Jet Flavour Classification with Graph Neural Networks

Gabriella Pesticci | Kenyon College

Dr. Conor Henderson, Dr. Nate Grieser | University of Cincinnati

August 7, 2024

Motivation

Jet Hadronization

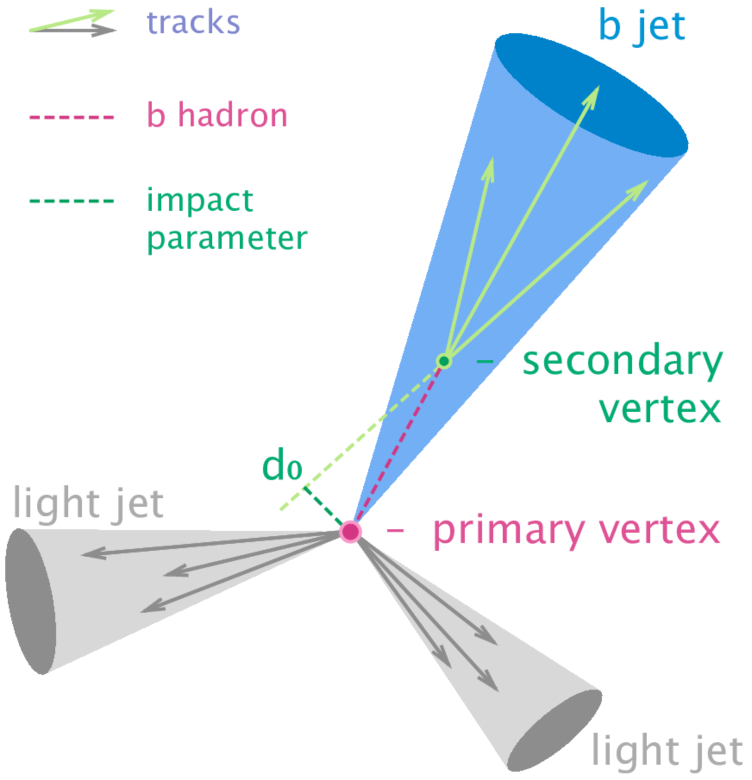
- p-p collisions at LHC produce quarks and gluons which hadronize due to QCD confinement
- The long lifetime of *b*-jets creates a characteristic secondary vertex
- We need to reconstruct the event and identify the original particle

Goal:

- Identify if jets are produced by *b*-quarks or not
- Utilize deep learning to accurately classify the jets in our data

Application:

- *b*-jet selection: h → bb, top quarks, Z + jets, …
- *b*-jet veto: WW, Z + jets, Higgs decays, …

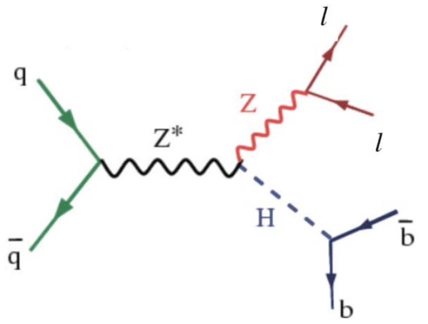


3-jet event

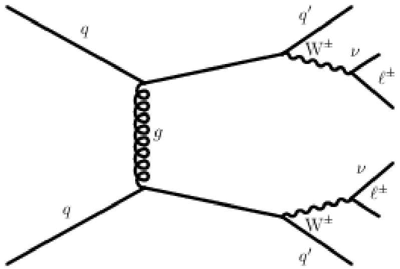
Selection

Veto

h → bb

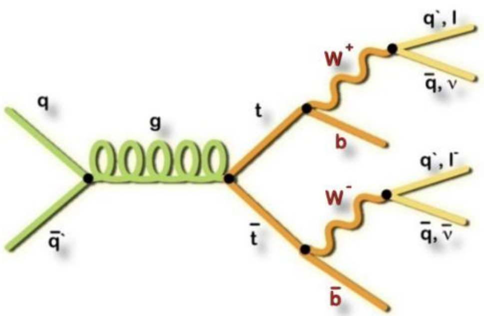


ATLAS Higgs Hunting

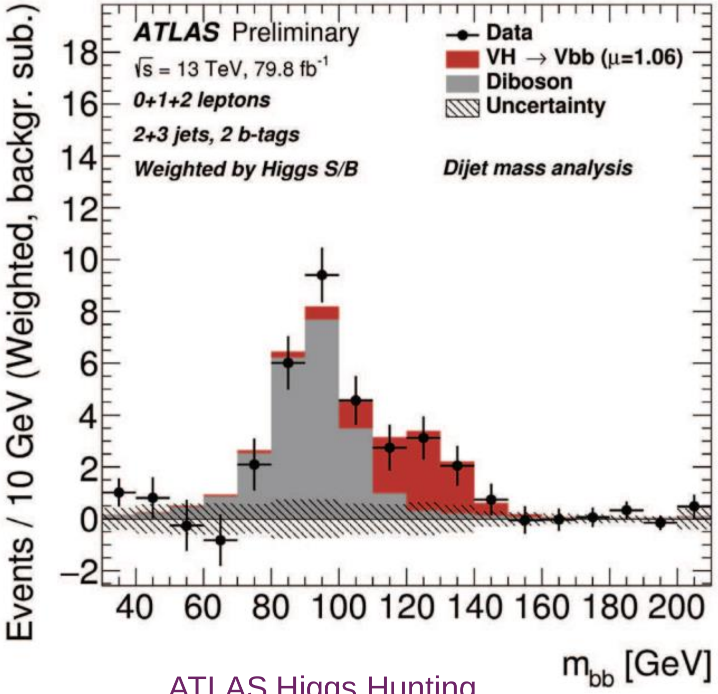


CMS-SMP-19-012

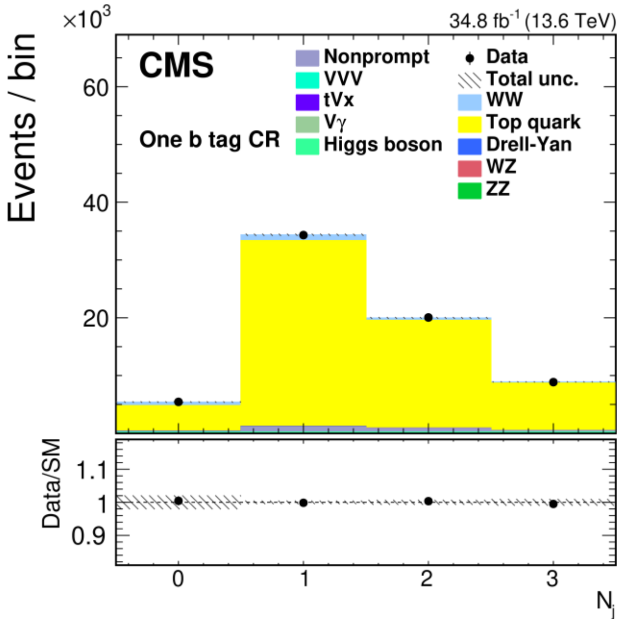
WW analysis



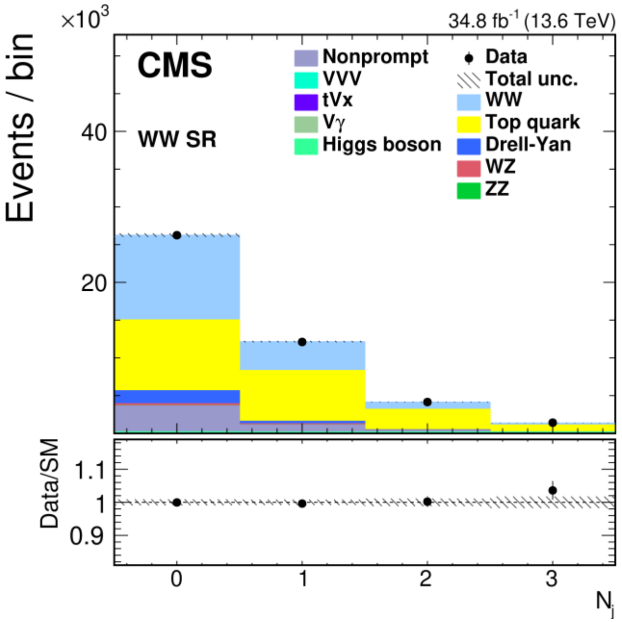
ATLAS b-tagging



ATLAS Higgs Hunting



$$N_{b-jet} \geq 1$$

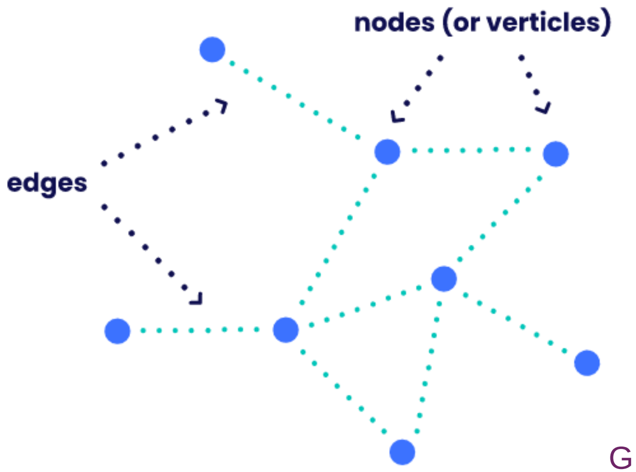


$$N_{b-jet} = 0$$

CMS-SMP-24-001

Graph Neural Networks (GNNs)

