A Search for Light Axion-Like Particles: Using $Z \to e^+e^-$ to Validate Transformer-Based Identification of Boosted Diphotons REU Program at Columbia University - Nevis Labs

Emily Centamore[1]

[1]The University of Alabama

August 6, 2026

Nevis Laboratories
Columbia University

Abstract

The conception of the axion as a solution to the strong CP problem gave rise to a new class of theoretical particles known as Axion-Like Particles (ALPs). ALPs share properties with axions, and the search for ALPs is motivated by their potential as dark matter candidates. This report focuses on the search for light ALPs using Run 3 data from the ATLAS detector at the LHC. Specifically, the focus is on the use of a transformer model to identify the merged photon final products produced in the X$\rightarrow$ $aa \rightarrow 4\gamma$ decay. This report introduces the $Z \rightarrow e^+e^-$ decay as a background event used to validate the classification ability of a transformer model across both data and Monte Carlo-simulated samples. Disagreements found in the transformer score across data and simulated samples are found not to be reflected in a raw comparison of the samples, although the disagreements show some correlation with kinematic values. Future trainings for the transformer model using entries re-sampled along p_T and η are expected to yield improved agreement between data and simulation score distributions.

Contents

1 Introduction **2**

1.1 The Standard Model . . . 2

1.2 Instrumentation . . . 3

1.2.1 The Large Hadron Collider . . . 3

1.2.2 The ATLAS Detector . . . 3

1.2.3 The Liquid Argon Calorimeter . . . 4

2 Identification of Boosted Dark Photon Pairs **4**

2.1 CP Symmetry Violation and Axion-Like Particles . . . 4

2.2 Boosted Photons in the Detector . . . 5

2.3 Transformer-Based Classification . . . 6

2.3.1 Cell-Level Information . . . 6

2.3.2 Data-Based Model Evaluation . . . 6

3 Transformer Validation with the $Z \rightarrow e^+e^-$ Background **7**

3.1 $Z \rightarrow e^+e^-$ Sample . . . 7

3.2 Initial Sample Investigation . . . 8

3.2.1 Kinematics . . . 8

3.2.2 Shower Shape Variables . . . 9

3.3 Transformer Performance . . . 10

3.3.1 Variation Across p_T Bins . . . 10

3.3.2 Variation Across Spatial Coordinates . . . 12

4 Summary and Conclusions **13**

5 Acknowledgements **14**

6 References **15**

A Classifier Comparison Table **16**

1 Introduction

1.1 The Standard Model

The Standard Model is the most comprehensive representation of our fundamental understanding of the universe. Finalized in the 1970s, the Standard Model has been tested and confirmed through experimentation and provides a good understanding of much of the physical world [1]. The Standard Model is primarily divided into two types of particles: fermions and bosons.

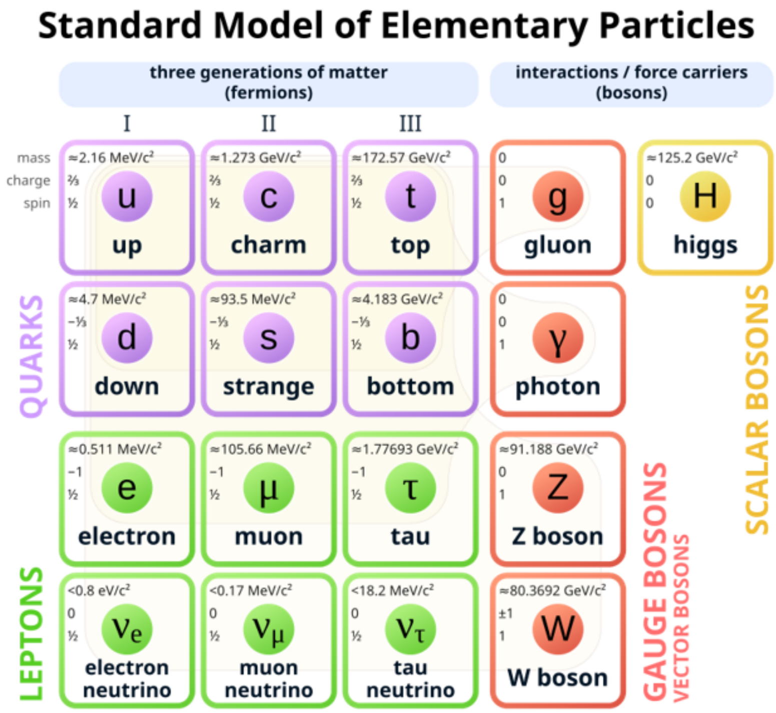


Figure 1: *The Standard Model is made of matter and force-carrier particles called fermions and bosons that define our current understanding of fundamental physics [2].*

Fermions are the particles that make up everyday matter, and they can be further categorized into quarks and leptons. Both quarks and leptons contain three generations of particles that increase in mass and decrease in stability from the lowest to the highest generation. Quark generations each contain two quarks; the up and down quarks belong to the first generation, the charm and strange quarks to the second, and the top and bottom quarks to the third. Up and down quarks are stable, coming together to create common particles like protons and neutrons, while the quarks in higher generations decay quickly and are only found in high-energy environments. Lepton generations refer each to one lepton and its corresponding neutrino. The leptons in order from lowest to highest generation are the electron and electron neutrino, the muon and muon neutrino, and the tau and the tau neutrino.

