DEPARTMENT OF PHYSICS & ASTRONOMY

Identifying the Origin of Jets with Machine Learning

PHAS0052: Group Report by

Group 29

Department of Physics and Astronomy
University College London
London,UK

28th March 2025

Abstract {CC}

In recent times, machine learning has been used to identify the origin of particle jets in the Large Hadron Collider. This project focuses on the LundNet graph neural network, with the primary objective of distinguishing the jets originating from boosted W and Z bosons from QCD jets. Firstly, 2 taggers were trained on narrow and wide mass windows for W jets. The wide-mass W tagger maintained the background mass distribution shape, while effectively suppressing QCD background jets, showing that mass sculpting effects were mitigated.

Next, the wide-mass W tagger was tested on Z jets, due to the similarity in jet structure, showing its applicability despite a lower background rejection rate, compared to the Z jet-specific classifier. The project then turns to distinguishing between W and Z bosons, using mass cuts and a new second-stage LundNet classifier, trained only on these vector bosons. Results showed moderate success in isolating W bosons, but distringuishing Z bosons proved more challenging.

Contents

Page

1. Introduction {CC and BB} . . . 3
   - 1.1 Background . . . 3
   - 1.2 Aims and Objectives . . . 3
   - 1.3 The Lund Plane and LundNet . . . 4
2. Normalization of the test samples {AA} . . . 5
3. Investigation of W jets {EE} . . . 6
   - 3.1 Tagger trained on regular-mass W sample . . . 6
   - 3.2 Tagger trained on wide-mass W sample . . . 6
   - 3.3 Performance of the W taggers . . . 6
4. Investigation of Z jets {CC} . . . 8
   - 4.1 Applying wide-mass W tagger to the Z sample . . . 8
   - 4.2 Performance of the Z taggers . . . 9
5. Distinguishing between W and Z jets {AA and DD} . . . 10
   - 5.1 Determining the mass cuts {AA} . . . 10
   - 5.2 Distinguishing W from Z Jets using Machine Learning {DD} . . . 11
     - 5.2.1 Data Preparation and Tagger Training . . . 11
     - 5.2.2 Tagger Score Distributions . . . 12
     - 5.2.3 Performance Evaluation . . . 13
6. Conclusion {AA and EE} . . . 15
   - 6.1 Future Outlook . . . 16

References **17**

1 Introduction {CC and BB}

1.1 Background

The largest and most powerful particle accelerator in the world is the Large Hadron Collider [1]. Experiments at the LHC can allow us to take precise measurements of the processes of the standard model, and to potentially look beyond the standard model [2]. The identification of electroweak scale heavy particles, such as W, Z and Higgs bosons, is a significant task in these investigations of the standard model [3]. Due to the high energies involved in LHC collisions, electroweak scale particles can become boosted to the point where their hadronic decay products are modelled as jets, a collimated spray of particles [4]. On the other hand, light quarks and gluons produce very similar QCD jets [5]. Identifying the origin of these jets will therefore allow us to identify particles produced at the LHC. To determine the physical origins of jets, machine learning is often used to investigate the radiation patterns within jets [3].

1.2 Aims and Objectives

In this project we focus on the LundNet tagger, which utilizes a graph neural network and the 'Lund Plane' to represent jets. LundNet has demonstrated a strong performance in distinguishing boosted jets from background processes compared to traditional tagging techniques[3]. We aim to use the LundNet tagger to identify W and Z bosons within the QCD background. We do this by training the tagger on different simulated samples of W and Z bosons, resulting in multiple classifiers and analysing the tagging performance of these neural network classifiers. We then investigate methods of separating between W and Z bosons themselves, first using the method of finding appropriate mass cuts and then by training a second-stage tagger to specifically differentiate between the two heavy particles. A key focus of our study is on mitigating mass sculpting, a phenomenon where a classifier learns to tag purely on jet mass rather than internal structure for tagging.

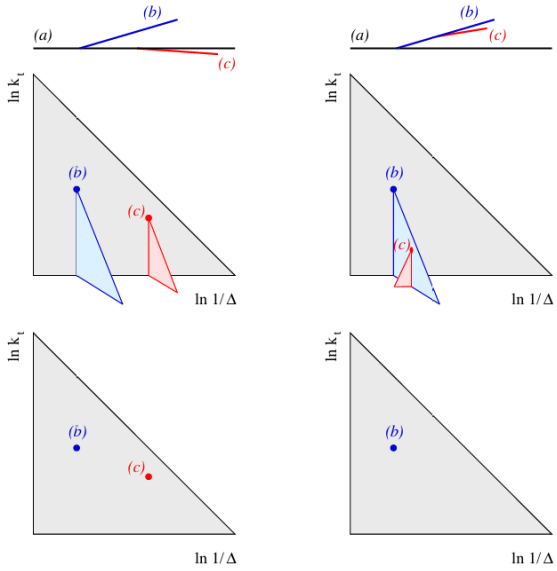


Figure 1 ***Lund Plane:*** Top: Jet Particles. Middle: Full Lund Plane. Bottom: Primary Lund Plane [6]

