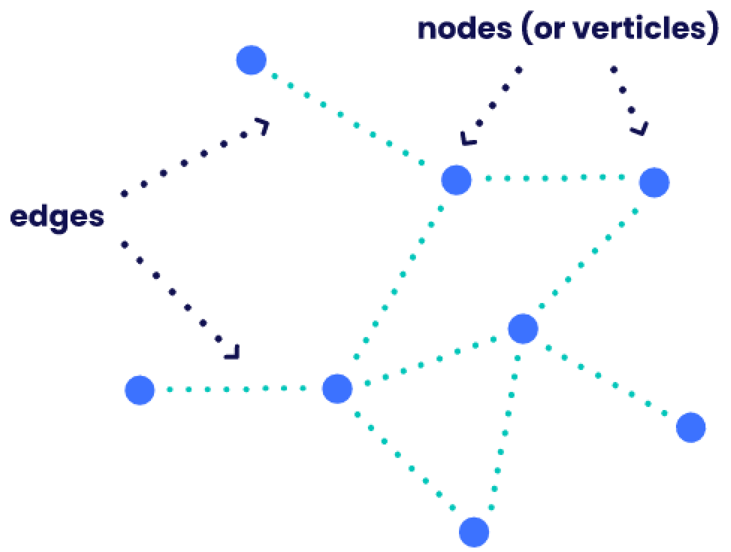


$N_{b-jet} = 0$

CMS-SMP-24-001

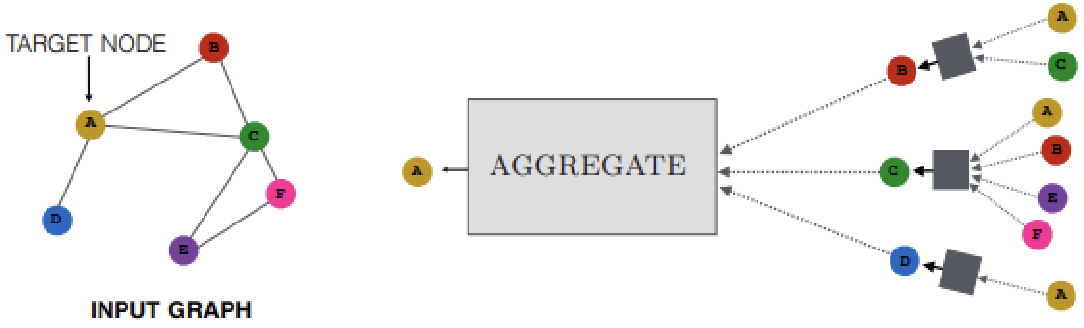
Graph Neural Networks (GNN)

- Variable number of nodes and edges
- Captures complex relationships to represent the system



Message Passing

- Node information aggregated from neighbors
    - Target node updated
    - Learn features of neighbors



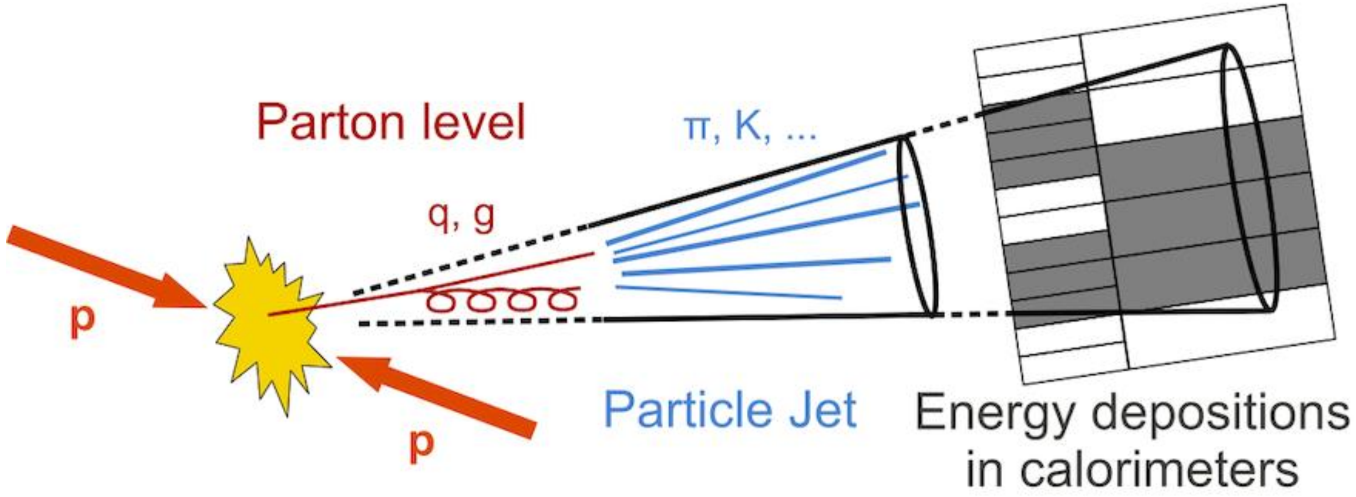
Graph Construction

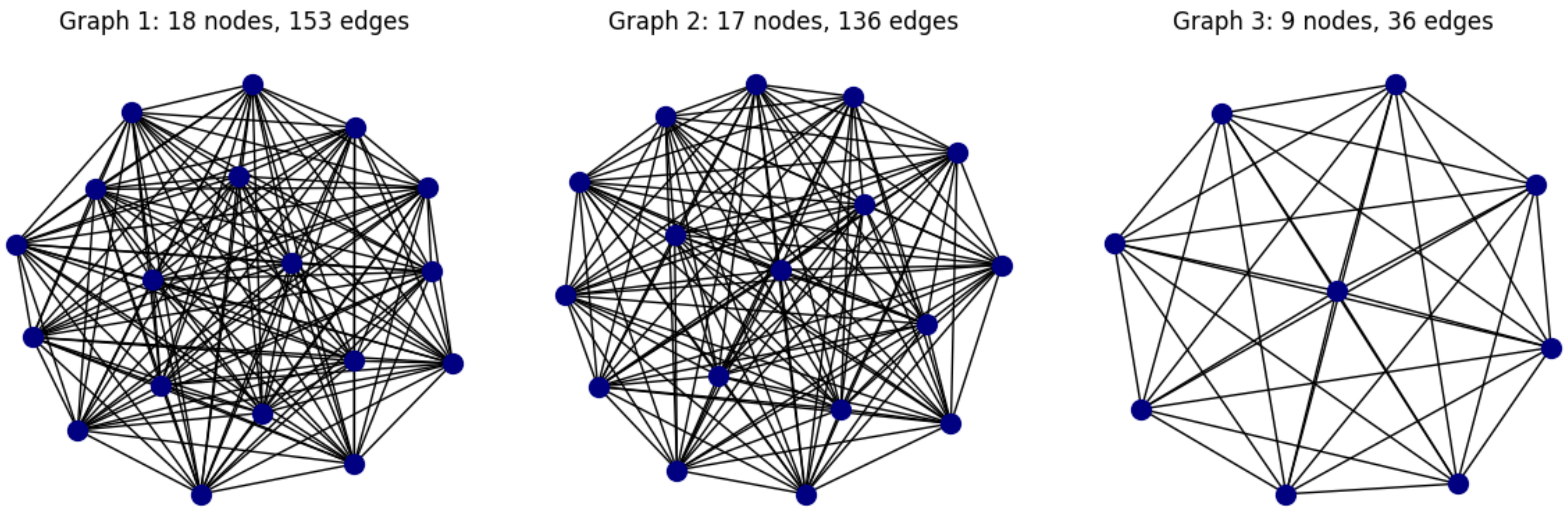
1 graph = 1 jet
1 node = 1 daughter particle