Bosons are force-mediator particles, and they are classified as either gauge bosons or scalar bosons. Gauge bosons include photons, which mediate the electromagnetic force, gluons, which mediate the strong force, and W± and Z bosons, which mediate the weak force. The only scalar boson in the Standard Model is the Higgs boson, and its namesake Higgs mechanism accounts for the acquisition of particle mass.

Despite its consistency under significant experimental validation, the Standard Model still fails to account for phenomena like neutrino mass, dark matter and dark energy, and the gravitational force. As a result, the search for physics beyond the Standard Model continues to motivate the development of new theories and physics searches, including high energy experiments performed at major particle colliders around the world.

1.2 Instrumentation

1.2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest particle accelerator in the world, built in 2008 by the European Center for Nuclear Research (CERN) to explore physics at the energy frontier as part of the search for physics beyond the Standard Model. The LHC is located 100 meters below ground at the French-Swiss border near Geneva, Switzerland. Roughly 27 kilometers of tunnels house the circular accelerator, which uses vacuum-insulated beam pipes and superconducting electromagnets to accelerate particles in both directions around the loop. Several smaller accelerators operated by CERN feed sequentially into the LHC, allowing particles in the LHC to reach energies of up to 13.6 TeV and travel at speeds of more than 99% of the speed of light [3].

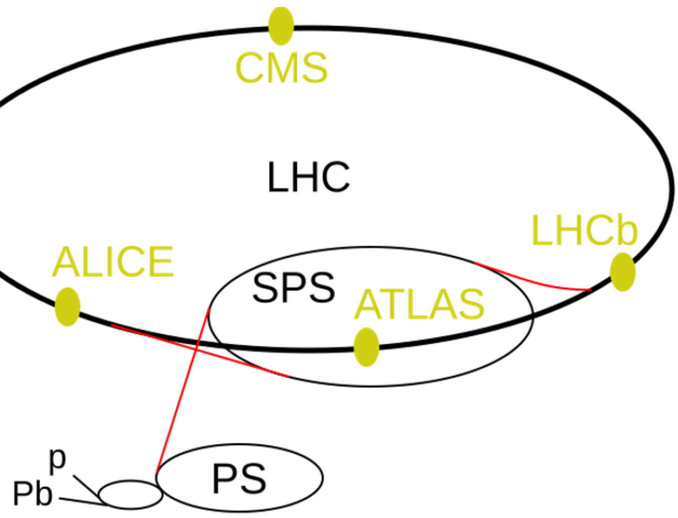


Figure 2: *The Large Hadron Collider accelerates particles to produce high-energy collisions at several sites around its circumference [4].*

There are four major collision points around the circumference of the LHC. At each collision point, one of the LHC's four major detectors - ATLAS, CMS, ALICE, and LHCb - measures the properties of particles produced in the collisions. Information from the collisions in the detectors is filtered, collected, and stored for future use in physics analyses.

1.2.2 The ATLAS Detector

This report focuses on samples collected by A Toroidal LHC ApparatuS (ATLAS) at the LHC. The ATLAS detector is one of the two general purpose detectors at the LHC, meaning that it is used to investigate a wide range of physics phenomena. Three concentric regions in the cylindrical ATLAS detector measure various quantities of particles in the detector, allowing for the reconstruction of collision events. The innermost layers of the detector make up the Inner Detector, which tracks the direction, momentum, and charge of charged particles as they fly outwards. The calorimeter layers lie just outside of the Inner Detector, and they measure the energies of electromagnetic (EM) and hadronic particles as they pass through the detector. The Muon Spectrometer is the outermost region of the detector, and it measures the momenta of heavy muons that frequently pass undetected through the inner layers of the detector.

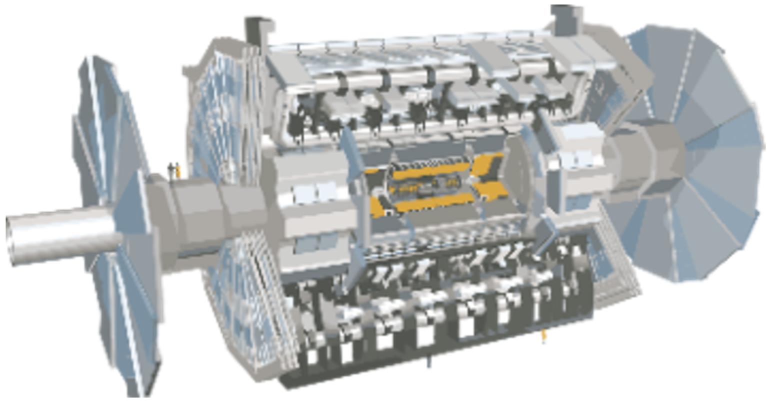


Figure 3: *The three major regions of the ATLAS detector are the Inner Detector, the calorimeters, and the Muon Spectrometer [5].*

1.2.3 The Liquid Argon Calorimeter

ATLAS's EM calorimeter components are integrated into the larger Liquid Argon (LAr) Calorimeter. Layers of metal in the LAr Calorimeter convert incoming particles into showers of lower energy particles, and these lower energy particles ionize the namesake liquid argon that surrounds each metal layer. The ionizations induce electric currents that are measured and combined to calculate the energy of the original particle [5].