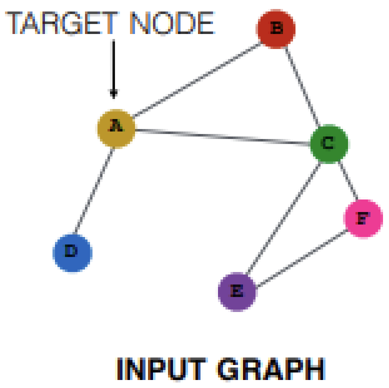
GitHub

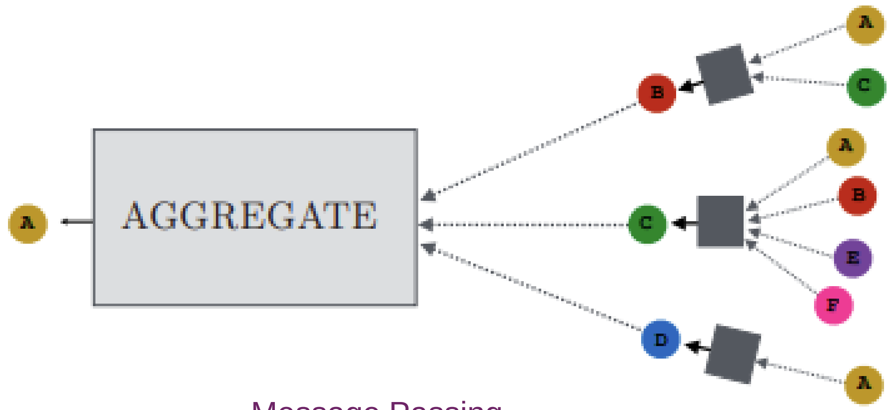


GNN

- **GNN Advantages**
  - Variable number of nodes and edges
  - Able to capture complex relationships and represent our system

- **Message Passing**
  - Node information aggregated from neighboring nodes
  - Target node is updated
  - Learn features of neighbors

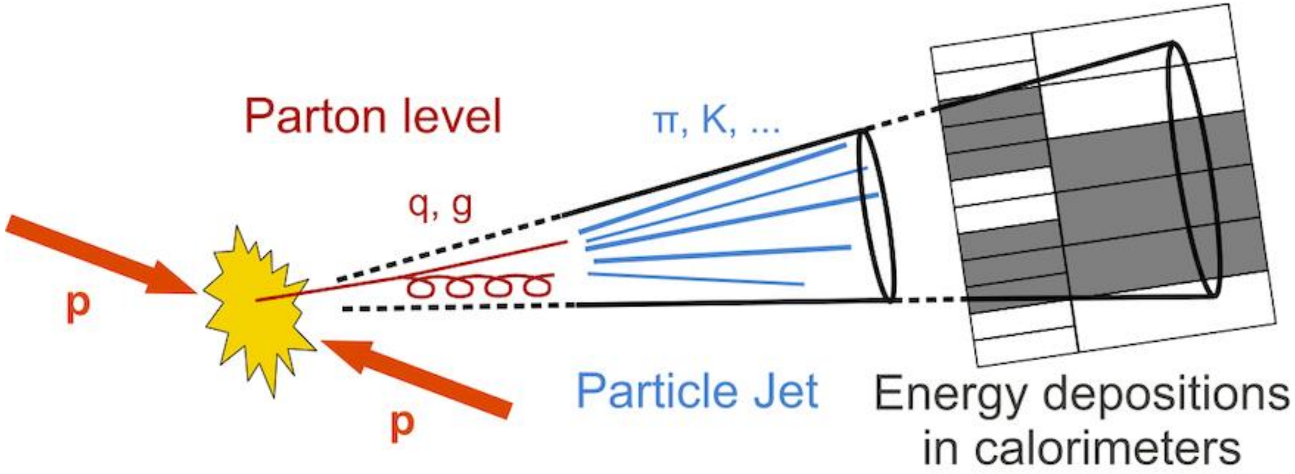




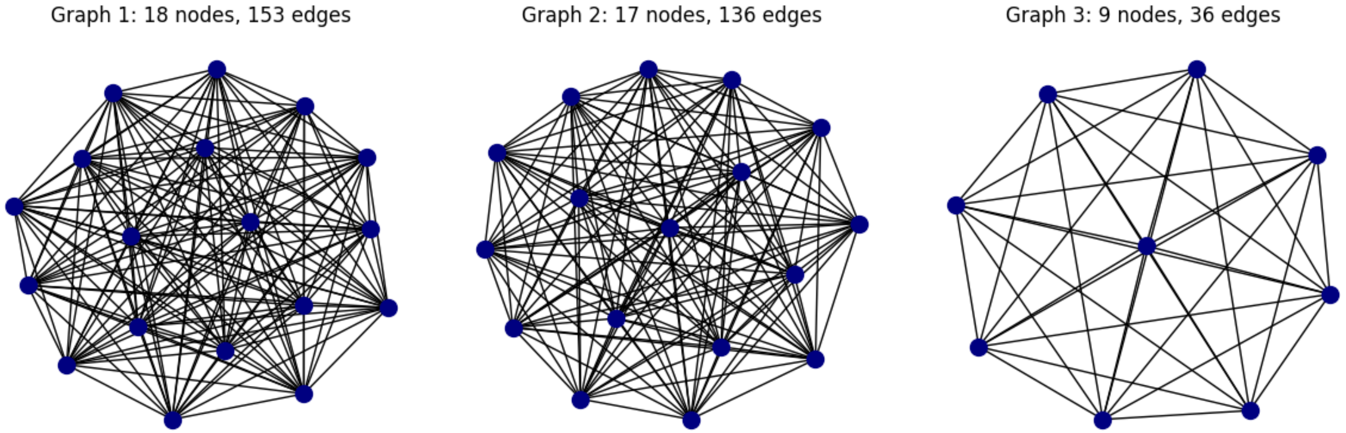
Message Passing

Graph Construction

- 1 graph = 1 jet
- 1 node = 1 daughter
- Fully connected edges
- Nodes vary between graphs
- Features: jet-level and daughter-level kinematics



ATLAS Jets



Data Preparation

Dataset

- 600k fully reconstructed events
- Dijet simulation, leading jet
- 80:20 training and testing split

Truth Matching

- Monte Carlo simulations to select signal (*b*-jets) and background (not *b*-jets)

Jet Selection

- Cuts on pT and η

Truth Matching	
Signal Data	**Background Data**
MC Match = 1 MC Jet EfB > 0.6	MC Match = 1 MC Jet EfB < 0.6 MC Jet EfD < 0.6

Selection Requirements
$20 \text{ GeV} < p_T < 50 \text{ GeV}$ $2.2 < \eta < 4.4$

Node Features

Jet Features

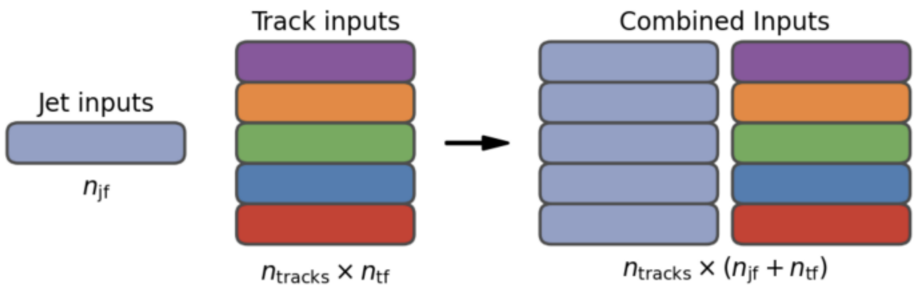
- Top-level jet kinematics
- Duplicated to each node