1.3 The Lund Plane and LundNet

The Lund Plane is a method of representing the internal structure of a jet. The jet particles are first clustered using an algorithm, grouping particles by their angle [3]. This results in a tree structure (Figure 1, top) where each branch represents a jet emission. At each branch splitting, the values of five kinematic variables are calculated: the angle between the two new branches Δ , the transverse momentum of the softer branch with respect to the harder branch k_t, the momentum fraction of the softer branch z, the invariant mass of the pair m and the azimuthal angle around the harder branches axis ψ. The branches are then plotted onto a graph with axes of ln k_t and ln $\frac{1}{\Delta}$. This jet representation efficiently holds information about jet substructure. The primary Lund sequence is especially useful in representing the radiation patterns within jets [6].

Furthermore the 2D graph can give us an indication of the particles from which the jet originated [2], for example hard, symmetrical splits tend to occur when a W boson decays into two quarks, leaving a clear signature in the upper left corner of the plane (the region of high-energy, wider-angle emissions). The Monte Carlo method is used to simulate collisions between protons [2]. LundNet is a graph neural network which utilizes the information within the Lund Plane representation at each node. This information is the set of kinematic variables, which are taken as the inputs to the graph neural network. We use LundNet 3 as our tagger, which takes the first three of the kinematic variables as inputs, $\ln k_t$, $\ln(\frac{1}{\Delta})$ and $\ln z$.

2 Normalization of the test samples {AA}

The testing W, Z, and QCD jet samples are generated from the Monte Carlo simulations. However, in real scenarios, these jet samples have different sizes relative to each other (e.g., there are much more QCD jets than W and Z jets [7]). Therefore, these jet samples are normalized by adding their corresponding weights [8]-[9].

For W and Z jets, their sample sizes are normalized by:

$$W_i = \frac{\sigma_i}{N_{gen_i}} \tag{1}$$

where W_i is the weight, σ_i is cross section of the jet, N_{gen_i} is the number of generated events (i denotes W or Z).

By contrast, the QCD jet sample (multiple event) have different cross sections for distinct numbers of generated events, where the corresponding generator filter efficiencies are also considered [8]:

$$W_{QCD} = \frac{\sigma_{QCD}}{N_{gen_{QCD}}} \times \epsilon_{gen} \tag{2}$$

where W_{QCD} are the weights applied to the QCD jet sample, σ_{QCD} are the cross sections, $N_{gen_{QCD}}$ are the numbers of generated events, and ϵ_{gen} are the generator filter efficiencies.

After normalization, the W, Z, and QCD jet samples are now in proper sizes relative to each other, as shown in Figure 2 below.

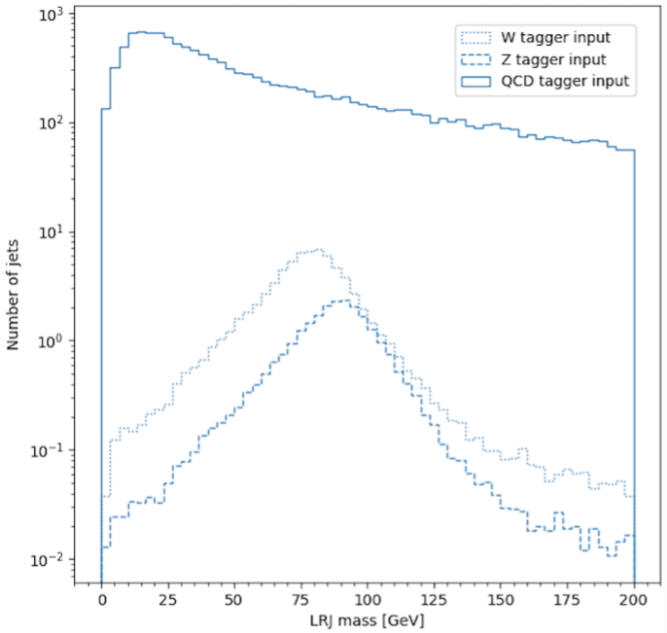


Figure 2 The normalized test samples of W, Z, and QCD jets with proper relative sizes.

3 Investigation of W jets {EE}

At the first stage, two Lund-jet-plane-based taggers are trained and applied to distinguish W jets from QCD dijets. The primary difference between these classifiers is the selection criterion for the W-boson jets used in training.