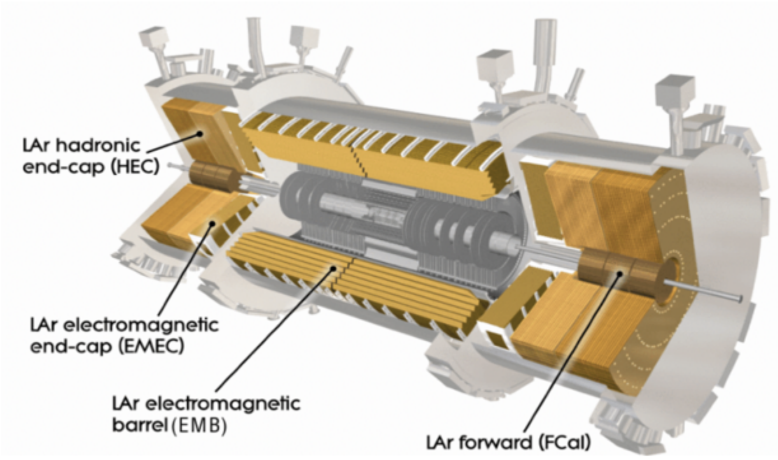


Figure 4: *The LAr calorimeter uses the currents induced by ionized liquid argon to measure the energy of electromagnetic and hadronic particles [6].*

Three concentric sub-layers make up the LAr Calorimeter, with variations in granularity and layer depth differentiating the three. The finely η-segmented inner layer provides higher granularity along the η dimension. The middle layer is the deepest, absorbing the majority of the energy from each particle, and the outermost layer collects any remaining energy resulting from high-energy showers.

2 Identification of Boosted Dark Photon Pairs

2.1 CP Symmetry Violation and Axion-Like Particles

One of the tensions in the Standard Model motivating the search for physics beyond the Standard Model is the strong Charge-Parity (CP) problem. CP violation has been observed in electroweak interactions like kaon and B meson decays [8]. However, despite the expectation that this sym-

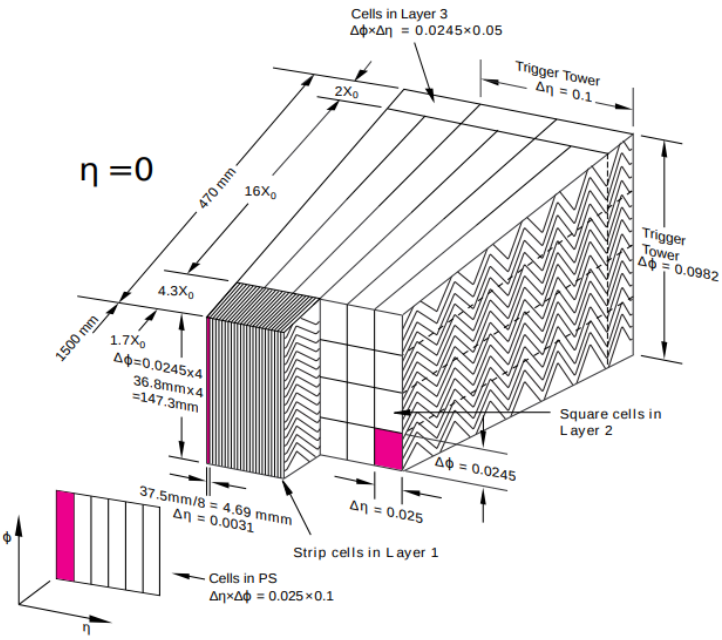


Figure 5: *Three concentric layers within the LAr calorimeter vary in granularity and depth [7].*

metry breaking would also extend to the strong force, no such violation has been recorded in strong force interactions. Further, according to quantum chromodynamics (QCD), any large CP violations by the strong force require an electric dipole moment in the neutron that falls beyond experimentally-determined bounds. A proposed solution to this conflict is the introduction of a new particle to the Standard Model, the axion, whose existence would naturally restrain the likelihood of strong CP violation to near-zero [9]. Efforts to discover the axion, further motivated by the axion's candidacy for dark matter, are currently ongoing.

The theory of the axion has led to the conception of a new class of theoretical particles: Axion-Like Particles, or ALPs. Although ALPs do not themselves solve the strong CP problem, they possess many of the same properties as axions and remain viable candidates for dark matter and evidence of physics beyond the Standard Model. The search discussed in this report focuses on the identification of ALPs produced by a heavy scalar X that decay into pairs of dark photons via the decay X$\rightarrow$ $aa \rightarrow 4\gamma$.