Daughter Features

- Kinematics unique to each daughter in the jet

SV Features

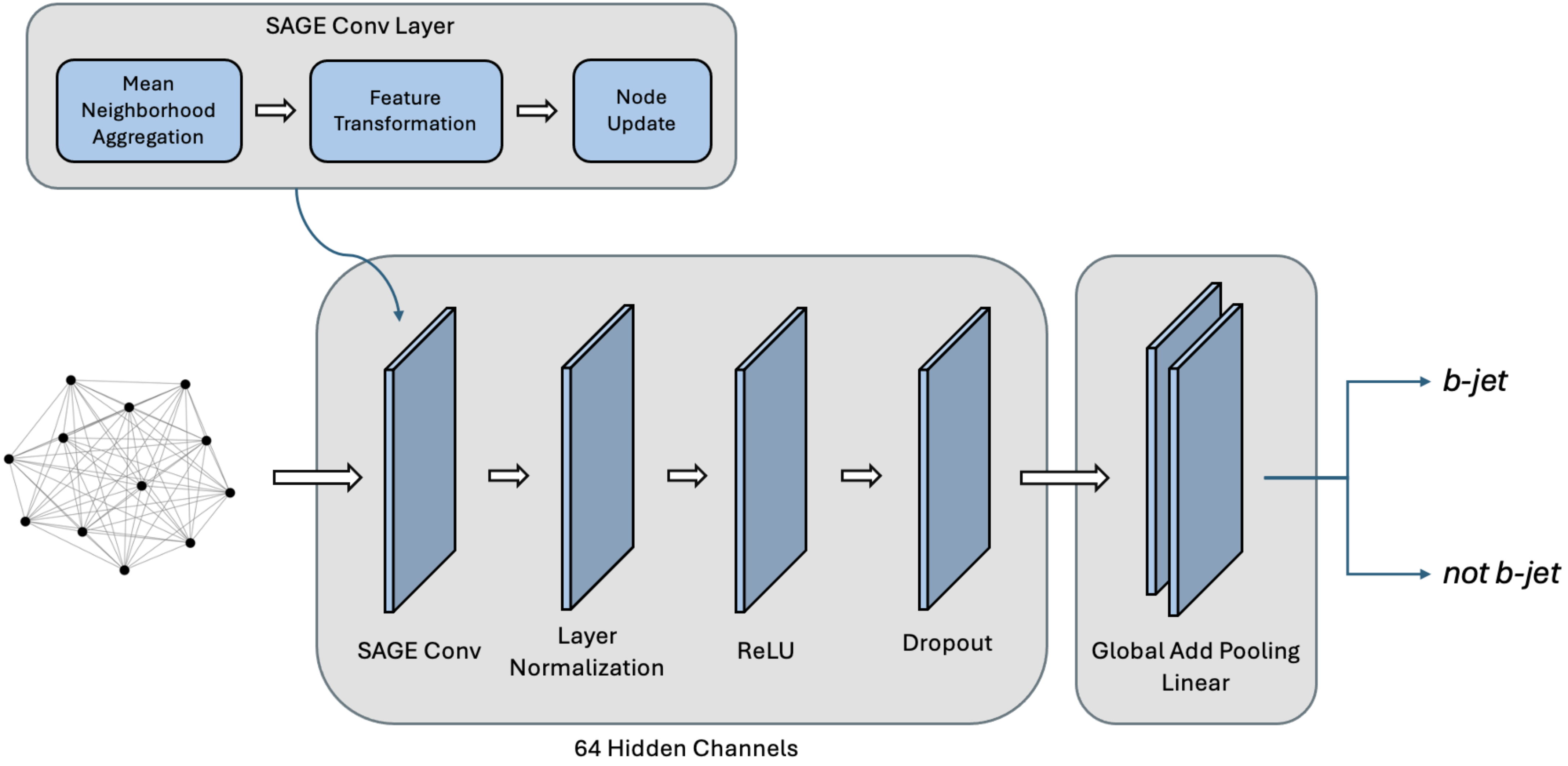
- Secondary vertex tagging variables
- Jet-level – duplicated to each daughter



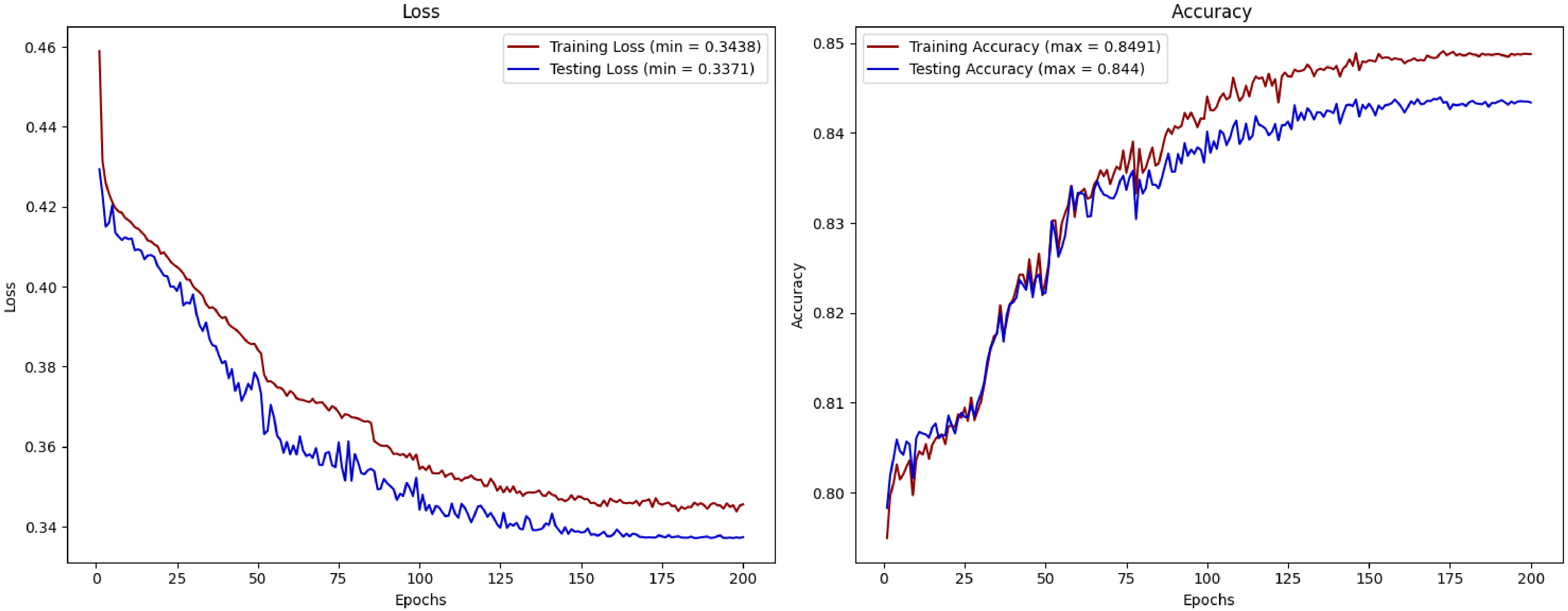
ATL-PHYS-PUB-2022-027

Jet Features	Daughter Features
Eta	E
pT	pT
	ID
	pX
SV Tagging	pY
fdrMin	pZ
ptSvrJet	Eta
nTrk	Phi
nTrkJet	Q
drSvrJet	IP
absQSum	IPCHI2
m	IPraw
mCor	NNe
fdChi2	NNk
ipChi2Sum	NNp
tau	NNpi
z	NNmu
pT	Chi2
	QoverP
	trackX
	trackY
	trackZ
	trackVX
	trackVY
	trackVZ
	CaloNeutralEcal
	CaloNeutralHcal2Ecal
	CaloNeutralE49
	CaloNeutralPrs