3.1 Tagger trained on regular-mass W sample

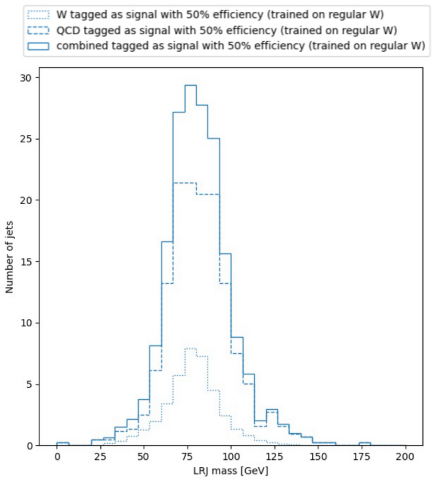
Tagger (1) is trained using regular-mass W samples (around 80 GeV). This focused selection facilitates high background rejection at a fixed signal efficiency by forcing the classifier to leverage the jet mass as a dominant feature. However, when tagger (1) is applied to broader test samples, its strong reliance on the narrow mass peak induces a distortion in the overall mass distribution, which is a phenomenon known as mass sculpting [2]. As shown in Figure 3 (a), this results in a background mass spectrum that clusters around the W mass peak, thereby complicating subsequent signal extraction and background modeling.

3.2 Tagger trained on wide-mass W sample

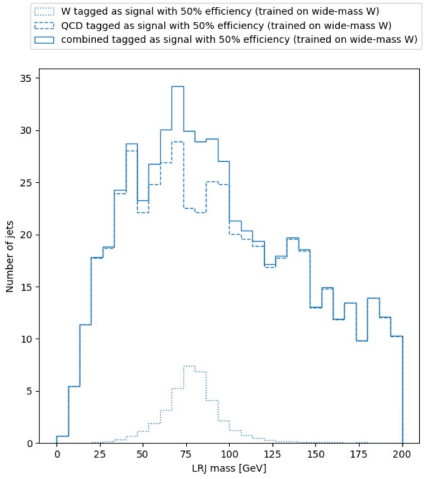
To mitigate the mass-sculpting issue, tagger (2) is trained on W jets spanning a wide-mass window (e.g., 50–110 GeV). By exposing the classifier to a wider spread of W jet masses, it is compelled to learn more nuanced features related to the jet's substructure and kinematic properties rather than overfitting to a narrow mass value. According to Figure 3 (b), tagger (2) achieves comparable background rejection power while preserving the intrinsic shape of the W-boson mass distribution. This quality is particularly beneficial when the classifier is applied to mixed samples containing W, Z, and QCD jets.

3.3 Performance of the W taggers

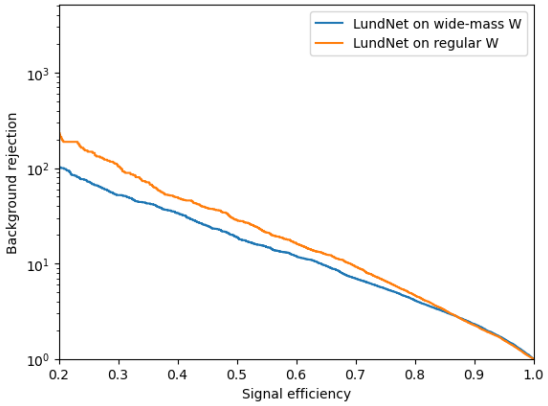
Following training, both classifiers are evaluated on a test set comprising a realistic mixture of W and QCD jets. Although tagger (1) exhibits strong background rejection in its training environment shown in Figure 3 (c), its heavy reliance on a narrow mass feature results in pronounced mass sculpting: QCD background jets are inadvertently shifted toward the W mass peak, distorting the overall mass spectrum. Conversely, tagger (2) not only achieves robust background rejection but also preserves the natural mass distribution of the W jets after tagging.



(a) Mass distribution of W, Z and QCD jets after passing through tagger (1).



(b) Mass distribution of W, Z and QCD jets after passing through tagger (2).



(c) Background QCD jets rejection versus signal efficiencies for tagger (1) and (2) in the mass range of 70-90 GeV.

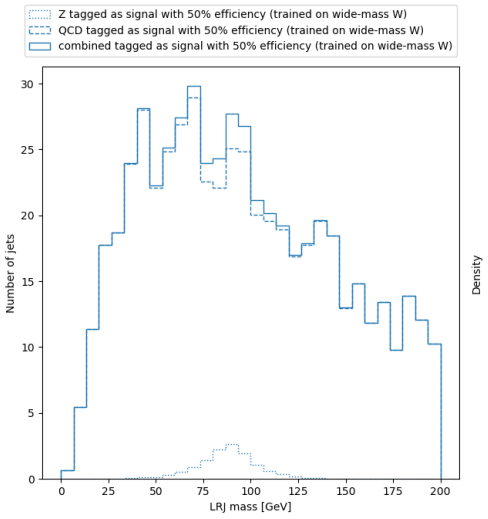
Figure 3 Output of tagger (1) and (2) on tagging W jets.

As shown in Figure 3 (c), tagger (1) shows a better performance by a factor of 1.7 compared to tagger (2) at 50% signal efficiency. However, despite tagger (1) learned to reject more QCD background jets due to its training on a regular-W sample, it failed to distinguish between W and QCD jets. This indicates that a narrower selection of jet mass in training leads to mass-sculpting as the tagger learned to pick out signals purely on selecting jets with a W mass, whereas tagger (2) achieves a similar background rejection while maintaining the ability to identify the signal jets.

