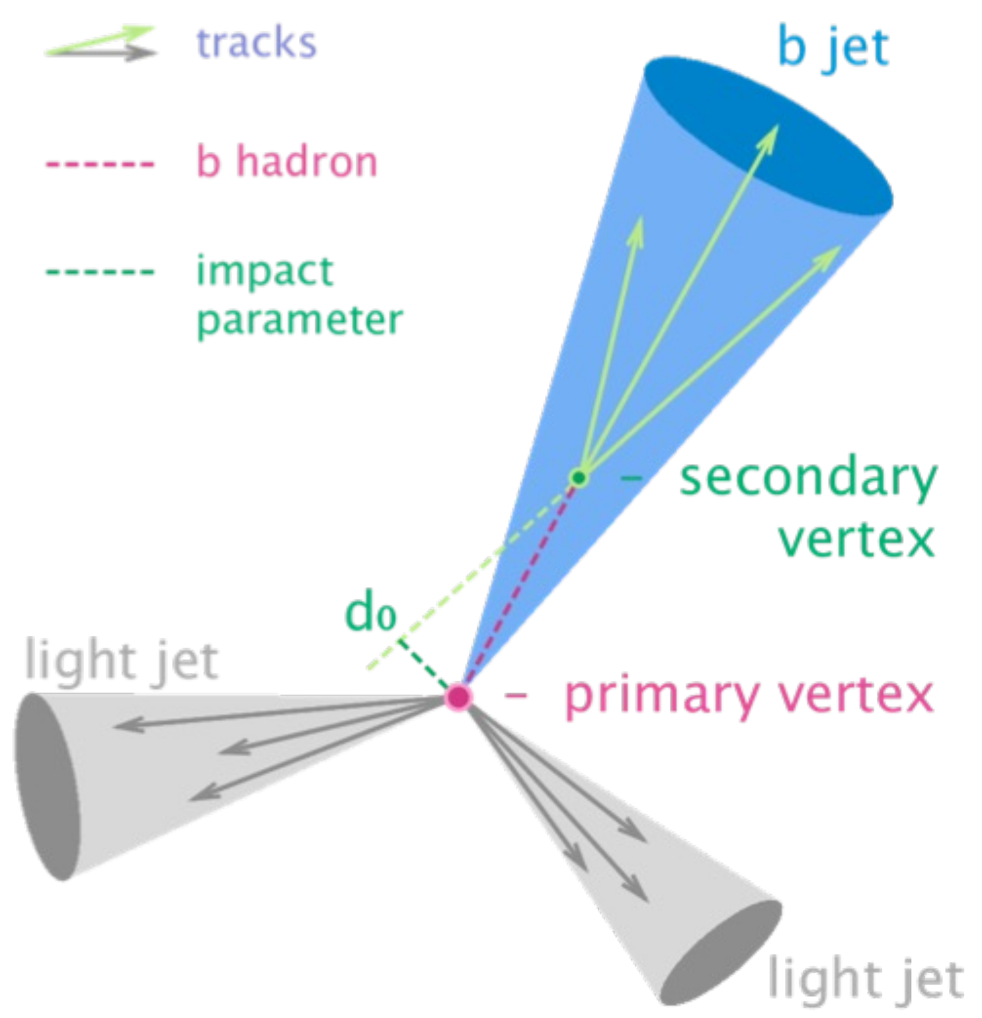
Jet Flavour Classification With Graph Neural Networks

Gabriella Pesticci | CERN Summer Student | Kenyon College
Dr. Conor Henderson | CERN | University of Cincinnati
Dr. Nathan Allen Grieser | CERN | University of Cincinnati