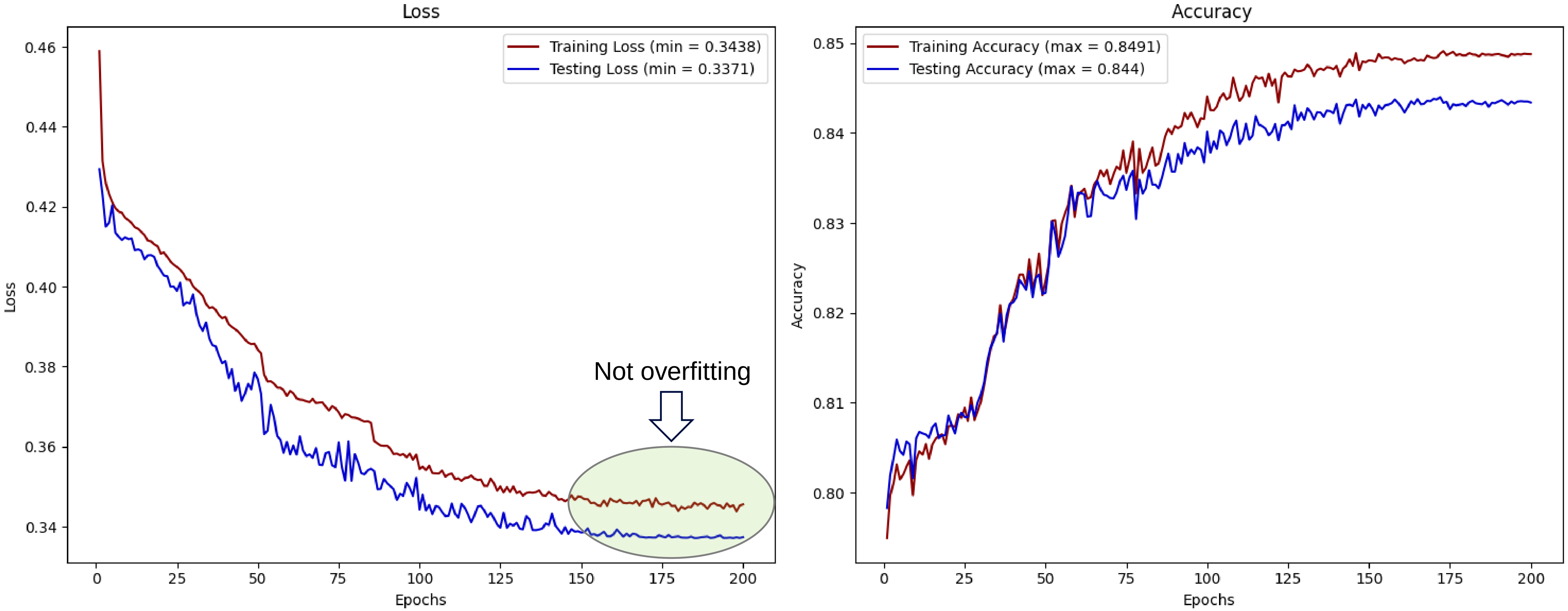
GNN Architecture



Results

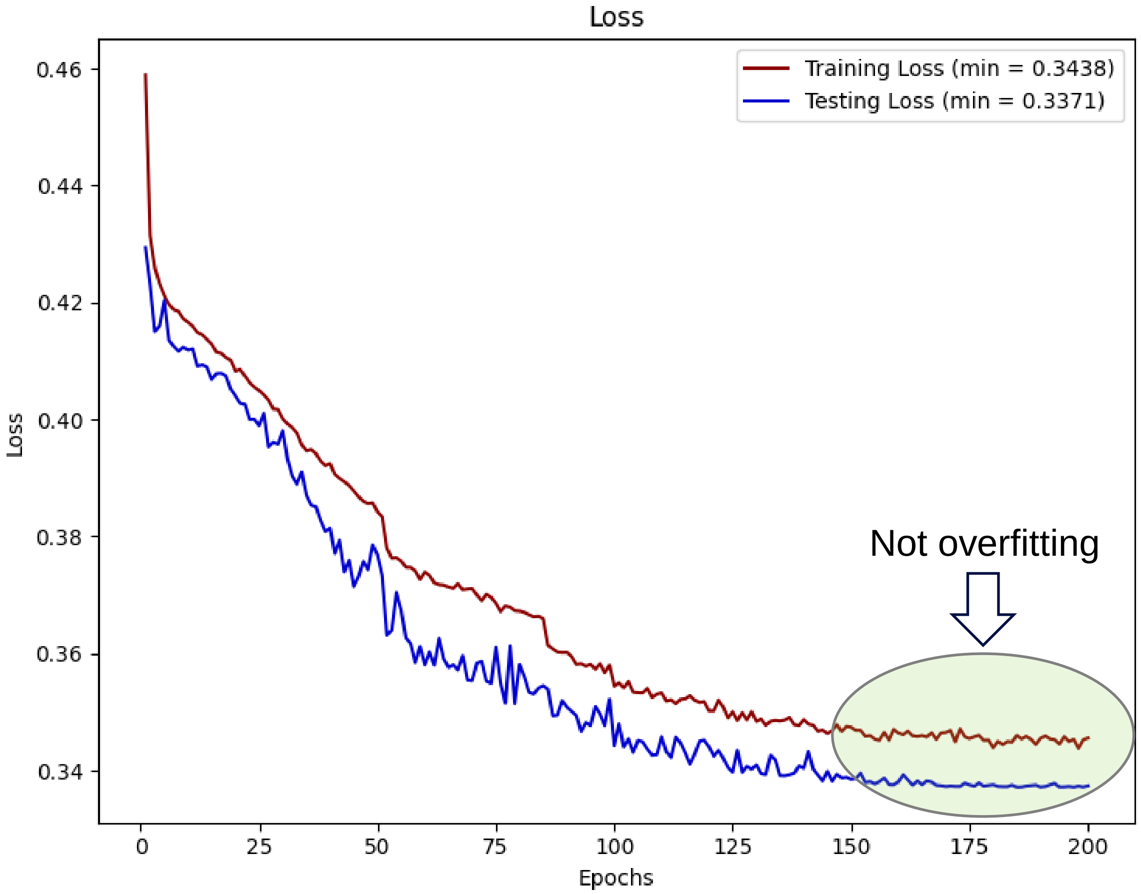

Loss
Training Loss (min = 0.3438)
Testing Loss (min = 0.3371)
Loss
0.46
0.44
0.42
0.40
0.38
0.36
0.34
0
25
50
75
100
125
150
175
200
Epochs
Accuracy
Training Accuracy (max = 0.8491)
Testing Accuracy (max = 0.844)
Accuracy
0.85
0.84
0.83
0.82
0.81
0.80
0
25
50
75
100
125
150
175
200
Epochs

Results

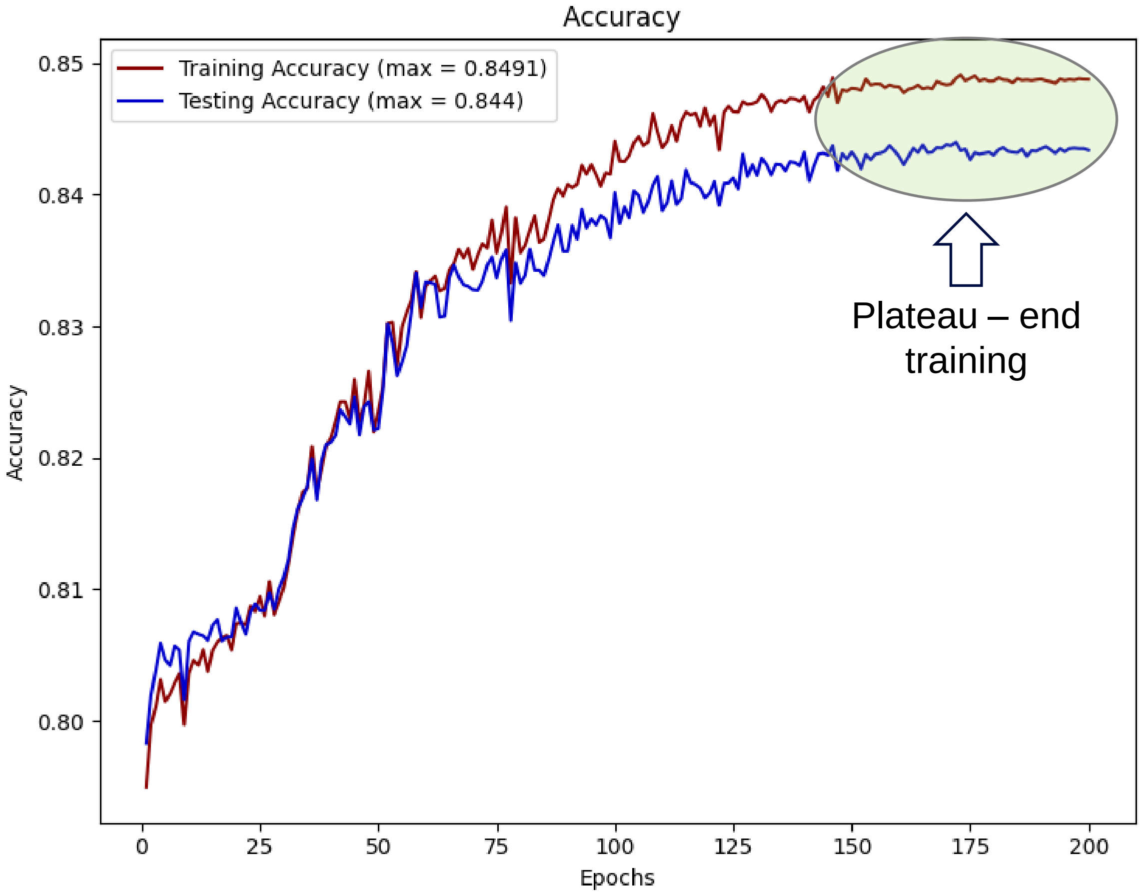


Loss
Training Loss (min = 0.3438)
Testing Loss (min = 0.3371)
Not overfitting
Loss
Epochs
Accuracy
Training Accuracy (max = 0.8491)
Testing Accuracy (max = 0.844)
Accuracy
Epochs