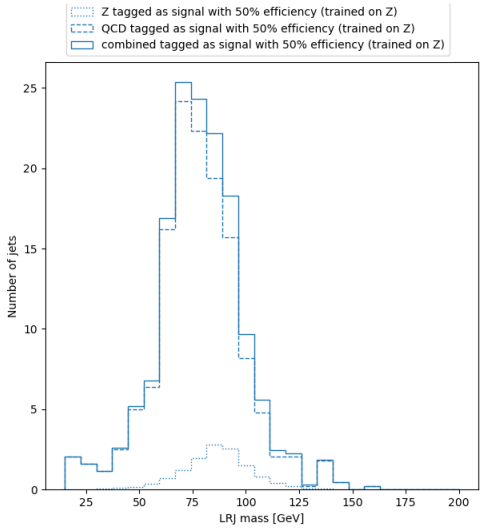
4 Investigation of Z jets {CC}

4.1 Applying wide-mass W tagger to the Z sample

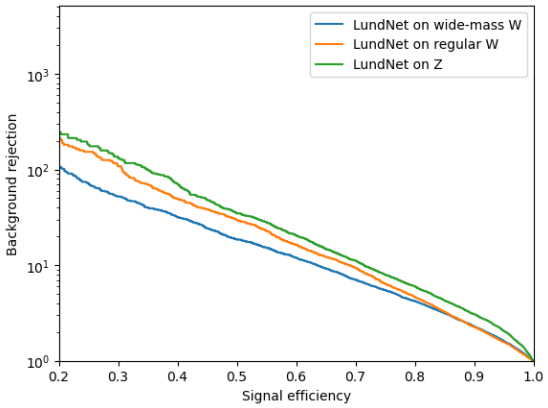
Next, tagger (2) was tested on a simulated data sample consisting of a mixture of QCD jets and jets originating from Z bosons, with Z jets treated as signal. This was done due to similarities in the substructures of jets originating from hadronically decaying W and Z bosons. The main difference would be the jet masses, as W bosons have a mass of roughly 80 GeV, while Z bosons have masses of around 91 GeV. To avoid mass-sculpting, tagger (2) was used rather than tagger (1).



(a) Mass distribution of Z and QCD jets after passing through tagger (2).



(b) Mass distribution of Z and QCD jets after passing through tagger (3).



(c) Background QCD jets rejection versus signal efficiencies for tagger (1), (2), and (3) in the mass range of 80-100 GeV.

Figure 4 Output of tagger (1), (2), and (3) on tagging Z jets.

Tagger (2) was trained on a wide mass sample, meaning it is less likely to ignore features of the jet substructure and classify jets purely by mass, therefore it should

potentially be able to successfully classify Z jets. In addition another tagger, tagger (3), was also trained using a realistic Z jet sample within a tight mass window. The narrow mass window of Z jets used resulted in a high background rejection rate at the 0.5 signal efficiency chosen, however this was also likely to cause a mass sculpting peak around the Z mass peak.

4.2 Performance of the Z taggers

As shown in Figure 4 (c), the tagger trained on the Z jet sample exhibited a higher background rejection rate at the same signal efficiency of 0.5 (within a mass window of 80-100 GeV). This means that tagger (3) was better at filtering out background QCD jets while keeping the same fraction of Z jets compared to tagger (2). On the other hand, in the overall mass distribution including the background jets in Figure 4 (b), we see a clear distortion, resulting in a peak similar to the signal peak. This shows that mass sculpting has occurred with tagger (3).

Meanwhile, we see the results of the tagger (2) test on the same sample in Figure 4 (a). The shape of the background distribution does not mimic the shape of the signal, meaning that mass sculpting has not occurred. There is a peak of tagged Z jets within the QCD background, meaning that the despite the fact that tagger (2) was tested on a wide-mass sample of W jets, it can successfully tag Z jets as predicted, although at a lower background rejection rate than tagger (3).

5 Distinguishing between W and Z jets {AA and DD}

In this section, two approaches will be investigated to distinguish between W and Z jets, which is challenging because both bosons decay hadronically and exhibit similar jet substructures. W and Z bosons have similar behaviors: they both decay into two jets producing two-prong structures [10].

5.1 Determining the mass cuts {AA}

From the above investigations it can be noticed that the wide-mass W tagger is the one that successfully picked out W or Z jets as signals from the QCD background jets. Therefore, the total contribution of W and Z jets can be obtained by applying this tagger onto a mixed sample of W, Z, and QCD jets, as shown in Figure 5.