2.2 Boosted Photons in the Detector

The identification of the X$\rightarrow$ $aa \rightarrow 4\gamma$ decay depends heavily on the identification of its final products. Conservation of energy dictates that, while energy may be split between momentum and mass in both the heavy scalar and the lighter ALPs, the energy of each photon must be entirely attributed to momentum. The result is that high-energy, collimated photon pairs become a frequent signature of the X$\rightarrow$ $aa \rightarrow 4\gamma$ decay, with the angular separation $\Delta\mathrm{R}_{\gamma\gamma}$ characterizing the separation of these pairs in the detector. The angular separation of the final state photon pairs depends on the mass $\mathrm{m_a}$ of the ALP and the transverse momentum $\mathrm{p_{T,a}}$ of the ALP, and it is given by the equation

$$\Delta R_{\gamma\gamma} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = \tfrac{2m_a}{p_{T,a}},$$

where η denotes pseudorapidity and ϕ denotes the azimuthal angle. The $\Delta\mathrm{R}_{\gamma\gamma}$ values observed in this study fall between 10^{-4} and 10^{-2}, while the finest granularity in the LAr calorimeter corresponds to a minimum cell radius of $\Delta\mathrm{R} \approx 3 \times 10^{-3}$ [10]. This means that many of the photon pairs produced by this subset of the X$\rightarrow$ $aa \rightarrow 4\gamma$ decay will be separated by a ΔR less than that of one calorimeter cell.

A problem resulting from such strong collimation of signal photons is the similarity of their detector signature to that of background single photons (Fig. 6). Multiple efforts have applied machine learning to this classification task, and one such effort is described here.

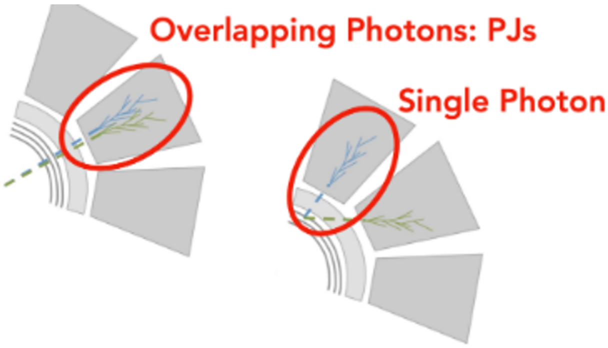


Figure 6: *Merged diphotons and single photons leave nearly identical signatures in the detector [11].*

2.3 Transformer-Based Classification

This study centers around a transformer model developed to identify the merged diphoton products of a light ALP decay [10]. The transformer was trained on a simulated sample of merged photon pairs and single photons. In the simulation, the merged photon pairs are prompt decay products of ALPs produced across a range of masses by a particle gun; the single photons are produced directly by the particle gun. Both single photons and merged photon pairs are measured with a simulated model of the ATLAS detector, and both truth-level and detector-level information is recorded for each event.

2.3.1 Cell-Level Information

Previous architectures have achieved success in the classification of merged photon pairs by training the model on shower shape variable (SSV) information [10]. When an EM particle hits the LAr calorimeter, it creates a multi-cell region of energy deposits, or shower, as visualized in Fig. 7. SSVs characterize these showers using a reduced set of variables, such as lateral shower width or hadronic leakage. SSVs simplify the process of creating classification models, but they also reduce the depth of information available during training relative to full cell-by-cell information.

The transformer discussed in this report is trained using full cell-level information, and its classification performance improves upon the performance of previous models applied to the same task. A comparison of the transformer's performance with existing models is included in Appendix A, and the transformer's score distributions for all signal and background samples is shown in Fig. 8.

2.3.2 Data-Based Model Evaluation

The transformer demonstrates a strong classification ability between single photons and collimated diphotons produced by ALPs in a simulated sample set. However, there are no metrics for how the transformer performs when classifying signal samples in data. Since the goal of the transformer is to confirm or deny the existence of the ALP in data, the existence of any dataset in which ALPs are labeled as such (and could thus be used to gauge the transformer's performance) would defeat the purpose of the model. Therefore, the transformer's classification ability in data must be evaluated by some other method - namely, by examining its performance on background event samples.