Results

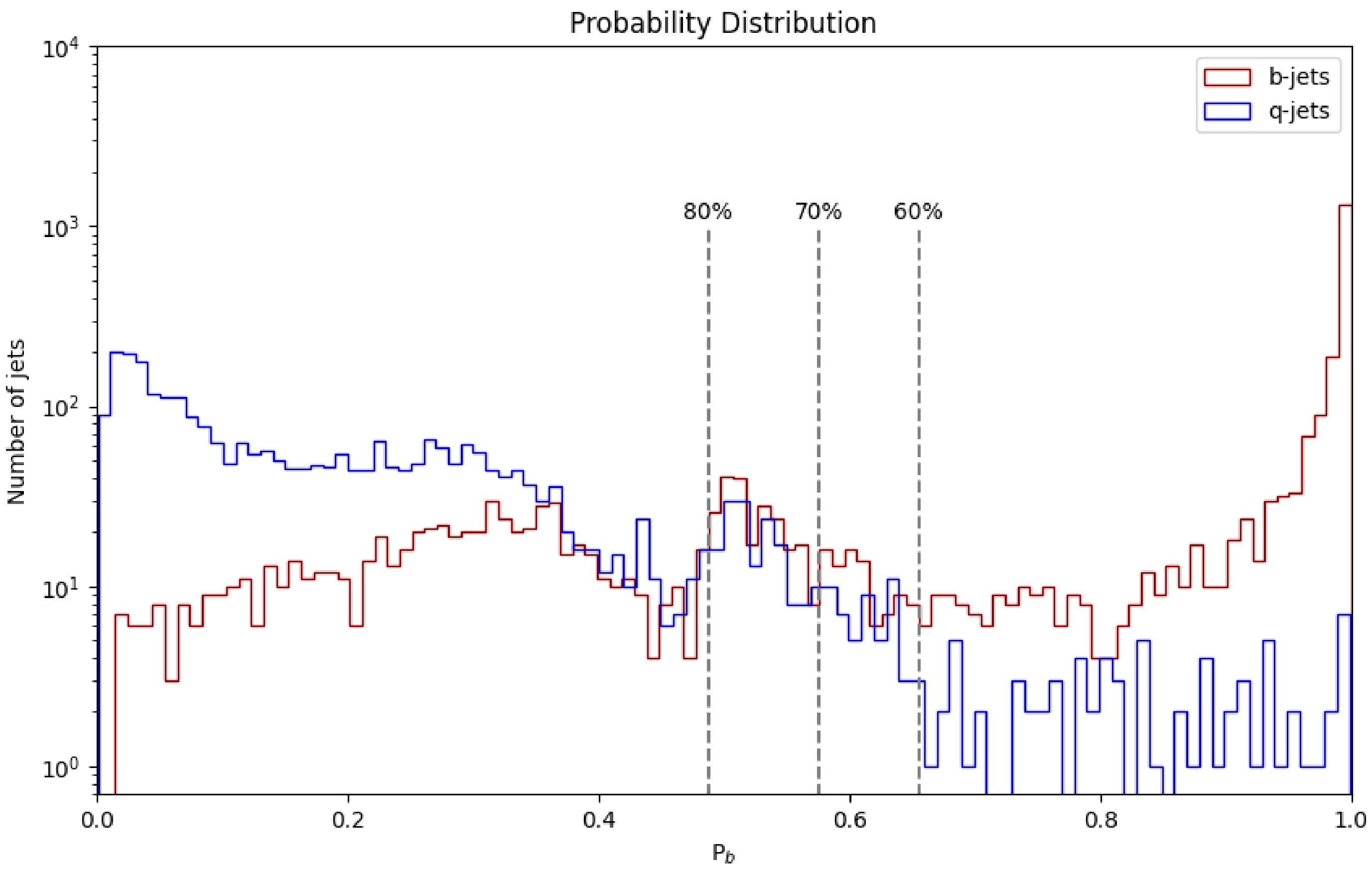


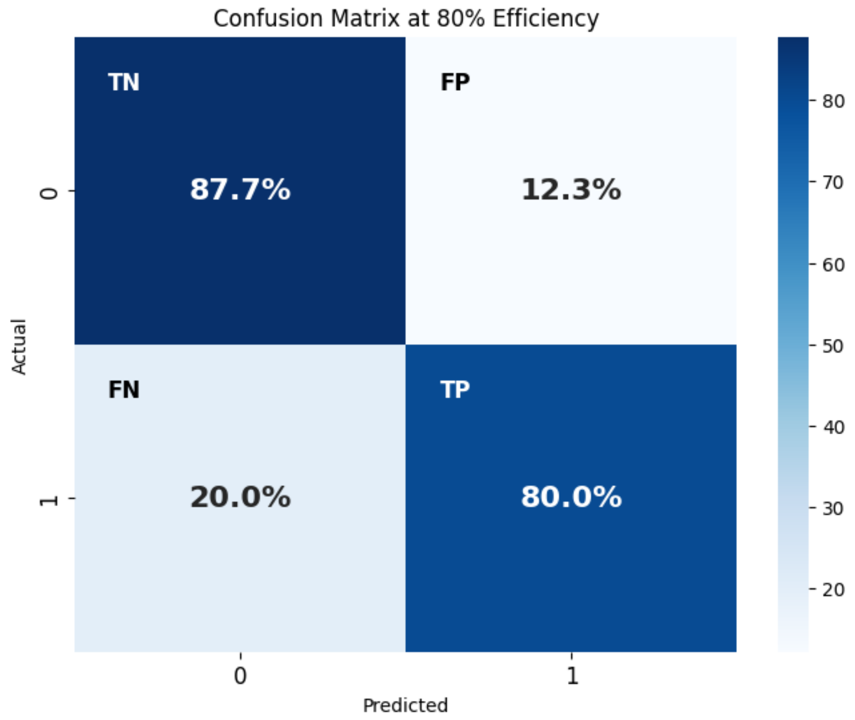
Loss
Training Loss (min = 0.3438)
Testing Loss (min = 0.3371)
Loss
Epochs
Not overfitting



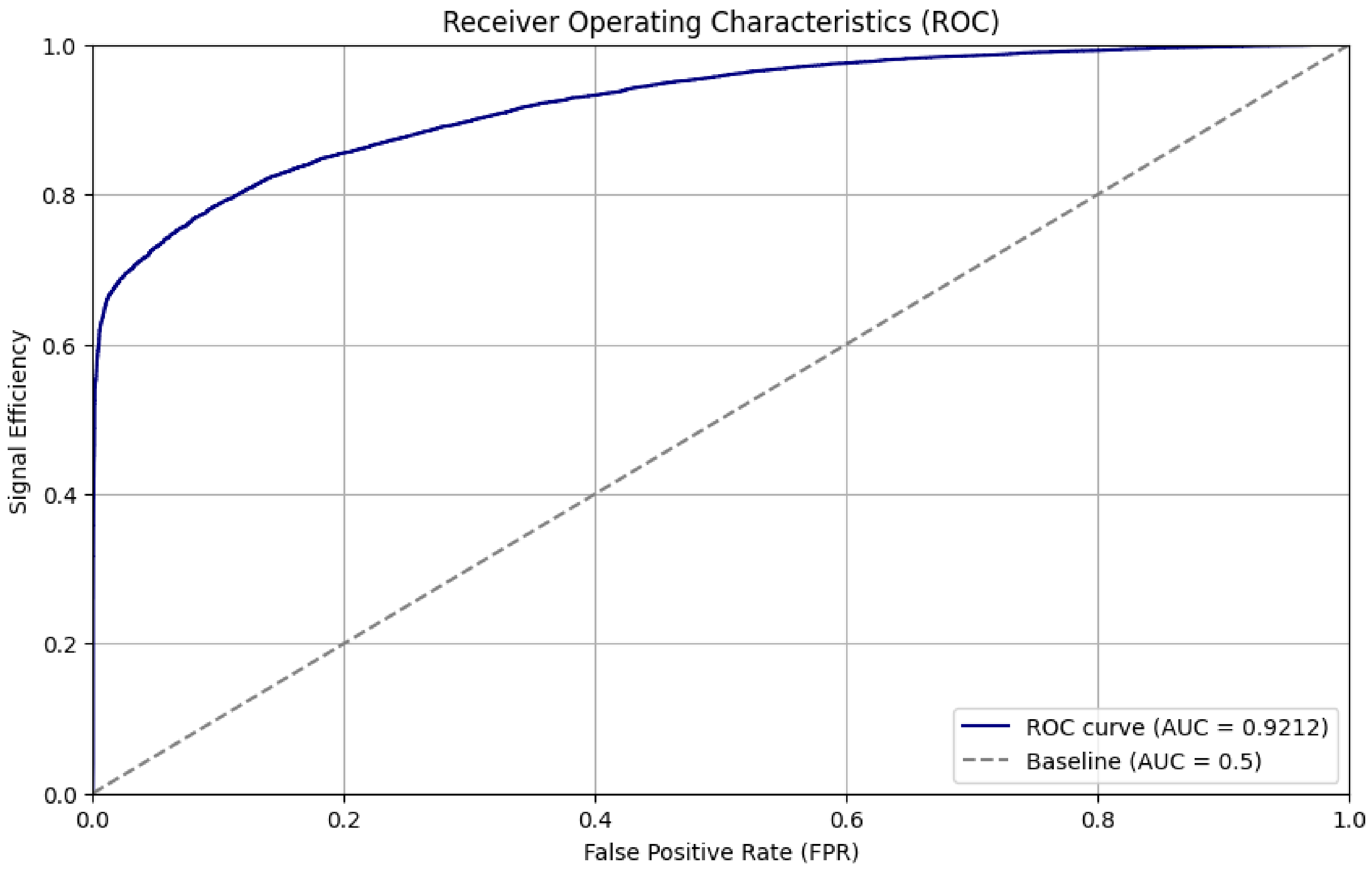
Accuracy
Training Accuracy (max = 0.8491)
Testing Accuracy (max = 0.844)
Accuracy
Epochs
Plateau – end training

Results


Probability Distribution
b-jets
q-jets
80%
70%
60%
Number of jets
10^0
10^1
10^2
10^3
10^4
0.0
0.2
0.4
0.6
0.8
1.0
P_b