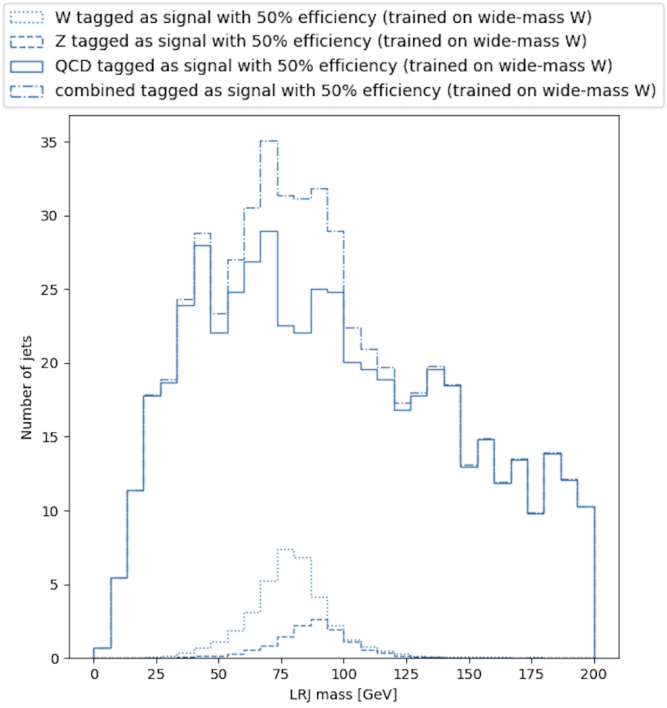


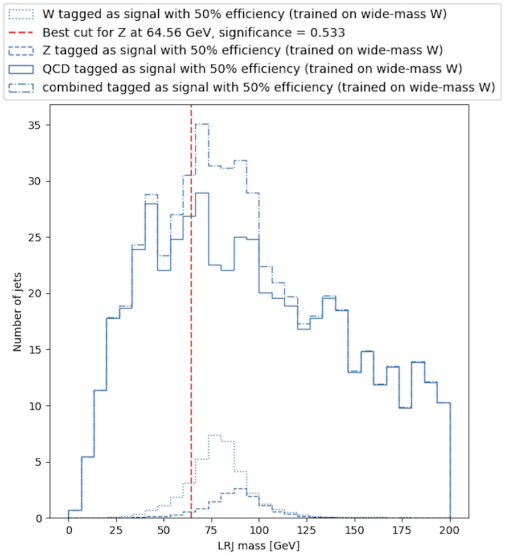
Figure 5 Tagging W, Z, and QCD jets using the wide-mass W tagger, the signal "bump" now represents the total contribution of W and Z jets.

To distinguish between W and Z contributions, two best mass cuts are determined to separate W jets from the combination of Z and QCD jets, and the other way around. In this case, the mass cut with "best" position is defined as the one with highest signal significance:

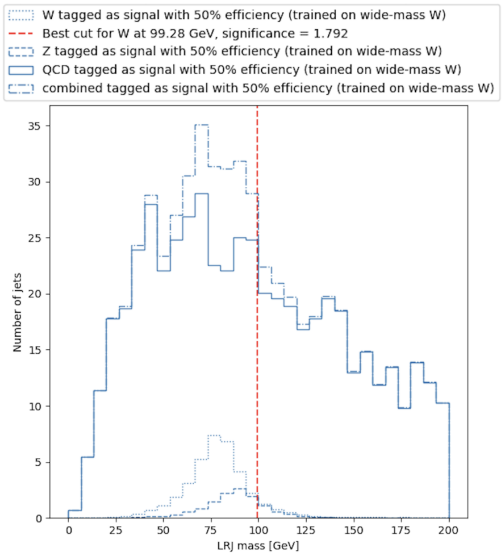
$$Z_{sig} = \frac{N_{signal}}{\sqrt{N_{signal} + N_{background}}} \tag{3}$$

where Z_{sig} is the significance, N_{signal} and $N_{background}$ are the total number of signal and background jets at that point (including weights). Note that when determining the mass cut for Z jets, the signal and background jets will be the corresponding ones on the right of the cut (under the signal bump and QCD background curve), and vice versa for W jets.

For the mass cuts in the range of 25 - 125 GeV, the ones with the highest values of Z_{sig} for Z and W jets are determined to be at 64.56 GeV and 99.28 GeV respectively, as shown in Figure 6. It can be noticed that the region on the left of the cut has a considerable selection of W signal jets and a good rejection of Z and QCD background jets, which indicates that the mass cut for W appears to perform well based on the graphs generated. Conversely, the mass cut for Z is selecting nearly the whole Z peak as a signal, which makes sense because it also rejects a significant number of QCD background jets.



(a) The mass cut (54.56 GeV) with highest significance (0.533) for classifying Z jet region (on the right of the cut).



(b) The mass cut (99.28 GeV) with highest significance (1.792) for classifying W jet region (on the left of the cut).

Figure 6 The best mass cuts determined for classifying Z and W contributions within the signal bump, (a) the best cut for Z jets, (b) the best cut for W jets.