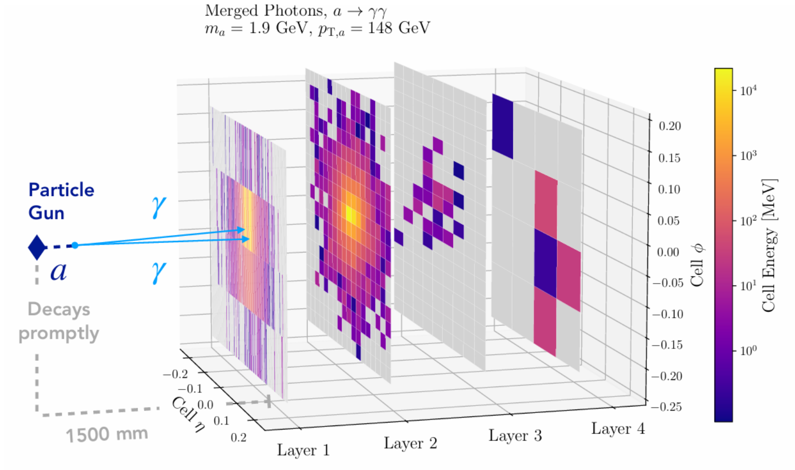


Figure 7: *Particles incident on the surface of the calorimeter leave multi-cell deposits of energy known as showers [10].*

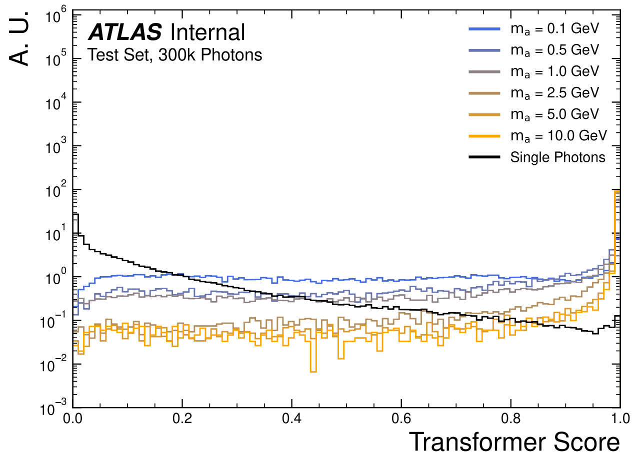


Figure 8: *The score distribution for signal photons peaks near zero, indicating their classification as background-like. The score distributions for merged photon pairs produced by a range of ALP masses all peak near one, indicating their classification as signal-like. [10]*

3 Transformer Validation with the $Z \to e^+e^-$ Background

3.1 $Z \to e^+e^-$ Sample

In this study, the $Z \to e^+e^-$ decay is introduced as a background event to validate the transformer. In an EM calorimeter, both electrons (positrons) and photons leave nearly identical signatures, meaning that for a model restricted to EM information from the LAr calorimeter, the ability to distinguish between single photons and merged photon pairs should be comparable to the ability to distinguish between single electrons and merged photon pairs. The $Z \to e^+e^-$ decay is an ideal candidate for this background sample because it is a precisely-measured Standard Model decay in the ATLAS detector.

The $Z \to e^+e^-$ data sample used in this study was taken in 2022 during Run 3 at the LHC, and it contains 19,478,724 events. The corresponding Monte Carlo-simulated sample used in this study contains 37,320,526 events. There were several cuts applied to both data and simulation to make the sample as pure as possible. The transverse momentum (p_T) of each electron (positron) was required to be greater than 20 GeV, the mass of the parent particle (m_{ee}) was required to be between 80 GeV and 100 GeV, and the electron-positron pairs in each event were required to have a net charge of 0. The sample was also restricted to events occurring within regions of good detector reliability, i.e. $|\eta| < 1.37$ and $1.52 < |\eta| < 2.5$. These cuts are summarized in Table 1.

Value	**Cut**
Transverse Momentum (p_T)	$p_T > 20$ GeV
Z Boson Mass (m_{ee})	80 GeV $< m_{ee} <$ 100 GeV
Pseudorapidity (η)	$|\eta| < 1.37$, $1.52 < |\eta| < 2.5$
Net Electron Pair Charge	Net Charge = 0

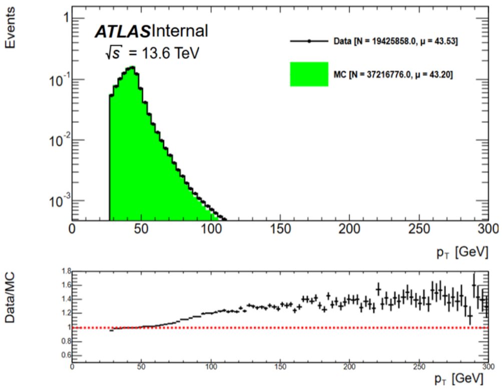
Table 1: *The cuts applied to the $Z \to e^+e^-$ sample help ensure the sample's purity.*

3.2 Initial Sample Investigation

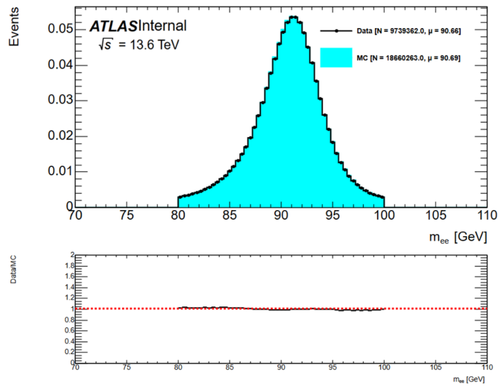
The creation of the simulation sample is done with the goal of mimicking data from the detector as closely as possible. If the simulation is made successfully, then, in the ideal case, the transformer will perform equally on data and simulation samples. The following examination of kinematics and SSVs in the sample compares the simulation to the data before the introduction of the transformer.