Confusion Matrix at 80% Efficiency
TN
87.7%
FP
12.3%
FN
20.0%
TP
80.0%
Actual
0
1
Predicted
0
1
80
70
60
50
40
30
20

Results



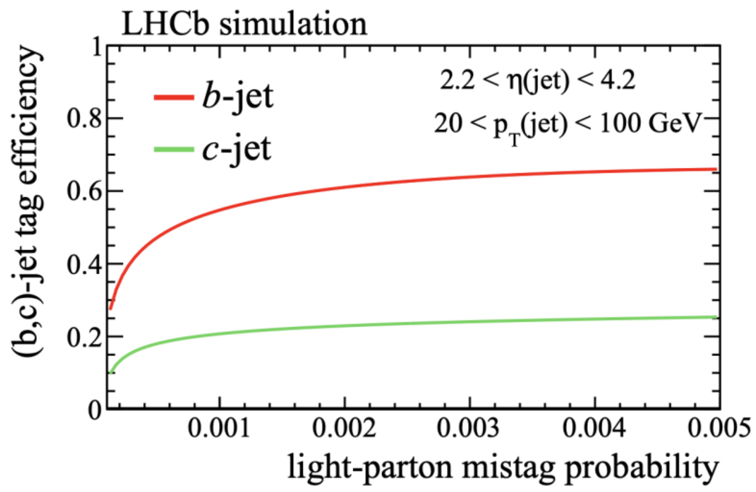
Efficiency	False Positive Rate
0.85	0.1880
0.80	0.1154
0.75	0.0662
0.70	0.0306
0.65	0.0102
0.60	0.0050

Veto

Selection

Comparing to current SV tagger:

- At 60%, similar FPR
- Ability to go to higher efficiency



LHCb-PAPER-2015-016

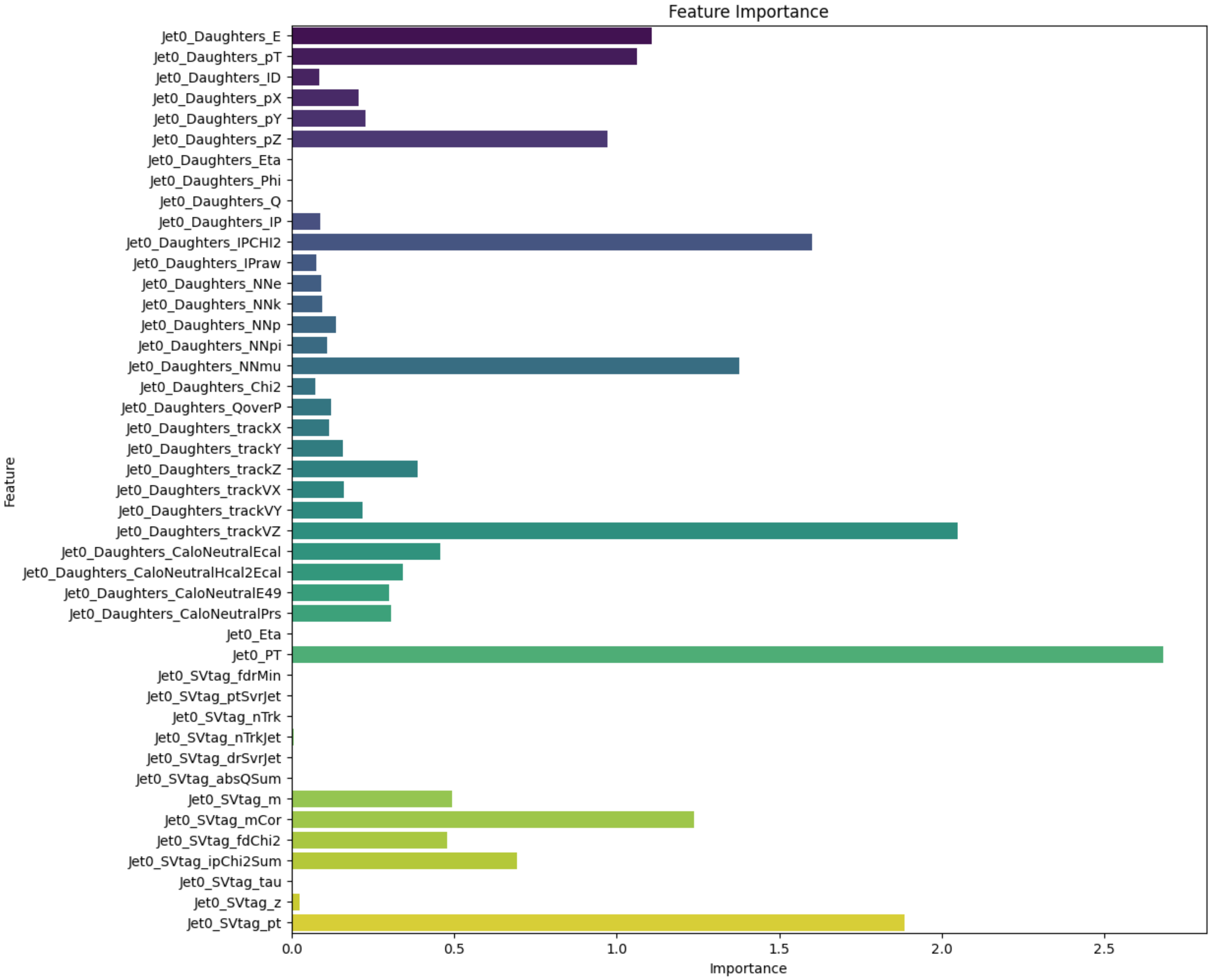
Results

Feature Ablation:

- Remove one feature at a time and compare predictions

Features with Importance > 0.25

Feature	Importance
Jet PT	2.680280
Daughters trackVZ	2.049136
SVtag pt	1.886200
Daughters IPCHI2	1.600275
Daughters NNmu	1.378289
SVtag mCor	1.236778
Daughters E	1.106416
Daughters pT	1.063479
Daughters pZ	0.972586
SVtag ipChi2Sum	0.692827
SVtag m	0.494002
SVtag fdChi2	0.477104
Daughters CaloNeutralEcal	0.457462
Daughters trackZ	0.388335
Daughters CaloNeutralHcal2Ecal	0.341928
Daughters CaloNeutralPrs	0.305089
Daughters CaloNeutralE49	0.298509



Conclusion

Summary

- First LHCb GNN jet tagger
- GNNs are desirable tools for jet tagging
- *b*-jets have SV which allow for improved identification
- Data processing: TM, selection, and graph creation
- Similar performance already to current SV tagger
- Flexible for physics scenarios

Next steps

- Apply to *b* vs *c*, *c* vs light (+ fat jets w/ HF jets inside)
- Migrate to Run 3 samples and retrain
- Integrate into Moore/ HLT2 HF jet selections
- Move to UC GPUs for further training → allow for faster turn around and larger training sample

Physics Applications

Veto	Selection
WW	h $\rightarrow b\bar{b}$ asymmetry
WZ	h $\rightarrow c\bar{c}$ asymmetry
ZZ	top quark measurements
Z + jets	Z + jets
h $\rightarrow$ WW	h $\rightarrow cx$ analysis