Motivation

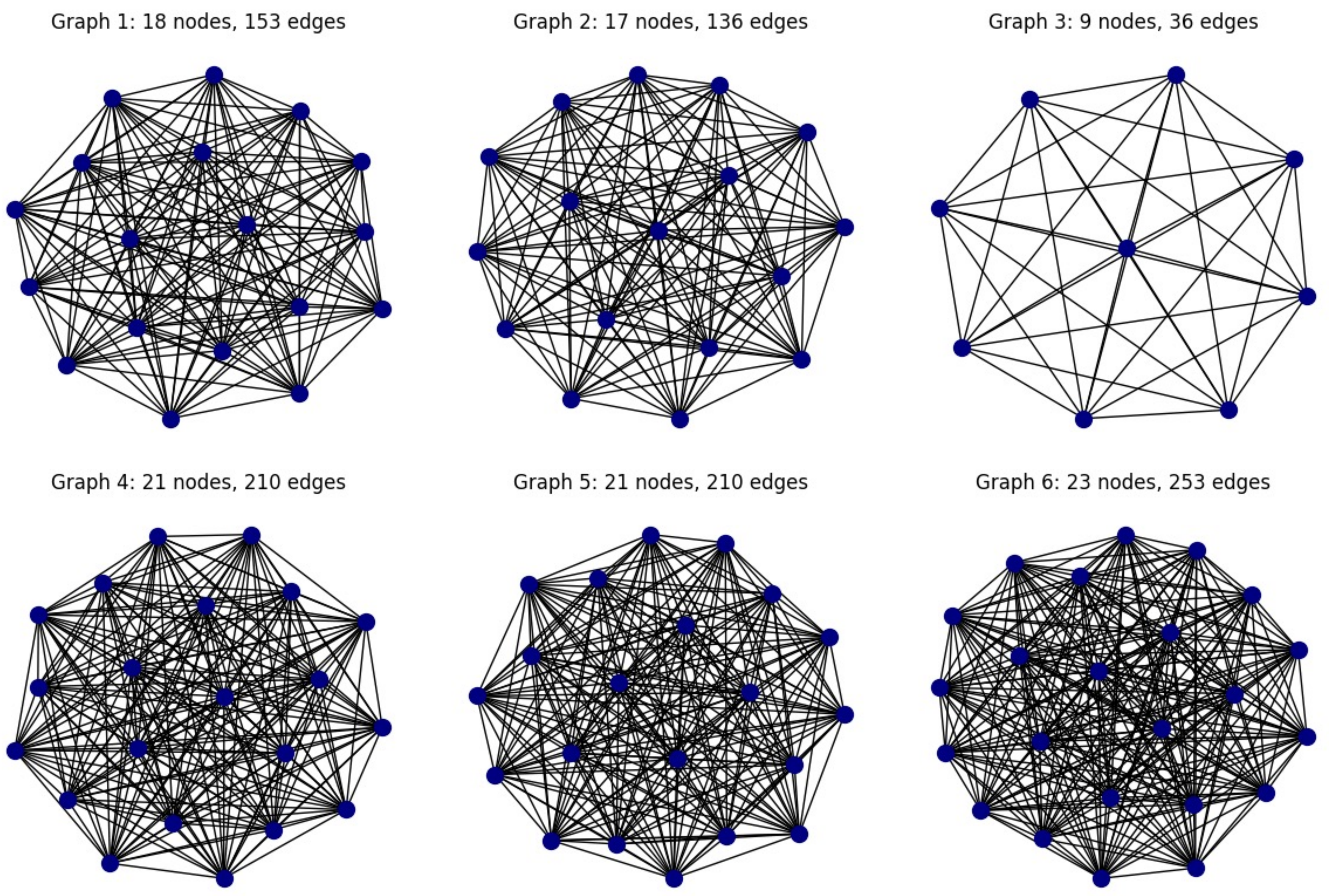
Proton-proton collisions at the Large Hadron Collider (LHC) generate various particles, including quarks and gluons. These particles hadronize due to QCD confinement, forming jets that are detected by the LHCb. Identifying the type of jet is crucial for reconstruction and understanding of collision events. My goal is to utilize deep learning, specifically graph neural networks, to classify *b*-jets, based on characteristics of both the jet itself, and the kinematics of the daughter particles. This GNN will be applied for further classification of *c*-jets and fat jets.