- Fully connected edges
- Variable # of nodes

Features: jet-level and daughter-level kinematics

PyTorch geometric





Node Features

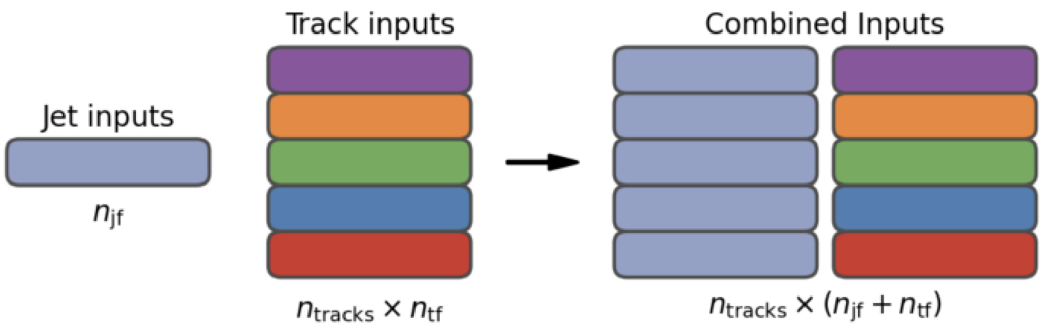
GitHub

Jet Features

- Top-level jet kinematics
- Secondary vertex (SV) tagging variables (LHCb-PAPER-2015-016)