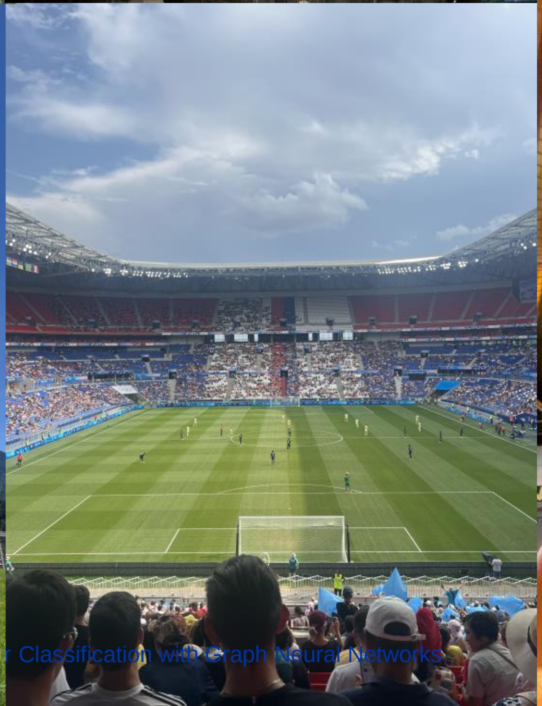
Acknowledgements

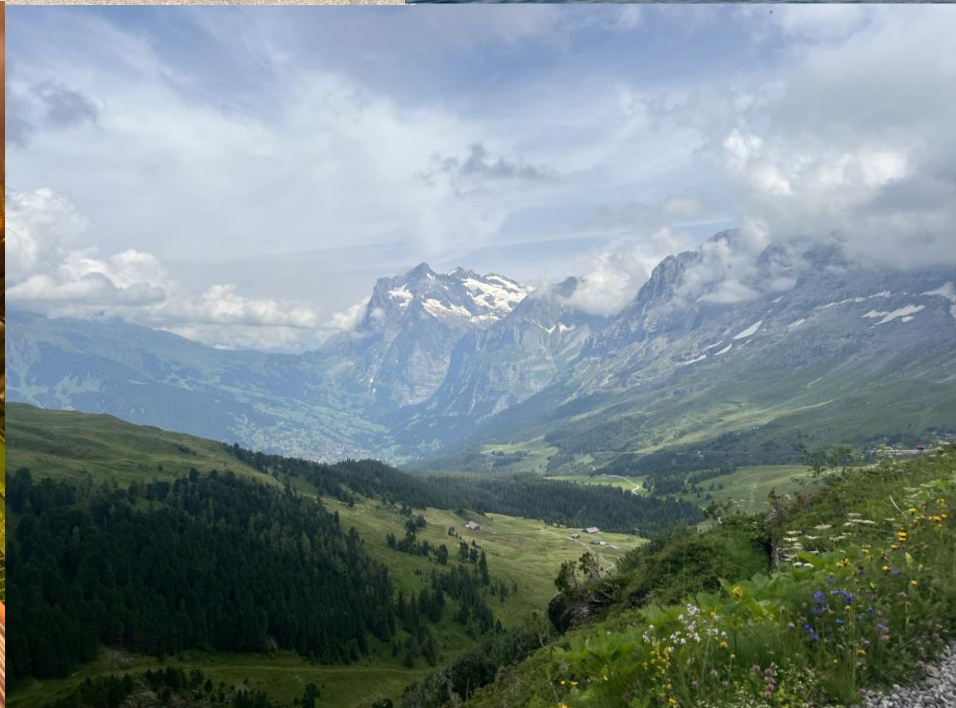
I would like to thank Dr. Conor Henderson and Dr. Nathan Allen Grieser for their mentorship and guidance this summer, as well as Dr. Simon Akar for his valuable input. Additionally, I would like to acknowledge the University of Michigan REU Program and the CERN Summer Student Program for their support.











References

Bain, Reggie. "Jets Aren't Just a Game of Tag Anymore." *ParticleBites*, 19 Feb. 2017, www.particlebites.com/?p=4007.

"Graph Neural Network (GNN)." *SEON*, 10 Nov. 2023, seon.io/resources/dictionary/graph-neural-network-gnn.

Graph Neural Network Jet Flavour Tagging with the ATLAS Detector, cds.cern.ch/record/2811135/files/ATL-PHYS-PUB-2022-027.pdf.

Hussein, Omar. "Graph Neural Networks Series: Part 4: The GNNS, Message Passing & over-Smoothing." *Medium*, The Modern Scientist, 17 May 2023, medium.com/the-modern-scientist/graph-neural-networks-series-part-4-the-gnns-message-passing-over-smoothing-e77ffee523cc.

Identification of Beauty and Charm Quark Jets at Lhcb - Arxiv, arxiv.org/pdf/1504.07670.pdf.

"Introduction to Jets." *ATLAS Software Documentation*, atlassoftwaredocs.web.cern.ch/AnalysisSWTutorial/obj_jet_intro/.

"Measurement of Inclusive and Differential Cross Sections for W + W − Production in Proton-Proton Collisions at √ s = 13.6 TeV." *CMS*, cms-results.web.cern.ch/cms-results/public-results/publications/SMP-24-001/index.html.

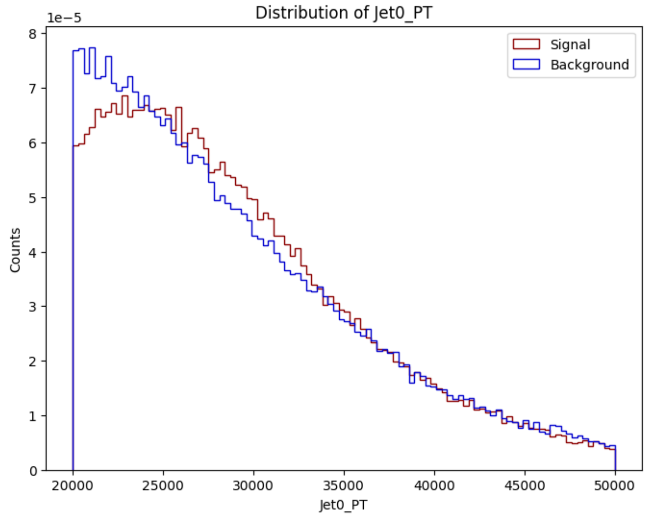
"Measurements of Production Cross Sections of WZ and Same-Sign WW Boson Pairs in Association with Two Jets in Proton-Proton Collisions At ." ArXiv, 4 May 2020, cds.cern.ch/record/2716981/.

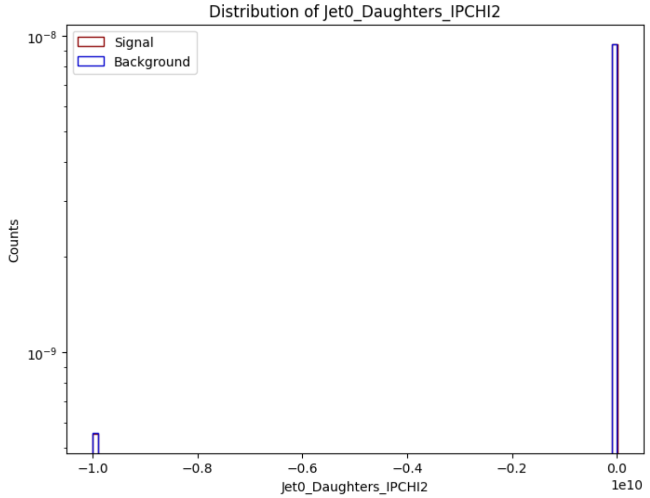
Nagai, Yoshikazu. *B-Tagging in Atlas*, indico.cern.ch/event/242419/contributions/520667/attachments/412165/572722/B-tag2012.pdf.

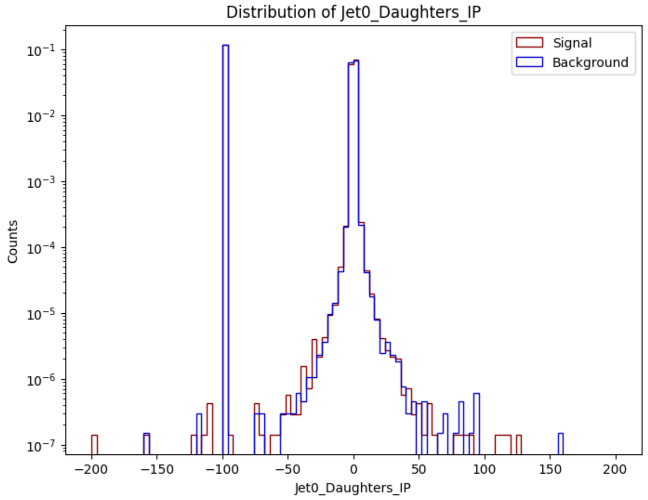
Vacavant, Laurent. *ATLAS Results on H(125) Fermion (Bb) Decays: Observation of Hbb Decays*, indico.ijclab.in2p3.fr/event/4754/contributions/15547/subcontributions/1336/attachments/13169/15826/\Hbb-HHunting2018-LV_FINAL.pdf.