3.2.1 Kinematics

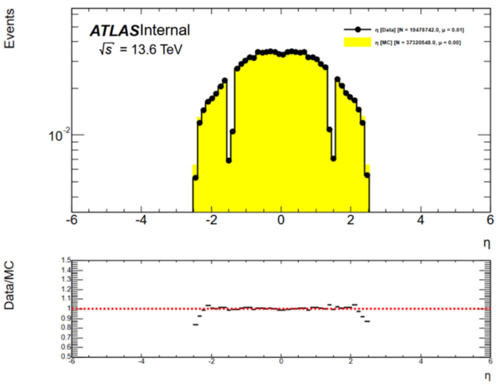
The top panels of the plots in Fig. 9 below show the distributions in the data and simulation samples for p_T, m_{ee}, η, and ϕ.



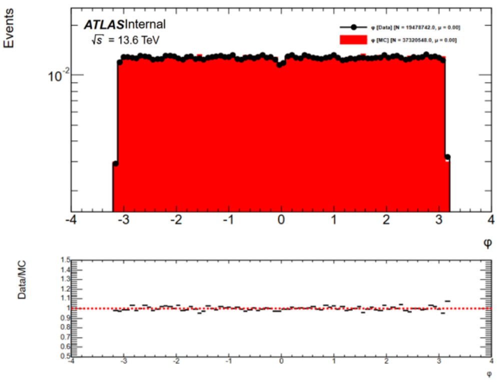
(a) *Transverse Momentum (p_T)*



(b) *Z Boson Mass (m_{ee})*



(c) *Pseudorapidity (η)*



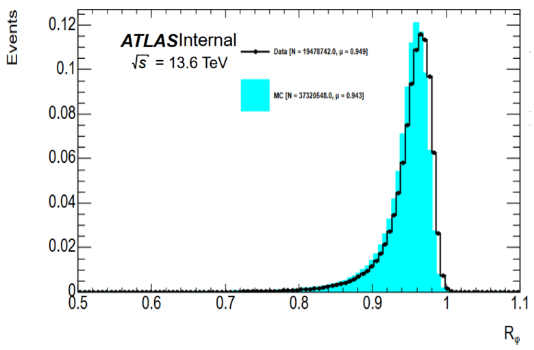
(d) *Azimuthal Angle (ϕ)*

Figure 9: *In each plot, a black line indicates the data distribution, and a filled region represents the simulation distribution. The ratio between data and simulation is shown in each lower panel across the domain of the kinematic distribution. Perfect agreement between data and simulation is indicated in the ratio plots by a dotted red line.*

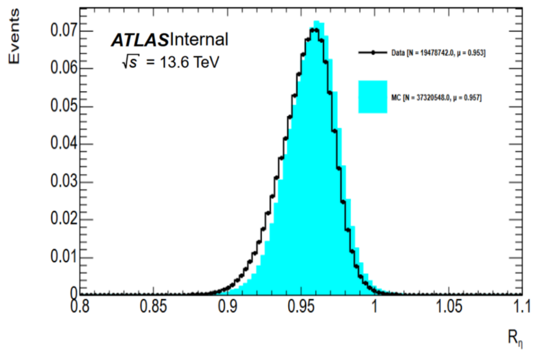
Each of these kinematic graphs demonstrates a strong agreement between data and simulation. The data distribution curves fall along the top edges of the simulation distributions, and each of the ratio plots largely adheres to the ideal agreement line at 1. The only notable exception to this is the agreement between data and simulation in the high p_T range. Here, there is visible disagreement between data and simulation in the ratio plot starting around 75 GeV. This can be attributed to a decrease in the number of events with increasing p_T.

3.2.2 Shower Shape Variables

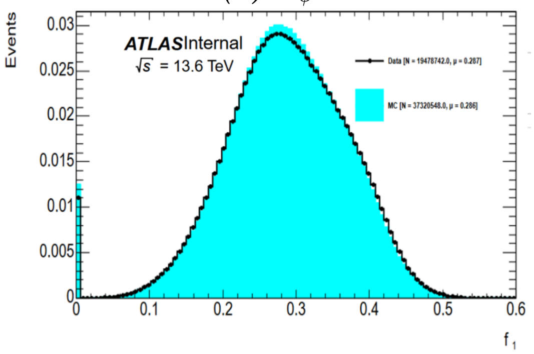
The agreement between the data and simulation samples is similarly examined by looking at the SSV distributions for the sample. A selection of these distributions is shown in Fig. 10 below.