Daughter Features

- Kinematics unique to each daughter in the jet

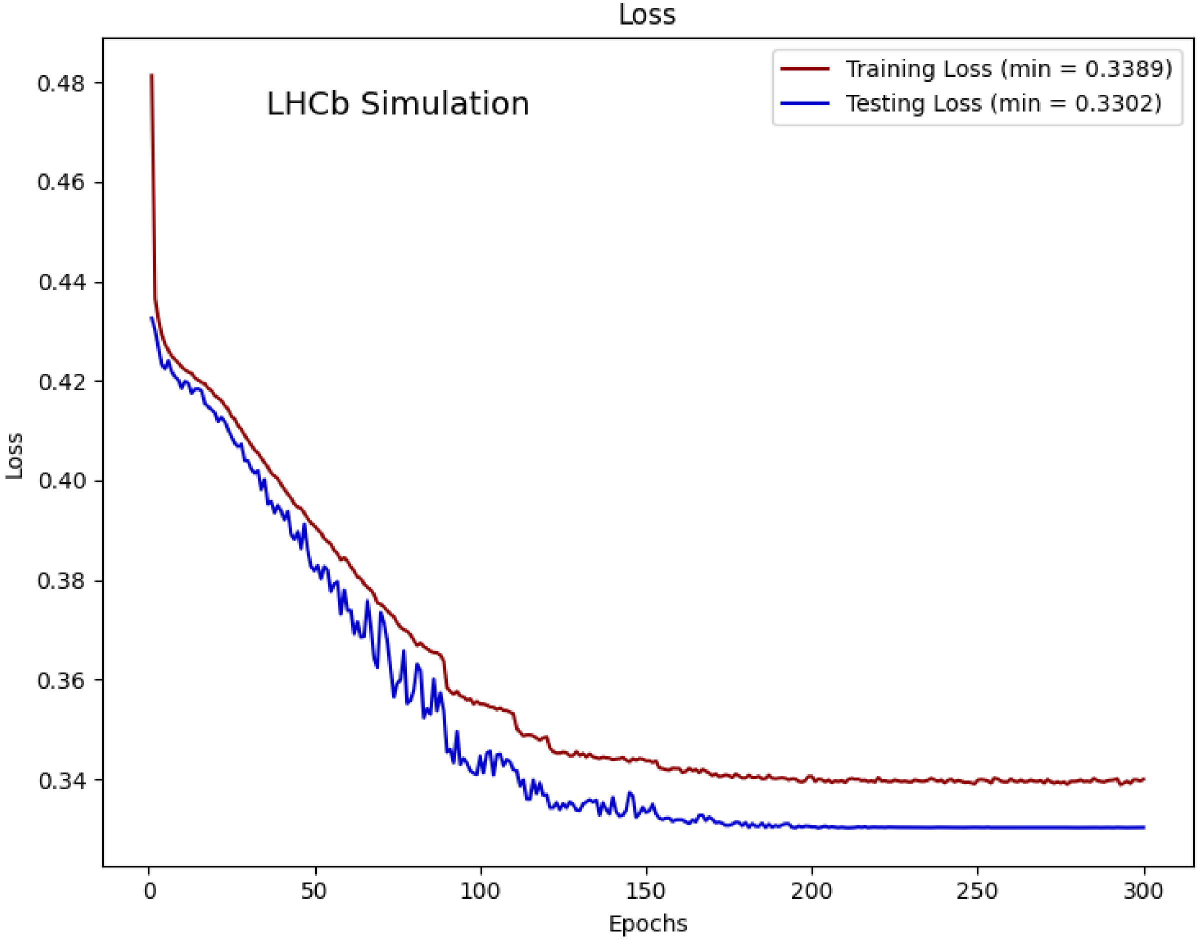


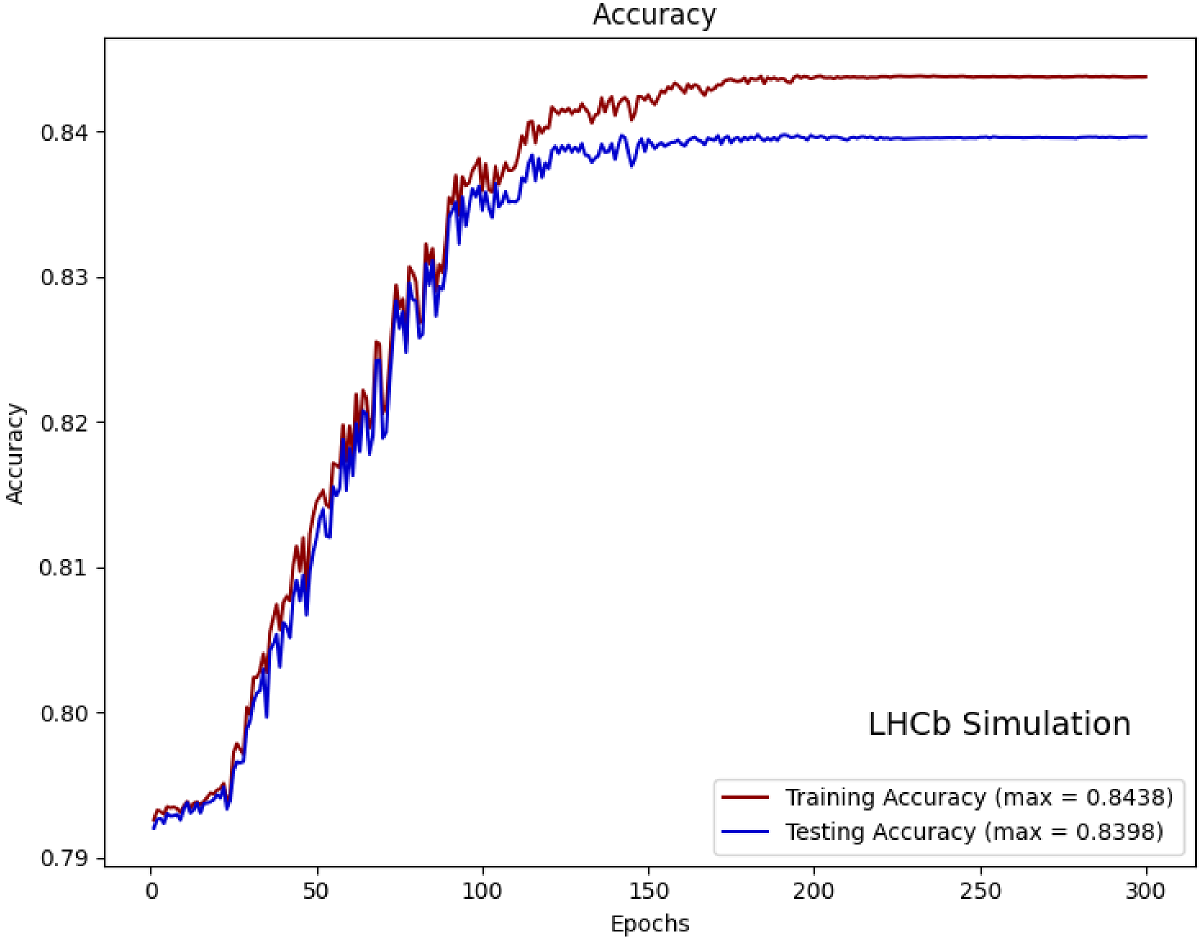
ATL-PHYS-PUB-2022-027

Jet Features	Daughter Features
Pseudorapidity (η)	Energy
Transverse Momentum (pT)	Transverse Momentum
	Particle ID
	Momentum (p)
SV Tagging	Pseudorapidity (η)
Min Radial Flight Distance	Azimuthal Angle (ϕ)
Jet Transverse Momentum (pT)	Charge (Q)
N Tracks	Interaction Point (IP)
N Jet Tracks	NN PID
ΔR	χ^2
Sum of Charges	Charge/Momentum (Q/p)
Mass	Track Direction
Corrected Mass	Vertex Track Direction
Flight Distance χ^2	Calorimetry Energy
Interaction Point χ^2 Sum	
Lifetime (τ)	
Longitudinal Position (z)	
Transverse Momentum (pT)	