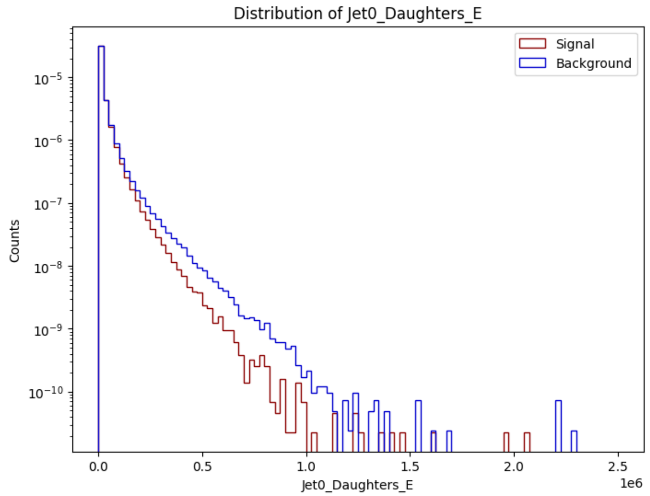
home.cern

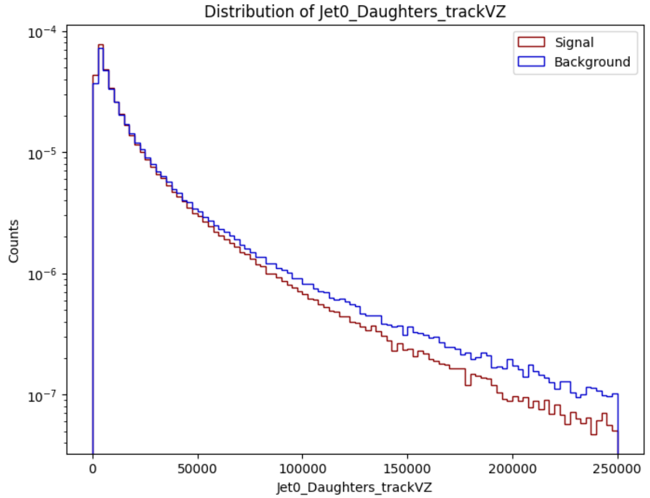
Results

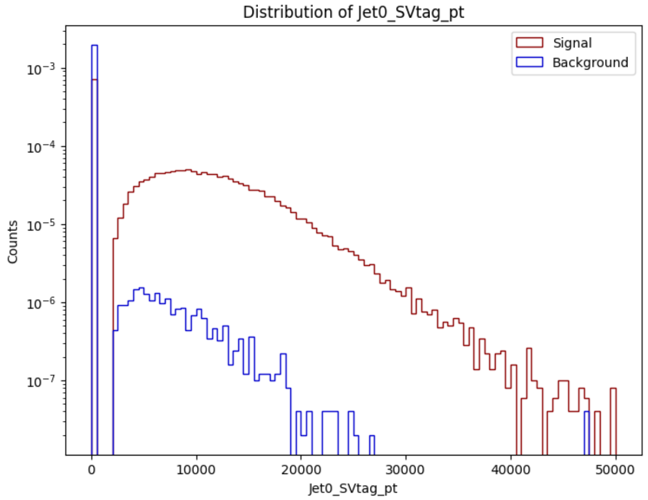

Distribution of Jet0_PT
1e−5
Signal
Background
Counts
Jet0_PT


Distribution of Jet0_Daughters_IPCHI2
Signal
Background
Counts
Jet0_Daughters_IPCHI2
1e10

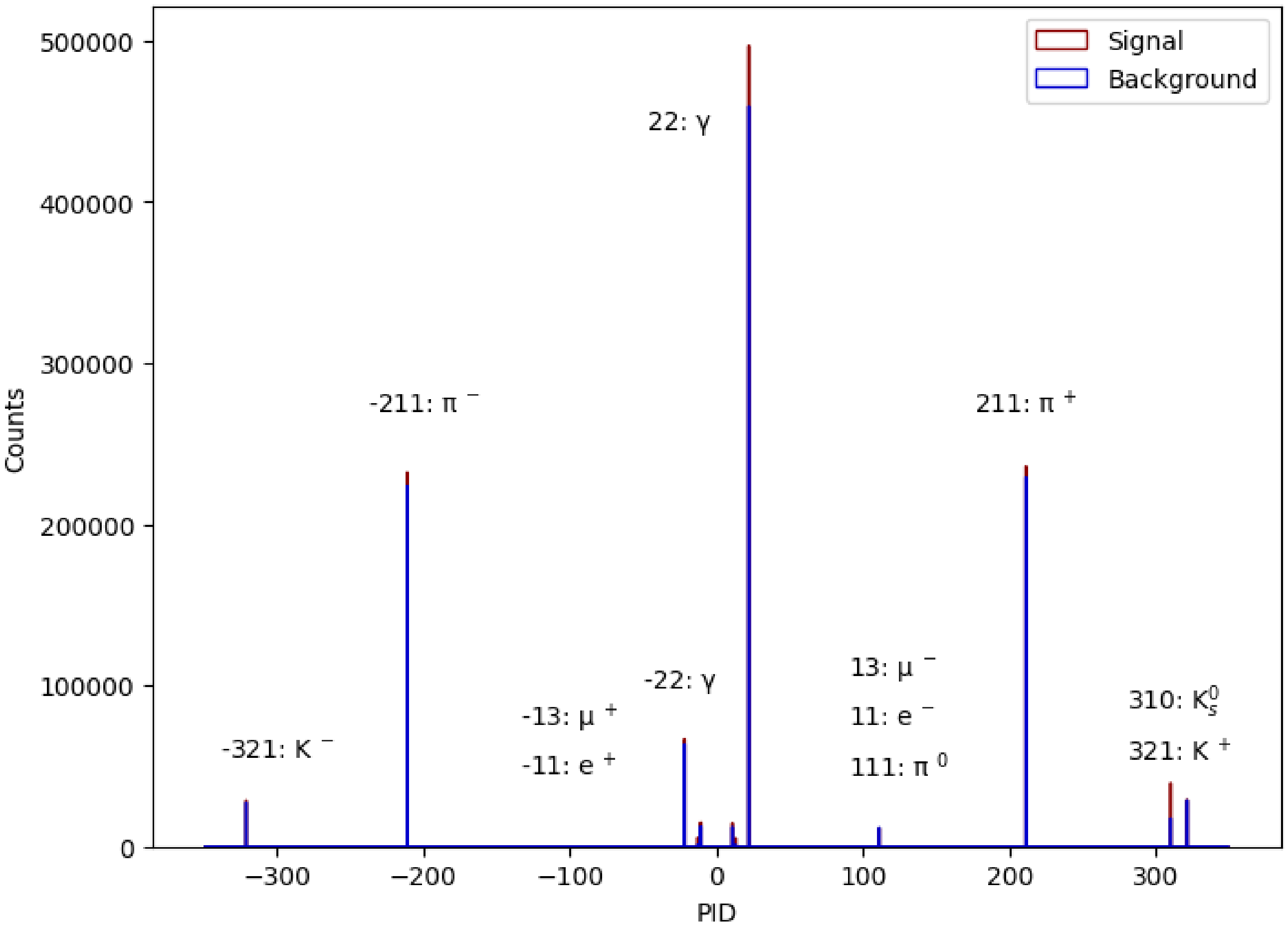

Distribution of Jet0_Daughters_IP
Signal
Background
Counts
Jet0_Daughters_IP


Distribution of Jet0_Daughters_E
Signal
Background
Counts
Jet0_Daughters_E
1e6


Distribution of Jet0_Daughters_trackVZ
Signal
Background
Counts
Jet0_Daughters_trackVZ


Distribution of Jet0_SVtag_pt
Signal
Background
Counts
Jet0_SVtag_pt

Particles Present in Jets



Particle ID	Signal Counts	Background Counts
-3122: Λ	4861	2501
-2212: p^-	9640	16193
-321: K^-	28599	27363
-211: π^-	232082	223814
-22: γ	66737	63545
-13: μ^+	5456	1114
-11: e^+	14963	12551
11: e^-	14550	12104
13: μ^-	5055	945
22: γ	496754	459141
111: π^0	11397	11790
211: π^+	235952	229267
310: K^0_s	39285	17327
321: K^+	28977	28592
2212: p^+	10255	17210
3122: Λ	4722	2409

Graph Representations

Node Features

- $[N_{nodes}, N_{features}]$
- Tensor representing features for each node

Edge Index

- $[2, N_{edges}]$
- Graph connectivity

Truth Labels

- One label per graph

Batching

- Multiple graphs are batched together
- Parallel processing in single forward pass
- Batch size = 64

GNN Layers – PyTorch Geometric

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

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

Loss and Optimization

Binary Cross Entropy Loss

$$\ell(x,y) = L = \{l_1, \dots, l_N\}^{\intercal}, l_n = -w_n[y_n \cdot \log \sigma(x_n) + (1-y_n) \cdot \log(1-\sigma(x_n))]$$

AdamW

- Stochastic gradient descent method
- Separates weight decay from gradients
- Better regularization than Adam

AdamW Algorithm

input : γ(lr), β_1, β_2(betas), θ_0(params), $f(\theta)$(objective), ϵ (epsilon)
λ(weight decay), *amsgrad*, *maximize*
initialize : $m_0 \leftarrow 0$ (first moment), $v_0 \leftarrow 0$ (second moment), $\widehat{v_0}^{max} \leftarrow 0$

for $t = 1$ **to** ... **do**
  if *maximize* :
    $g_t \leftarrow -\nabla_\theta f_t(\theta_{t-1})$
  else
    $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$
  $\theta_t \leftarrow \theta_{t-1} - \gamma\lambda\theta_{t-1}$
  $m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1) g_t$
  $v_t \leftarrow \beta_2 v_{t-1} + (1-\beta_2) g_t^2$
  $\widehat{m_t} \leftarrow m_t / \big(1-\beta_1^t\big)$
  $\widehat{v_t} \leftarrow v_t / \big(1-\beta_2^t\big)$
  if *amsgrad*
    $\widehat{v_t}^{max} \leftarrow \max(\widehat{v_t}^{max}, \widehat{v_t})$
    $\theta_t \leftarrow \theta_t - \gamma \widehat{m_t} / \big(\sqrt{\widehat{v_t}^{max}} + \epsilon\big)$
  else
    $\theta_t \leftarrow \theta_t - \gamma \widehat{m_t} / \big(\sqrt{\widehat{v_t}} + \epsilon\big)$

return θ_t

Hyperparameter Tuning

Random Search

- Set ranges of hyperparameters
  - Number of Convolutional Layers (0 → 5) = 4
  - Hidden Channels (64, 128, 256) = 128
  - Dropout Rate (0.1 → 0.5) = 0.2
  - Weight Decay $(10^{-6} \rightarrow 10^{-2}) = 10^{-4}$
- Choose best performance: highest validation accuracy
- Repeat with narrower range of hyperparameters and more epochs