5.2 Distinguishing W from Z Jets using Machine Learning {DD}

5.2.1 Data Preparation and Tagger Training

Following the application of the wide-mass W tagger to identify vector boson jets from QCD [3, 6], a second-stage tagger was developed to separate W and Z jets within the boson-like category. This classifier was trained on a refined dataset composed exclusively of jets that had passed the first-stage tagger's signal threshold, ensuring a sample with

high vector boson purity. These jets were then relabeled — Z jets as signal and W jets as background — to facilitate binary classification.

As both bosons decay into two-prong topologies and have similar masses, distinguishing them relies on subtle substructure features[11]. The same Lund plane variables ($\ln k_T$, $\ln \Delta$, $\ln z$) used in the first tagger were reused[3, 6]. A LundNet-based graph neural network was trained on this dataset using previously optimized hyperparameters. Training was conducted on the Dias high-performance cluster with model selection based on validation performance.

5.2.2 Tagger Score Distributions

The effectiveness of each tagger is illustrated through score distributions. Figure 7 shows the score from the initial wide-mass W tagger, where W and Z jets appear mostly overlapping — indicating difficulty in separation at this stage. This is expected, as the tagger was trained only to separate vector bosons (W/Z) from QCD background, and not to distinguish between the bosons themselves. The output score is therefore not well-optimized to identify differences between W and Z jets, which have highly similar substructures due to their shared two-prong decay topologies.

Figure 8 shows the output from the second-stage classifier trained specifically to separate W and Z jets. A clearer distinction is observed: W jets are now clustered near higher scores, while Z jets are pushed toward lower scores. Although the overlap is still present, this improvement demonstrates that the network has learned to exploit subtle differences in the substructure — possibly in the splitting scales or radiation patterns captured in the Lund plane variables ($\ln k_T$, $\ln \Delta$, $\ln z$) [3, 6].

Nonetheless, even with this separation in score space, full discrimination remains difficult. In previous analyses, mass distributions of the tagged jets at the 50% signal efficiency threshold showed that a substantial portion of Z and QCD jets were still retained in the tagged sample — reinforcing the idea that the structural differences between W and Z jets are subtle and overlapping in feature space. This aligns with experimental findings from early hadronic measurements of W and Z decays [10], where clean separation was also challenging in dijet topologies. While the classifier succeeds in reshaping the score distributions to better isolate W jets, it still operates under fundamental physical constraints imposed by the similarity of the decay products.

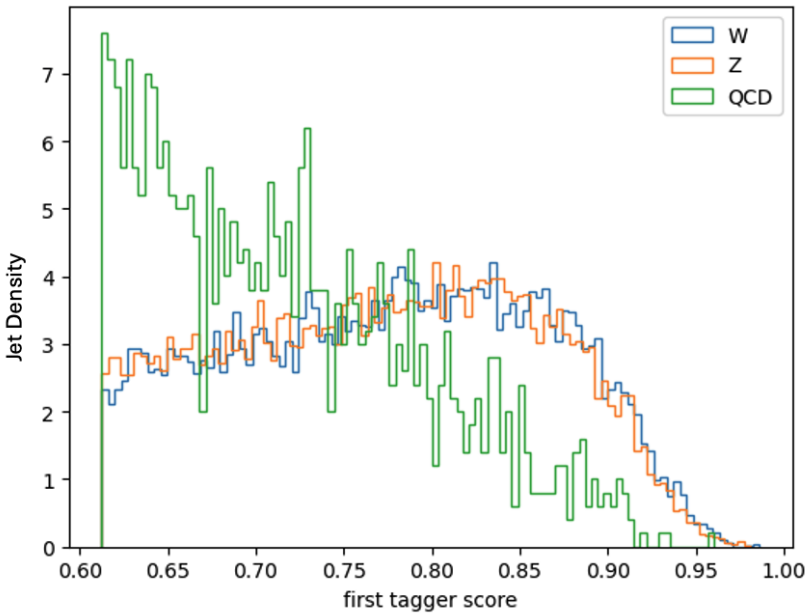


Figure 7 Score distribution from the first-stage (wide-mass W) tagger. W and Z jets show overlapping response, as the tagger was not trained to distinguish between them.