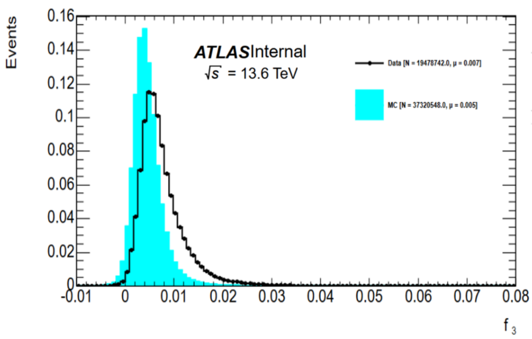
(a) R_ϕ



(b) R_η



(c) f_1



(d) f_3

Figure 10: *Black lines show the SSV distributions in data, and filled regions show the SSV distributions in simulation.*

Each of the SSV distributions shown above represents a ratio of energy values. In Fig. 10a and Fig. 10b, showing R_ϕ and R_η, the ratio represented is the fraction of the total energy spread across a given rectangle of calorimeter cells for which a smaller region of calorimeter cells centered within the larger can also account. In R_ϕ, the expansion from the central rectangle to the larger rectangle occurs along the ϕ dimension; in R_η, it occurs along the η dimension. In Fig. 10c and Fig. 10d, showing f_1 and f_3, the ratio represented is the fraction of total energy that is deposited in the first or third layer of the calorimeter, respectively.

Although the agreements between data and simulation in several of the SSVs are not as exact as those observed in the kinematic distributions, they are sufficient to continue with the study under the assumption of general sample agreement between data and simulation. For all SSVs, the shapes of the data and simulation distributions are comparable, with disagreements largely limited to small offsets along the horizontal axis. The application of small "fudge factors" is therefore likely sufficient to correct the majority of disagreements, and any further investigation on this front is left to future work.

3.3 Transformer Performance

With the assurance of reasonable agreement between the data and simulation samples, the focus of this report now turns to the performance of the transformer on the background sample. Upon examination of the transformer score distribution (Fig. 11), it becomes apparent that the transformer's performance on data samples does not match its performance on simulation samples.

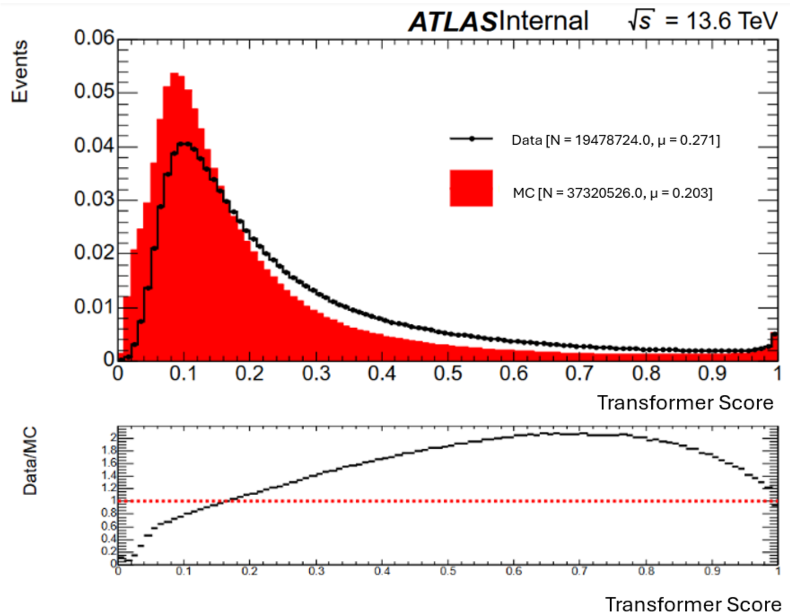


Figure 11: *The transformer score distributions for the $Z \to e^+e^-$ sample differ across data and simulation.*

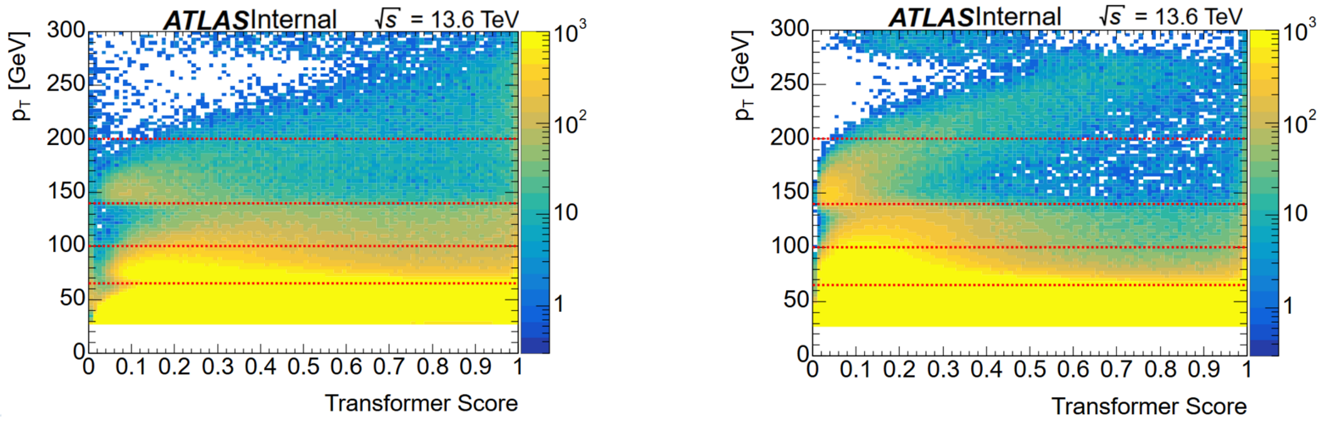
The transformer score ranges from 0 to 1, from most background-like to most signal-like. Both the data score distribution and the simulation score distribution peak near zero, indicating that the transformer score correctly identifies the $Z \to e^+e^-$ decay as a background event. However, the clear difference between the score distribution for data and the score distribution for simulation indicates that the transformer has identified and exploited some feature of the simulation that is not consistent with data. Although no disagreement between the data and simulation samples appeared in the initial comparison across kinematic variables and SSVs, an examination of relationships between the transformer score and physical quantities in the sample may reveal an underlying pattern resulting in the transformer's inconsistency.