3-jet event producing a *b*-jet [1]

Graph Neural Networks

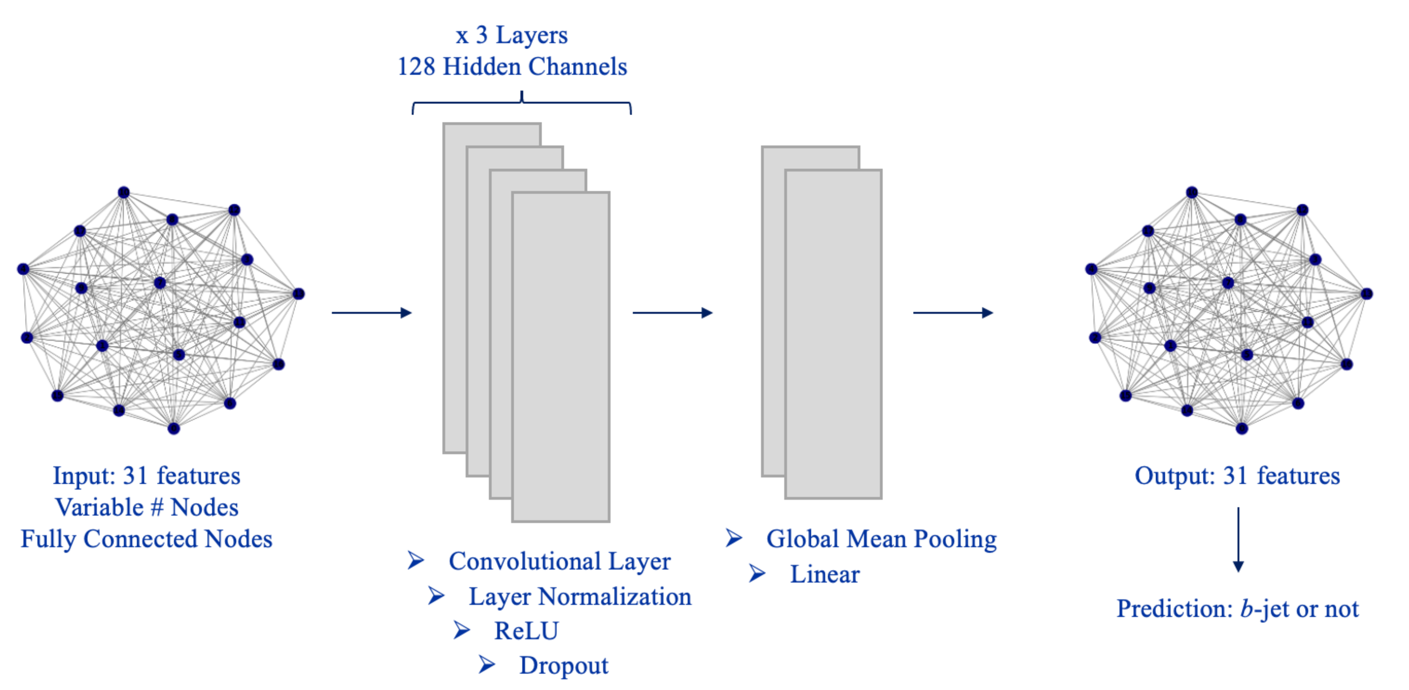
Graph neural networks (GNN) are powerful deep learning algorithms which work with graph-structured data. GNNs are advantageous for performing binary classification of jet flavours due to their ability to capture complex relationships within data and handle inhomogeneous data.



6 graphs included in the sample

Graphs: one graph represents a leading jet in the dijet sample
Nodes: one for each daughter in the jet
Edges: fully connected nodes
Features: 31 total, description of overall jet and individual daughter kinematics

GNN Architecture



Jet Features
η
p_T
Track Features
E
p_T
ID
p (x,y,z)
η
ϕ
Q
IP
IP χ^2
IP raw
NNe
NNk
NNp
NNπ
NNμ
χ^2
Q/P
Track (x, y, z)
Track V (x, y, z)
CaloNeutralEcal
CaloNeutralHcal2Ecal
CaloNeutralE49
CaloNeutralPrs

Data Preparation

Jet Features:
- Describes overall jet kinematics
- Shared by all daughters in the jet

Daughter Features:
- Unique kinematics for each daughter in the jet

Truth Matching:
- All data:
  MC Match = 1
- Signal data:
  MC Jet EfB > 0.6
- Background data:
  MC Jet EfB < 0.6
  MC Jet EfD < 0.6