b vs q – Training

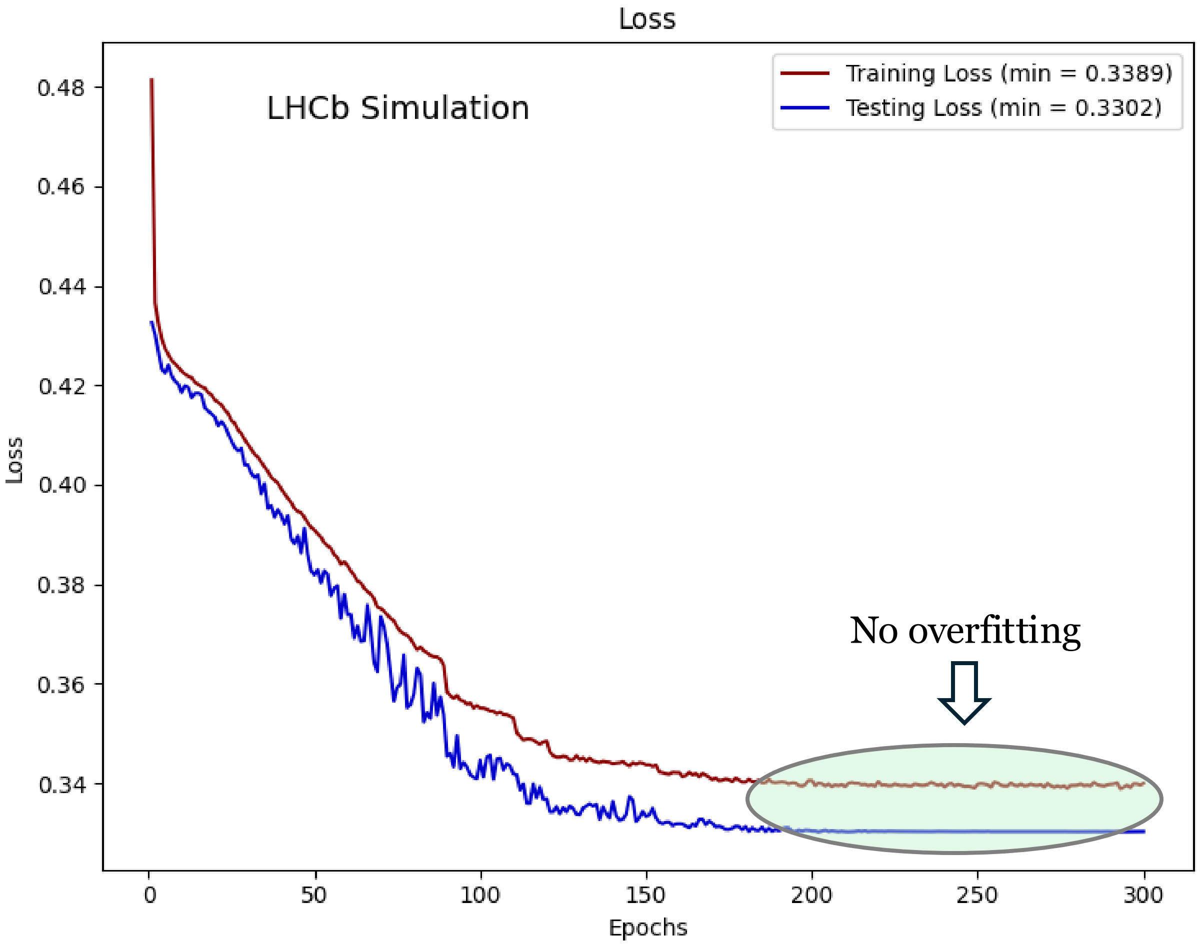
~ 5h
NVIDIA GeForce RTX2080Ti GPU

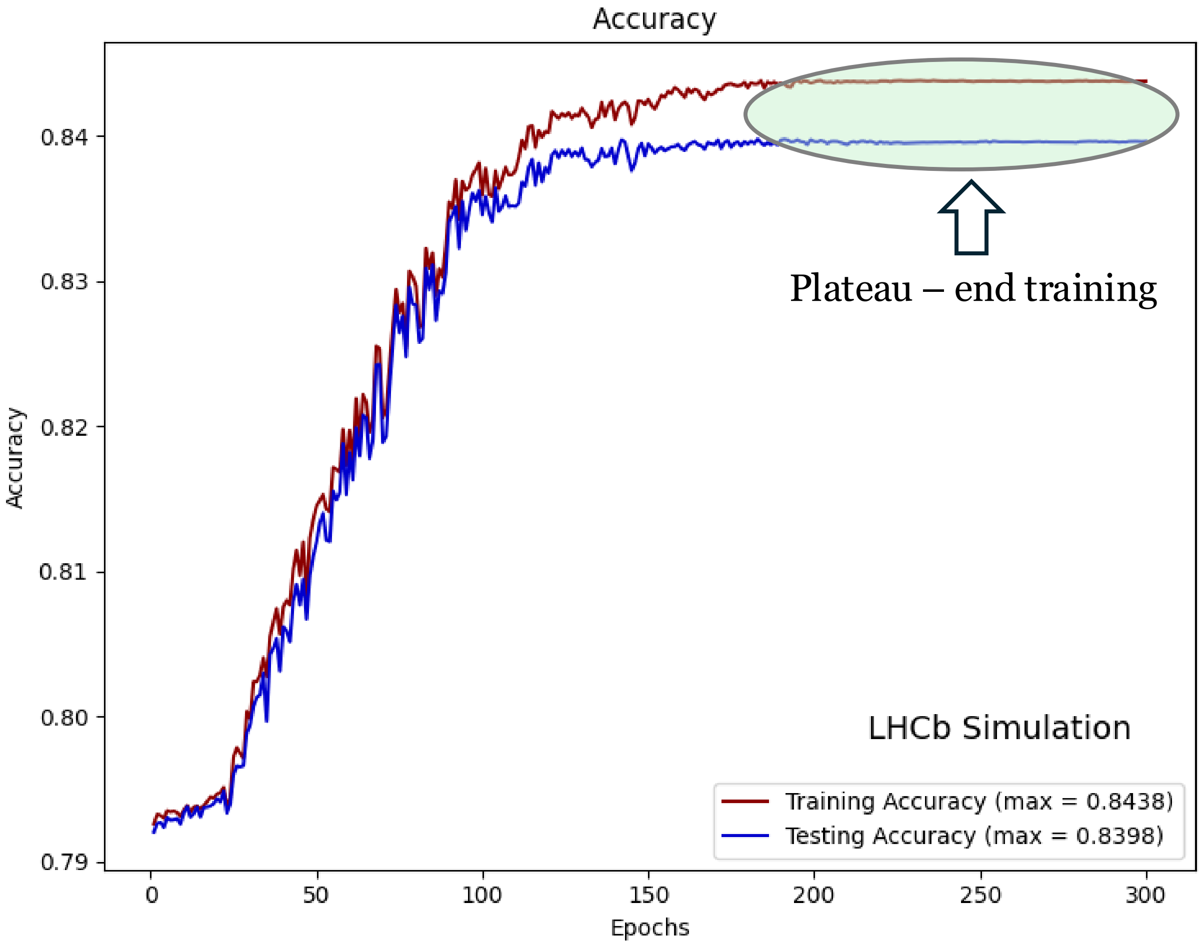




b vs q – Training

~ 5h
NVIDIA GeForce RTX2080Ti GPU

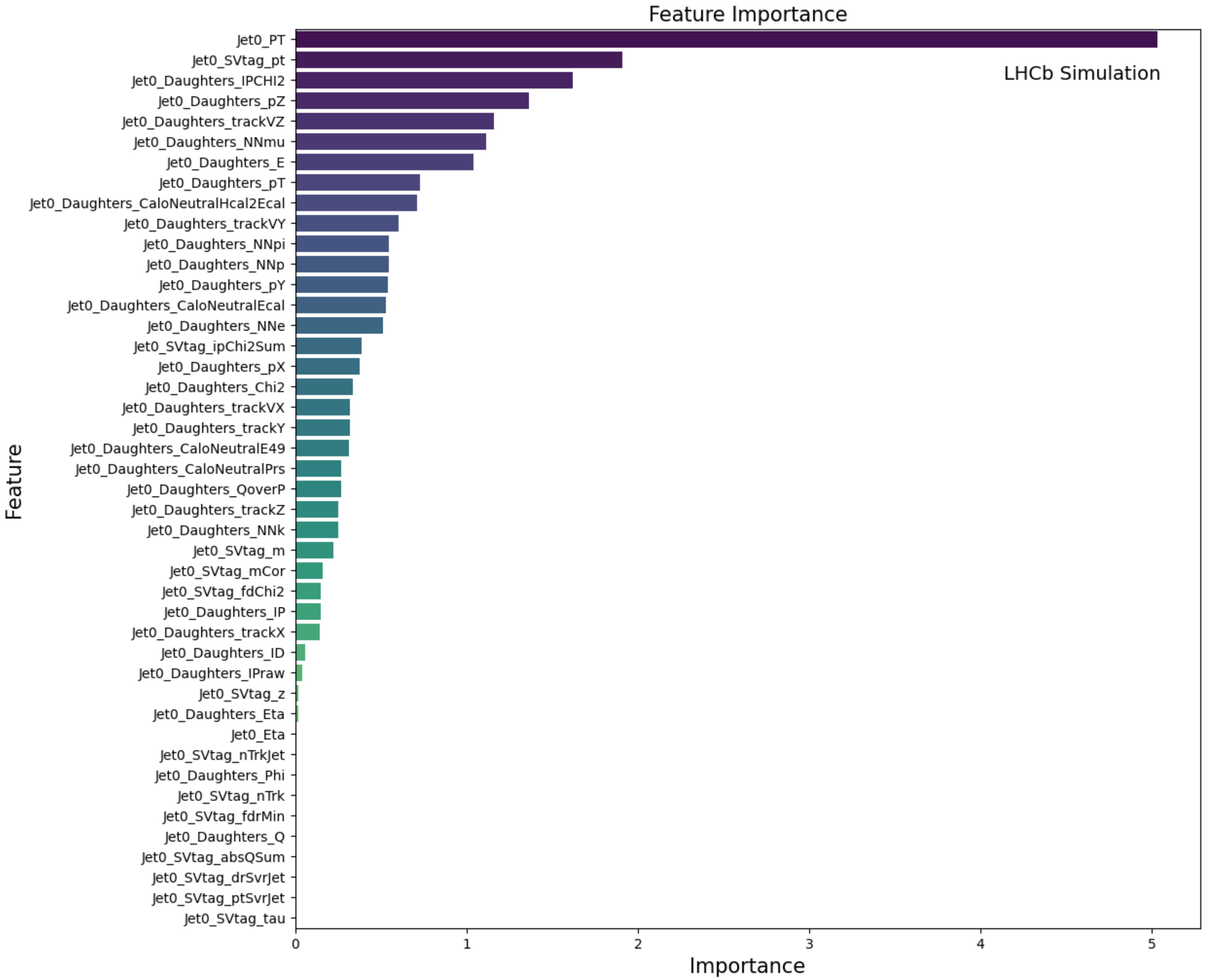




b vs *q* - Feature Importance

Feature Ablation – remove one feature at a time and compare predictions

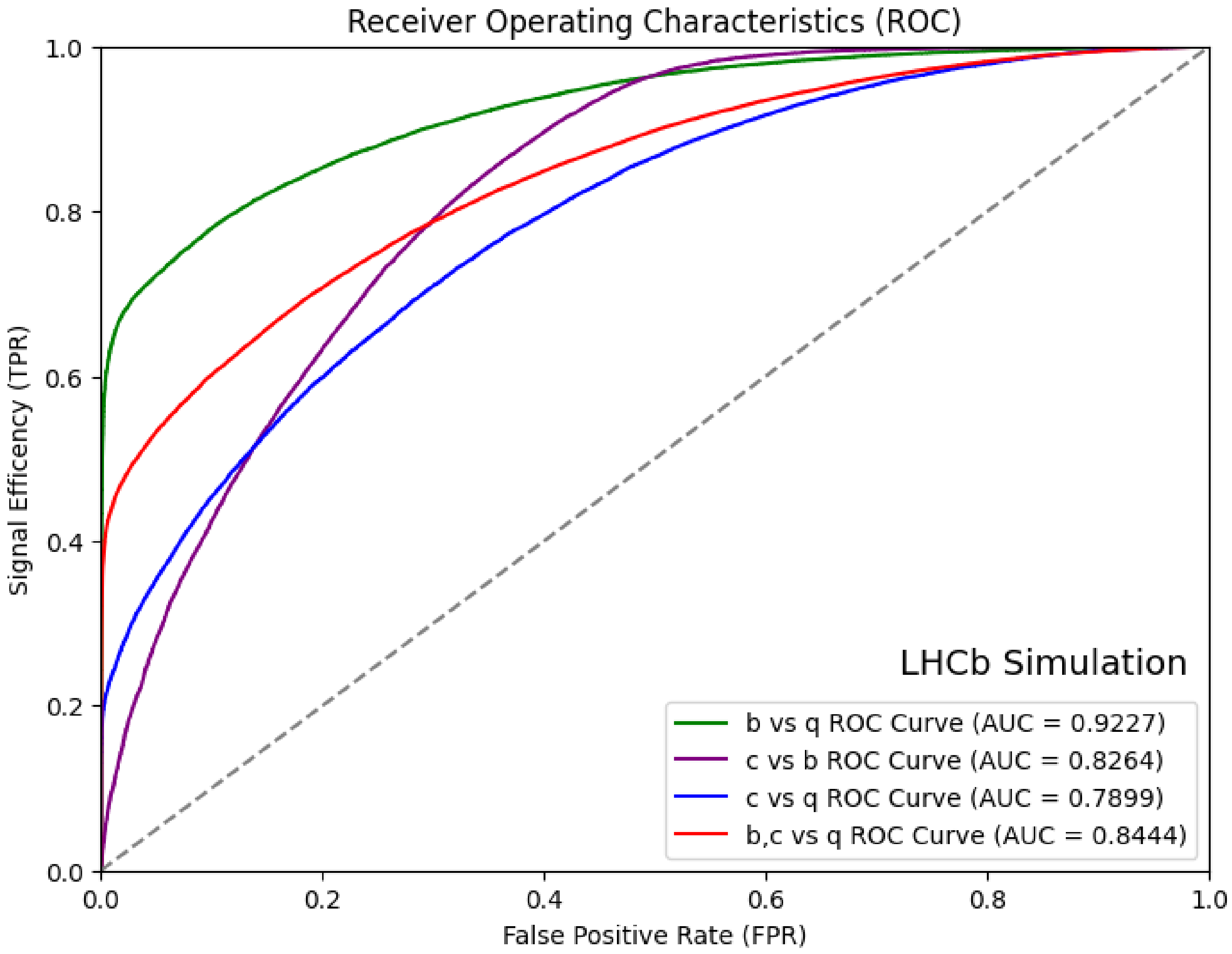
Feature	Importance
Jet0_PT	5.030720
Jet0_SVtag_pt	1.909004
Jet0_Daughters_IPCHI2	1.620610
Jet0_Daughters_pZ	1.366060
Jet0_Daughters_trackVZ	1.160426
Jet0_Daughters_NNmu	1.113793
Jet0_Daughters_E	1.038598
Jet0_Daughters_pT	0.725996
Jet0_Daughters_CaloNeutralHcal2Ecal	0.710647
Jet0_Daughters_trackVY	0.601507
Jet0_Daughters_NNpi	0.547010
Jet0_Daughters_NNp	0.544873
Jet0_Daughters_pY	0.537589
Jet0_Daughters_CaloNeutralEcal	0.526994
Jet0_Daughters_NNe	0.510969
Jet0_SVtag_ipChi2Sum	0.384352
Jet0_Daughters_pX	0.378024
Jet0_Daughters_Chi2	0.336792
Jet0_Daughters_trackVX	0.317194
Jet0_Daughters_trackY	0.317135
Jet0_Daughters_CaloNeutralE49	0.311731
Jet0_Daughters_CaloNeutralPrs	0.269899
Jet0_Daughters_QoverP	0.265553