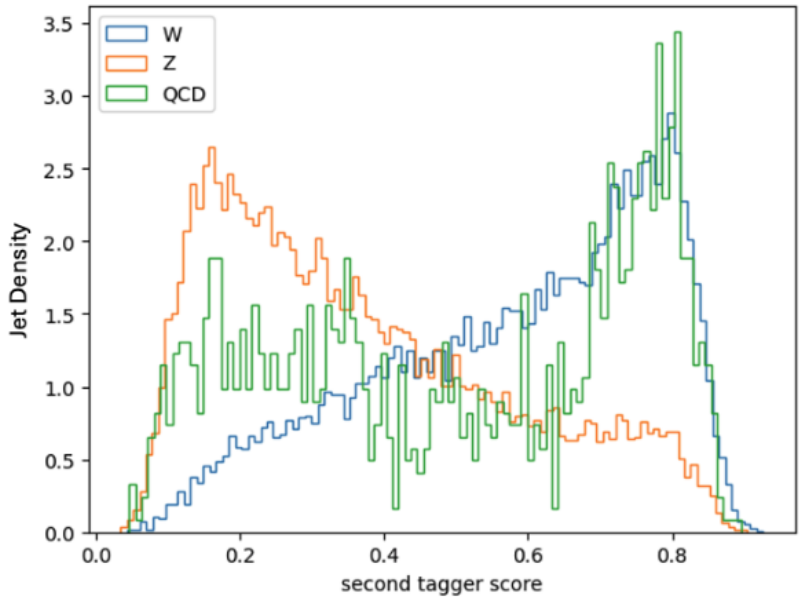


Figure 8 Score distribution from the second-stage W/Z classifier. The network shows improved ability to distinguish Z jets (low scores) from W jets (high scores).

5.2.3 Performance Evaluation

After training, each jet in the test set was assigned a classifier score indicating the probability of being a Z boson. A threshold corresponding to 50% signal efficiency for W jets was selected, and the same threshold was applied to the Z and QCD samples to assess performance uniformly.

Figure 9 shows the ROC curve comparing the true positive rate (W correctly identified) against the false positive rate (Z and QCD jets misidentified as W). The performance of the second-stage classifier is visualized as a smooth curve obtained from scanning over various score thresholds. While the classifier clearly outperforms random guessing (indicated by the dashed diagonal), its separation power remains limited. The operating point at 50% W efficiency lies in a relatively flat region of the curve, with a false positive rate of around 0.3–0.4, indicating that a substantial fraction of Z and QCD jets are still tagged as W.

Complementing this, Figure 10 shows the background rejection (inverse of false positive rate) as a function of signal efficiency. At the 50% operating point, the background rejection reaches just above 10, corresponding to a false positive rate of approximately 0.1. This moderate rejection confirms that the classifier distinguishes some structural differences between W and Z jets, but not decisively — especially given their similar masses and decay topologies.

Both figures include the 88 GeV mass cut as a reference point. This simple cut yields similar, albeit slightly worse, performance compared to the trained classifier, suggesting that while machine learning offers incremental improvement, the overlap between W and Z substructures remains a fundamental bottleneck.

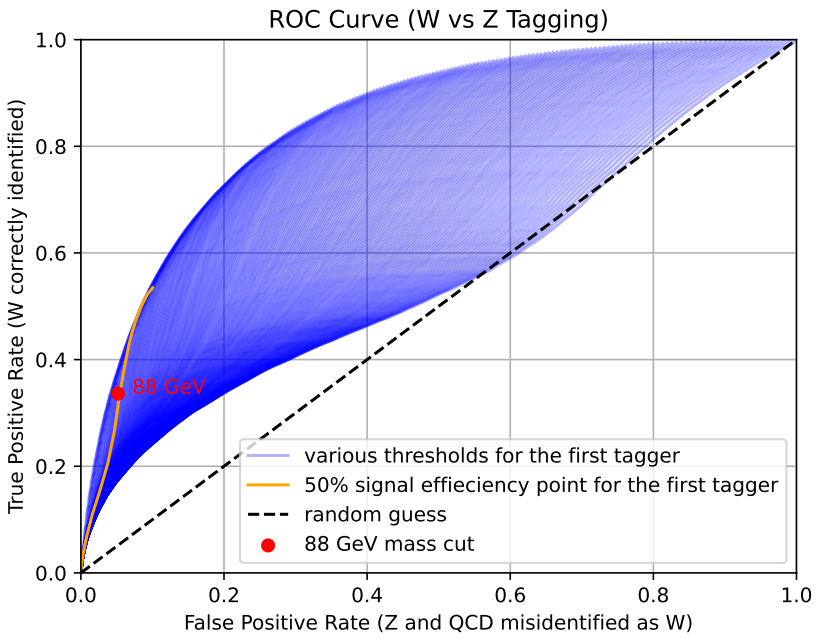


Figure 9 ROC curve for distinguishing W from Z jets. The 50% signal efficiency point and the 88 GeV mass cut are marked for comparison. While better than random guessing, separation remains modest.