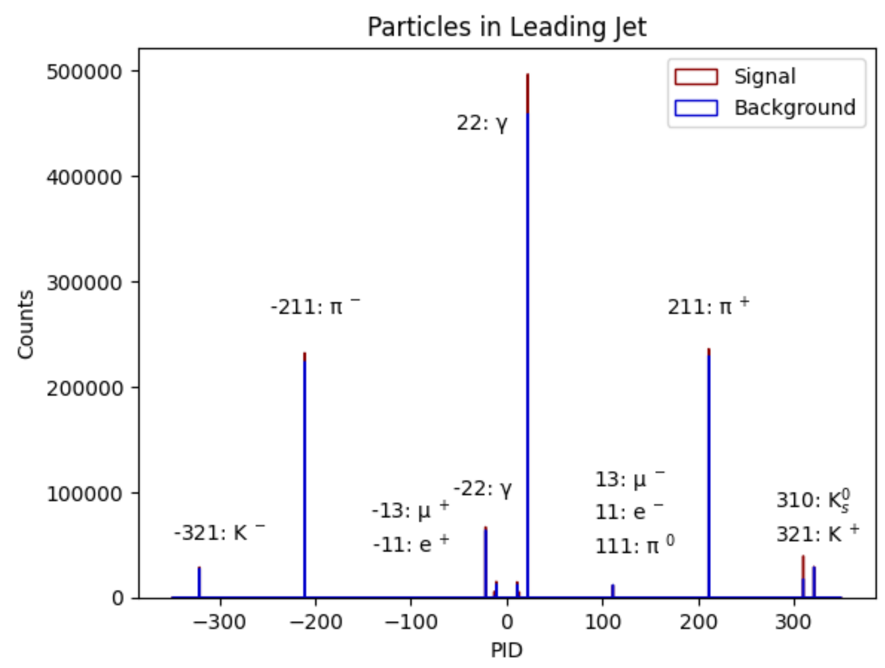
Jet Selection:
- pT > 20 GeV
- 2.2 < η < 4.4

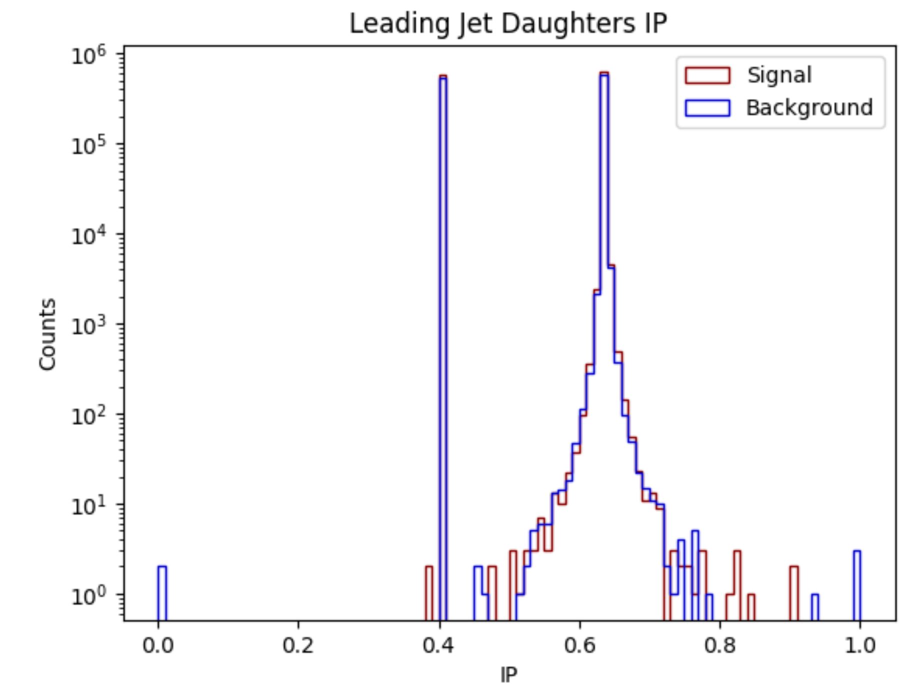
Muon Features:
- Currently not included
- N muons
- Prob NN mu
- Muon IP Chi2