PID Information

Particle ID	b Counts	q Counts	Ratio (b/q)
$\mu^\pm$	77381	17184	4.50
K_s^0	257905	141741	1.82
$\Lambda^\pm$	74696	43770	1.71
$e^\pm$	207242	203155	1.02
γ	3538058	3833703	0.92
$\pi^\pm$	3171630	3582468	0.89
π^0	92794	107348	0.86
$K^\pm$	368460	430964	0.86
$p^\pm$	139609	251848	0.55

Performance Results



b **Efficiency**	q **FPR**
0.85	0.1922
0.8	0.1212
0.75	0.0723
0.7	0.0332
0.65	0.0118
0.6	0.0046

c **Efficiency**	b **FPR**
0.40	0.0909
0.30	0.0560
0.20	0.0296

Conclusions

Summary

- First GNN jet tagger for LHCb
- Training strengthened by PID information
- Broader application without SV dependence

Further Steps

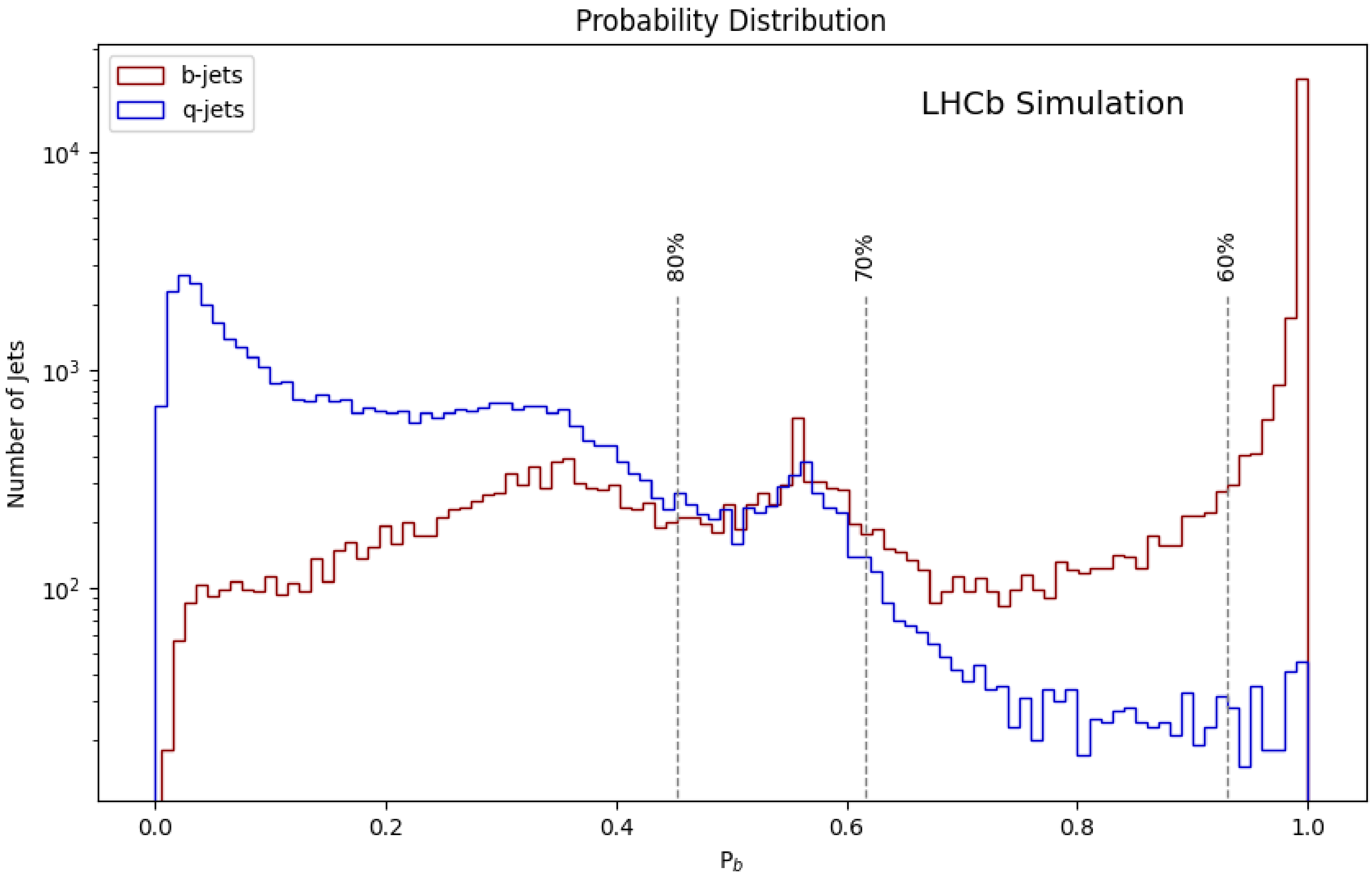
- Expand for fat jets + HF-jets inside
- Migrate to Run 3 samples and retrain
- Integrate into Moore/ HLT2 HF jet selections