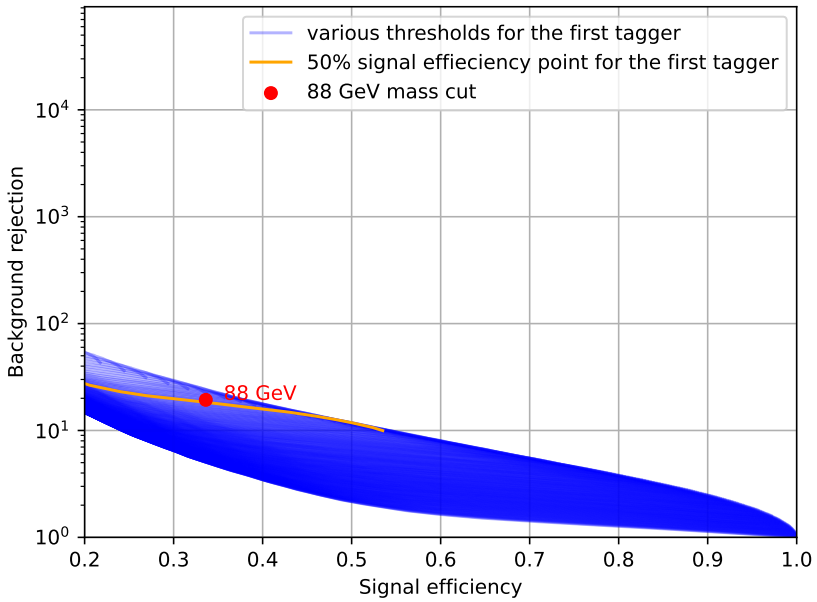


Figure 10 Background rejection versus signal efficiency. The second-stage classifier achieves moderate Z+QCD rejection at 50% W efficiency, slightly outperforming a simple 88 GeV mass cut.

In summary, while the sequential approach (first suppressing QCD and then distinguishing W from Z) does improve classification fidelity for W jets, it struggles to significantly reduce contamination from Z jets. The second-stage LundNet classifier achieves modest gains over simple mass cuts but does not decisively separate the two boson classes. Further improvements may require deeper architectures, alternative features, or a multi-class classification scheme that explicitly models all three jet types simultaneously.

6 Conclusion {AA and EE}

In conclusion, our LundNet-based graph neural network approach has successfully identified W and Z boson jets. By training on a wider mass window, we reduced the artificial peak (mass sculpting) in the background, yet maintained similar rejection power (at a 50% signal efficiency) compared to a narrower mass window of 70–90 GeV, balancing classification performance and preserving the natural mass distribution. Meanwhile, applying a Z-specific classifier within 80–100 GeV provided strong background rejection but induced a noticeable peak around 91 GeV, mirroring the mass sculpting issue observed for the narrow-window W tagger.

When examining a combined sample of W, Z, and QCD jets, we found that simple mass cuts at 64.56 GeV for Z and 99.28 GeV for W gave the highest significance values

of 0.533 and 1.792, respectively, indicating a partial ability to separate these bosons from each other and from the background. On the other hand, while the two-stage classifier retained around half of the W jets as desired, it did not dramatically lower the fraction of Z jets compared to a mass-cut strategy, showing the similarity of W and Z jet substructures.

Our results highlight the advantages of wide-window training. Nonetheless, fully disentangling W and Z remains challenging in adjacent mass ranges, and it may require deeper neural networks or a richer set of substructure variable.

6.1 Future Outlook

Testing these methods in realistic detector simulations (including pileups) will show how robust they are under actual collider conditions and simulating more QCD events would reduce statistical noise and help the model to learn background features better. Additionally a Multi-Class Tagger can be designed for extending this approach to classify multiple particles (e.g., top quarks, the Higgs boson) in a single network, potentially streamlining the tagging process and improving efficiency, while also providing a single framework to handle different systematic uncertainties.

References

1. Achenbach, J. At the heart of all matter: The hunt for the God particle. *National Geographic (USA),* 90–105. https://cds.cern.ch/record/1091244 (2008).
2. The ATLAS Collaboration. *Tagging boosted W bosons applying machine learning to the Lund Jet Plane* tech. rep. (CERN, Geneva, 2023). https://cds.cern.ch/record/2864131.
3. Dreyer, F. A. & Qu, H. *Jet tagging in the Lund plane with graph networks* 2021. arXiv: 2012.08526 [hep-ph]. https://arxiv.org/abs/2012.08526.
4. Kirschenmann, H. Jets at CMS and the determination of their energy scale. *CERN, July,* 60 (2012).
5. Furuichi, A., Lim, S. H. & Nojiri, M. M. Jet classification using high-level features from anatomy of top jets. *Journal of High Energy Physics* **2024,** 1–36 (2024).
6. Dreyer, F. A., Salam, G. P. & Soyez, G. The Lund jet plane. *Journal of High Energy Physics* **2018,** 1–42 (2018).
7. collaboration, A. *et al. Performance of W/Z taggers using UFO jets in ATLAS* tech. rep. (LHC/ATLAS Experiment, 2021).
8. Larkoski, A. J. QCD analysis of the scale-invariance of jets. *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **86,** 054004 (2012).
9. Berger, C. *et al.* Next-to-leading order QCD predictions for W+ 3-jet distributions at hadron colliders. *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **80,** 074036 (2009).
10. Alitti, J. *et al.* A measurement of two-jet decays of the W and Z bosons at the CERN p p collider. *Zeitschrift für Physik C Particles and Fields* **49,** 17–28 (1991).
11. Institute of Physics. *The Standard Model* Accessed: 2025-03-06. https://www.iop.org/explore-physics/big-ideas-physics/standard-model#gref.