Calculate ΔR for muons:
- $\Delta R^2 = \eta^2 + \phi^2$
- Select ΔR < 0.5

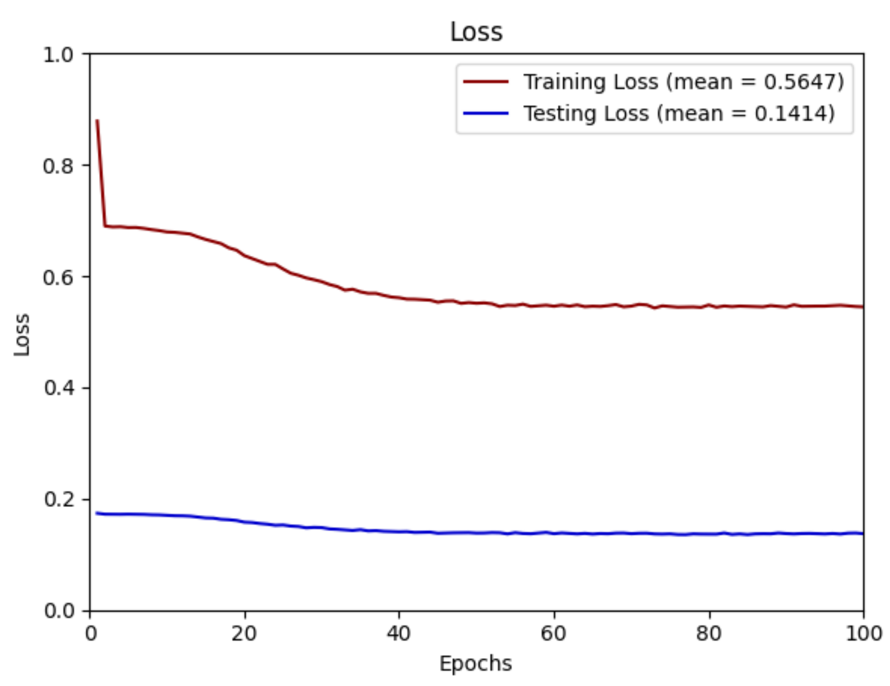
Particles Present in Leading Jet

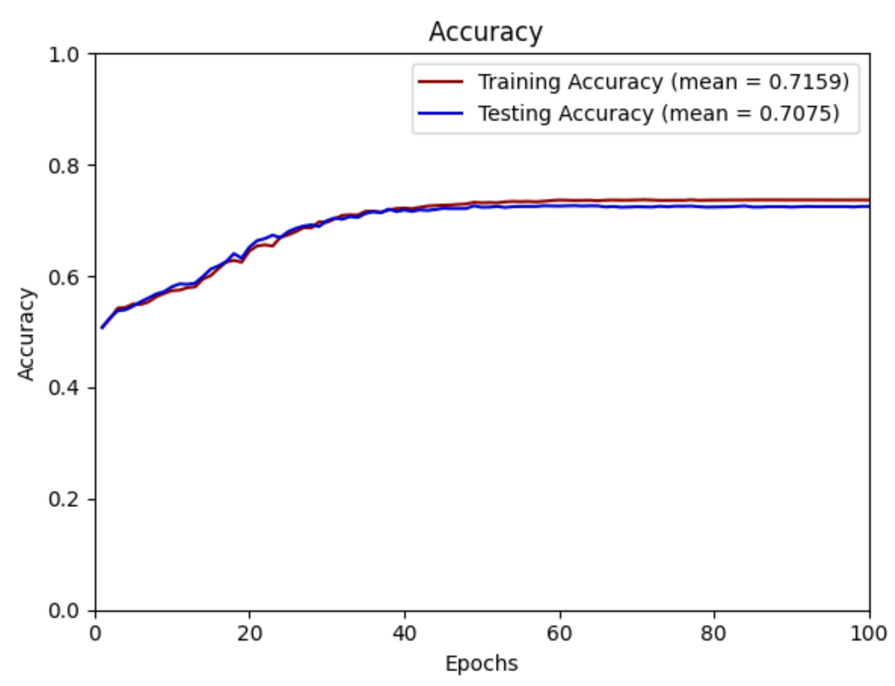
Particle ID	Signal Counts	Background Counts
-3122: Λ	4861	2501
-2212: p^-	9640	16193
-321: K^-	28599	27363
-211: π^-	232082	223814
-22: γ	66737	63545
-13: μ^+	5456	1114
-11: e^+	14963	12551
11: e^-	14550	12104
13: μ^-	5055	945
22: γ	496754	459141
111: π^0	11397	11790
211: π^+	235952	229267
310: K_s^0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