Physics Applications

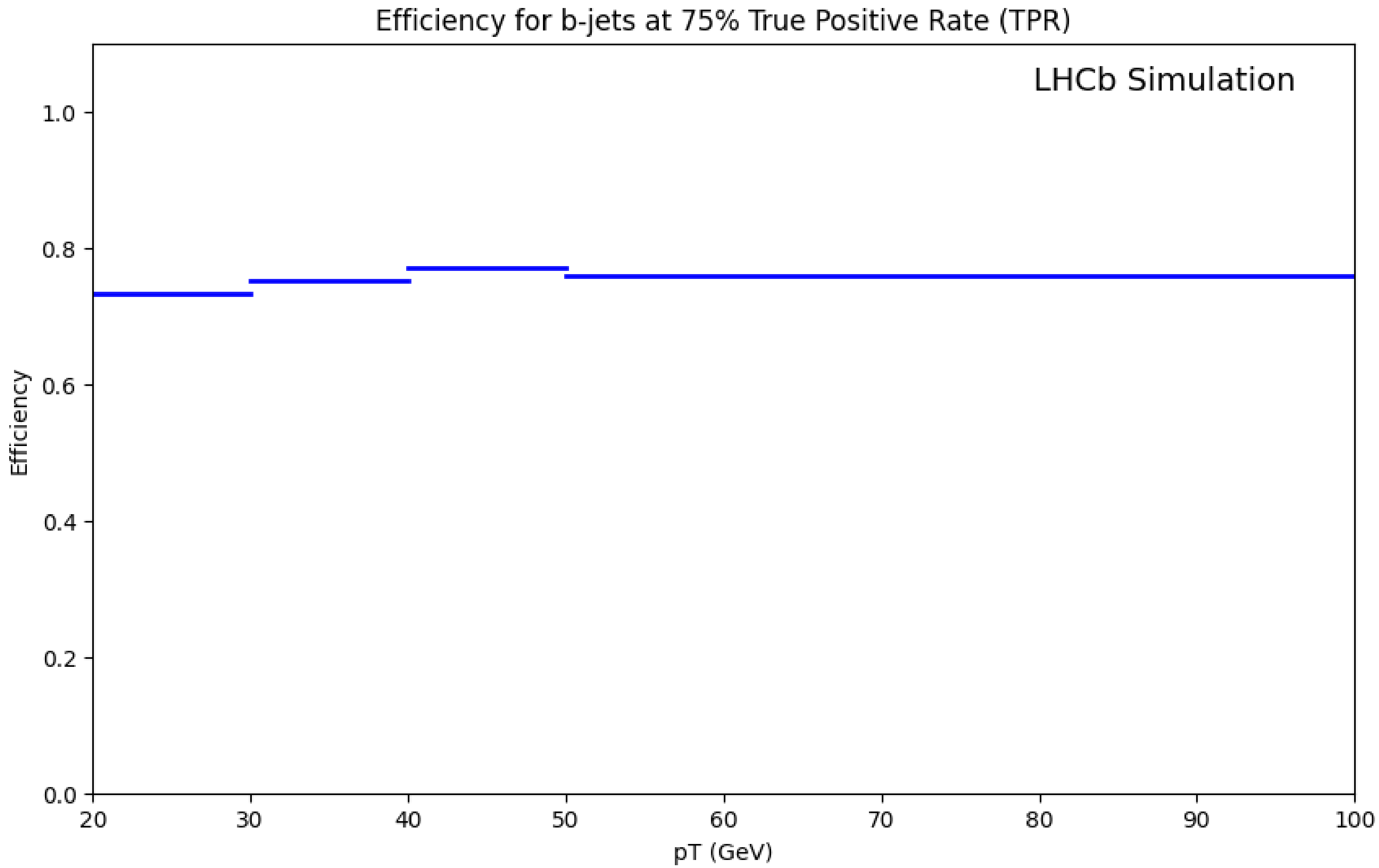
Veto	Selection
WW	$h \to b\bar{b}$
WZ	$h \to c\bar{c}$
ZZ	top quark measurements
Z + jets	Z + jets
$h \to WW$	$W \to cX$ analysis