Results



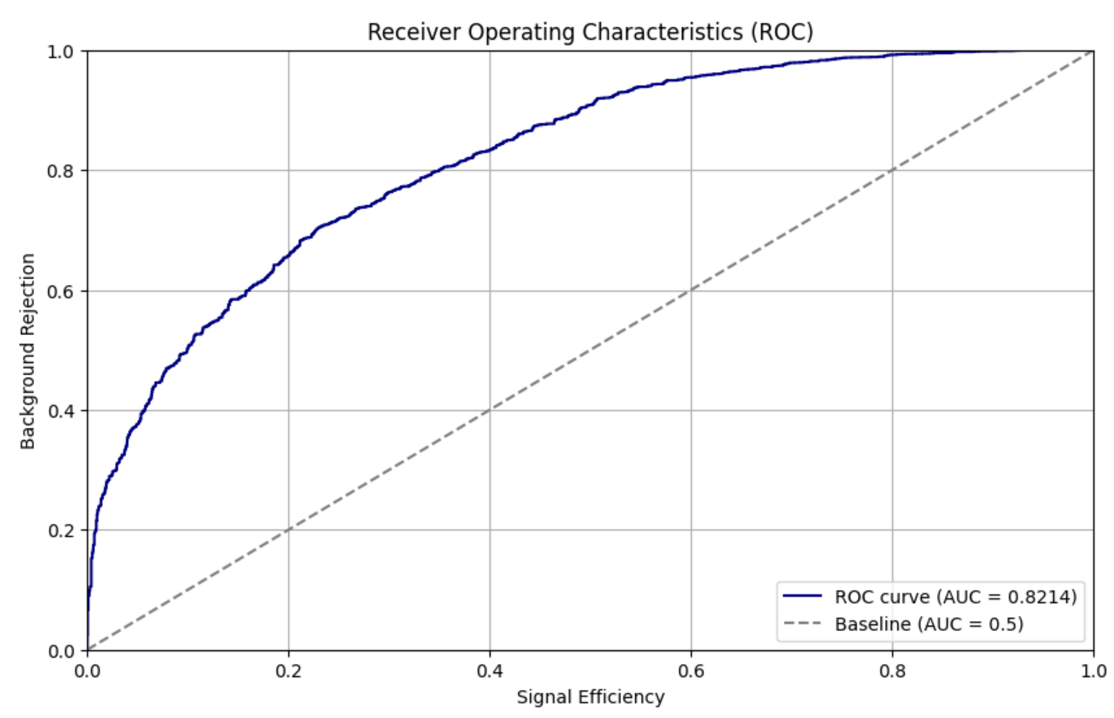


Loss
- Training: 0.8786 → 0.5462
- Testing: 0.1738 → 0.1366

Accuracy
- Training: 0.5086 → 0.7366
- Testing: 0.5073 → 0.7250

AUC
- AUC = 0.8214

Next Steps
- Tune hyperparameters
- Calculate ΔR
- Apply muon features
- Add SV tagging variables



References

[1] Bain, Reggie. “Jets Aren’t Just a Game of Tag Anymore.” *ParticleBites*, 19 Feb. 2017, www.particlebites.com/?p=4007.

Acknowledgements

I would like to thank Dr. Conor Henderson and Dr. Nathan Allen Grieser for their mentorship and guidance throughout this summer. Additionally, I would like to acknowledge the University of Michigan REU Program, the CERN Summer Student Program, and my fellow summer students for their support.