Backup

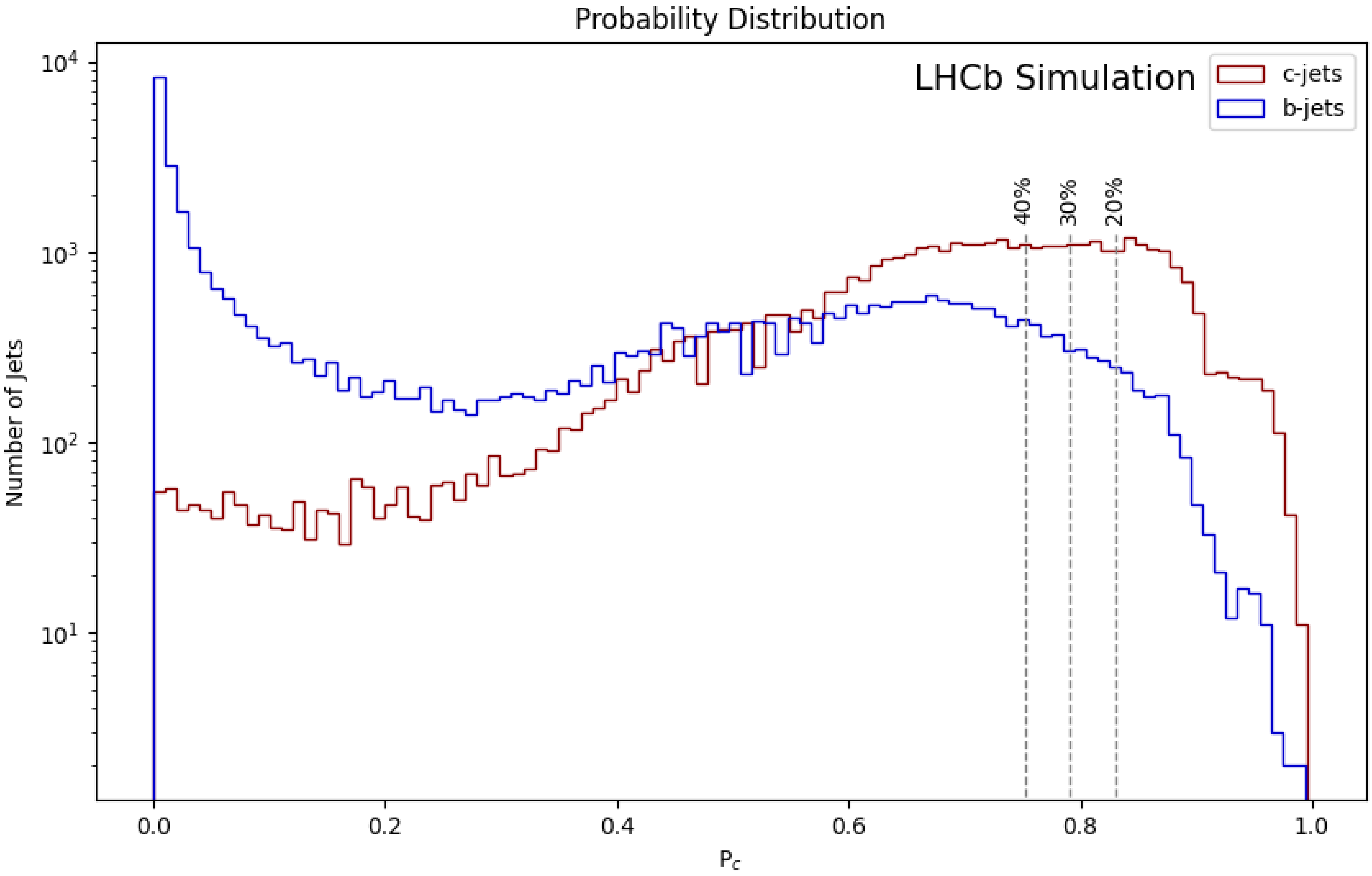
Classifier Application (b vs q) - P_b



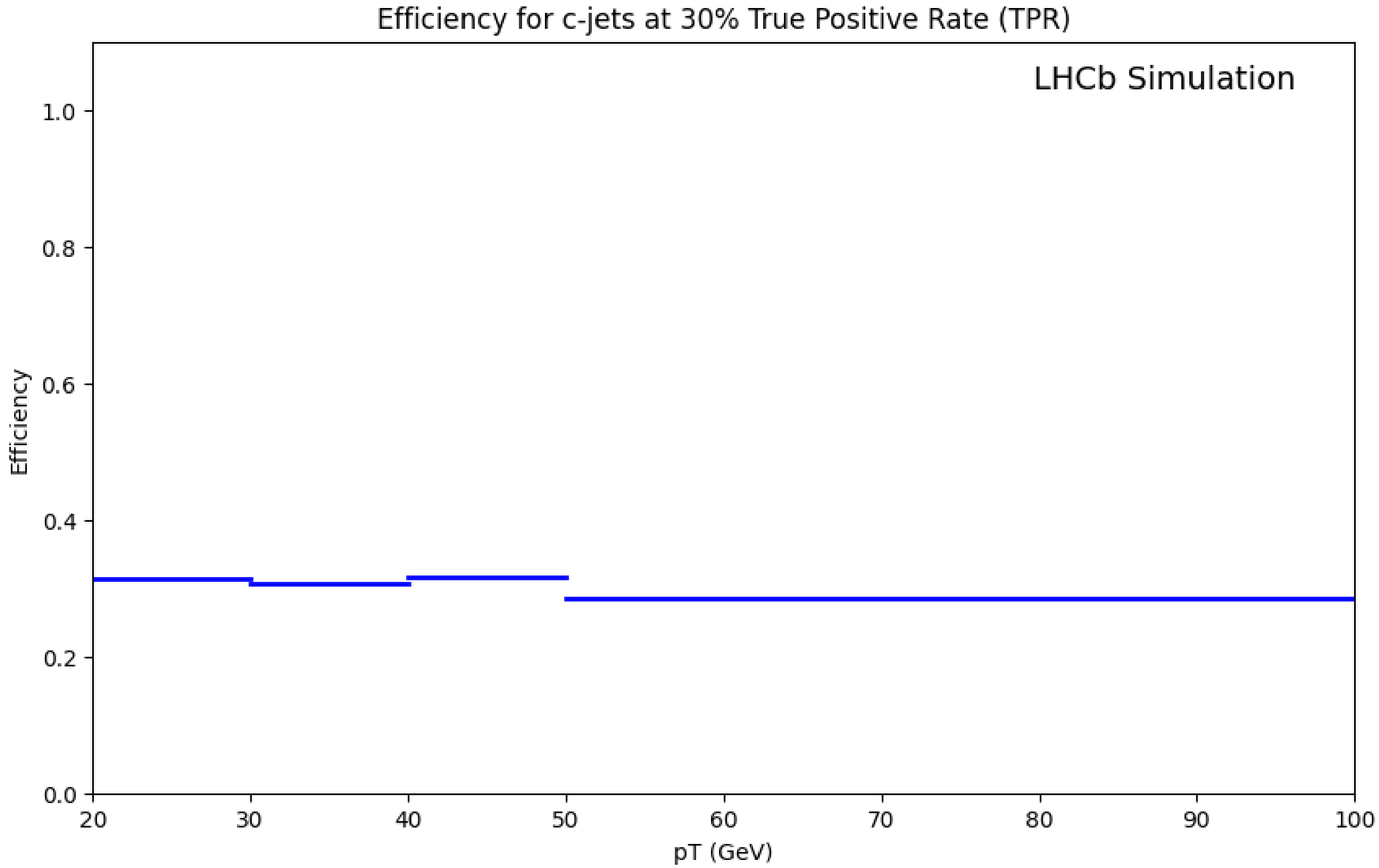
Classifier Application (*b* vs *q*)



Classifier Application (c vs b) - P_c



Classifier Application (*c* vs *b*)



Data Preparation

Dataset

- 1.2M fully reconstructed di-jet events per flavour
  - Leading jet only
  - 80:20 training and validation split

Truth Matching

- Reco jet matched to truth jet
- Energy fraction of daughters used for flavour selection

Classifiers

- b vs q
- c vs q
- c vs b
- b/c vs q

General Selection Requirements
$p_T > 20$ GeV $2.2 < \eta < 4.4$

Truth Matching		
b **Selection**	c **Selection**	q **Selection**
MC Match = 1 MC Jet EfD > 0.6	MC Match = 1 MC Jet EfB > 0.4	MC Match = 1 MC Jet EfB < 0.2 MC Jet EfD < 0.2

PyTorch Geometric Layers

SAGEConv

- Aggregates information from neighbors – mean
- $\boldsymbol{x}_i' = \boldsymbol{W}_1 \boldsymbol{x}_1 + \boldsymbol{W}_2 \cdot \mathrm{mean}_{j \in \mathcal{N}(i)} \boldsymbol{x}_j$

LayerNorm

- Normalize inputs across all features independently
- $y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} * \gamma + \beta$

ReLU

- Introduces non-linearity
- $R(z) = \max(0, z)$

Dropout

- Zero elements with probability, p
- Scale by factor of $\frac{1}{1-p}$

Global Add Pooling

- After convolutional layers, add outputs
- $\boldsymbol{r}_i = \sum_{n=1}^{N_i} \boldsymbol{x}_n$

Linear

- Reduce dimensionality of outputs
- $y = xA^T + b$

Binary Cross Entropy Loss (with sigmoid layer)

- Computes difference between prediction and truth labels
- $\ell(x, y) = L = \{l_1, \dots, l_N\}^\top, l_n = -w_n[y_n \cdot \log \sigma(x_n) + (1 - y_n) \cdot \log(1 - \sigma(x_n))$

AdamW Optimizer

- Minimizes loss function – stochastic gradient descent
- Separates weight decay from gradients