3.3.1 Variation Across p_T Bins

One approach to take in exploring the transformer's behavior is to examine the score distributions over distinct regions of single-electron p_T. These p_T regions are derived from the plot of p_T against transformer score (Fig. 12), where the colored contours on the z axis indicate the event count in each bin. A comparison of data and simulation at this stage already reveals significant p_T-correlated disagreement between data and simulation scores, particularly at p_T values above 140 GeV. However, the contours in both plots create distinct p_T regions suitable for binning the transformer score that are consistent between the data and simulation samples.

With these bins defined, the transformer score can be examined across five distinct p_T regions. The plot in Fig. 13 shows this stratification.

Although there are noticeable variations in the distributions across each p_T bin, the majority of the score distributions exhibit a background-like shape consistent with the plot of the overall transformer score. The exception lies in the score distribution for data in the highest p_T bin,



(a) Data *(b) Simulation*

Figure 12: *The bin boundaries are indicated by red dotted lines on the contour plots, and they fall at 65 GeV, 100 GeV, 140 GeV, and 200 GeV.*

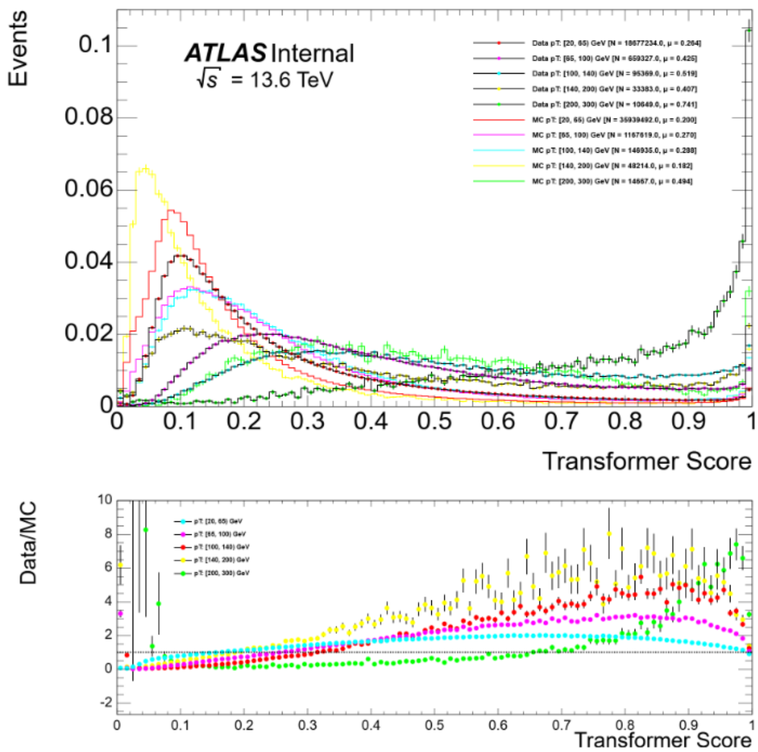


Figure 13: *Each color in the plot corresponds to a different p_T bin. Data distributions are plotted using black lines with colored markers, and simulation distributions are lined in color.*

which covers the 200-300 GeV range. This distribution looks distinctly signal-like, in contrast with even the simulation distribution for the same bin. Although the number of events contained in the highest p_T bin is only $\sim 0.1\%$ of the sample in either data or simulation, making it unlikely to affect the overall transformer score distribution, such a significant error in the transformer's treatment of data in this range may be an indication of a deeper issue in the model.

To ensure that the events in the highest p_T range of the sample are indeed $Z \to e^+e^-$ events, the mass of the parent particle was also binned according to the same p_T bins (Fig. 14). Across all five p_T bins, the m_{ee} distributions remain consistent with the Z boson: both the peaks and the means of each distribution indicate a parent particle mass near that of the Z boson at 91.2 GeV. Therefore, the purity of the $Z \to e^+e^-$ sample is seen to be consistent across all p_T bins.

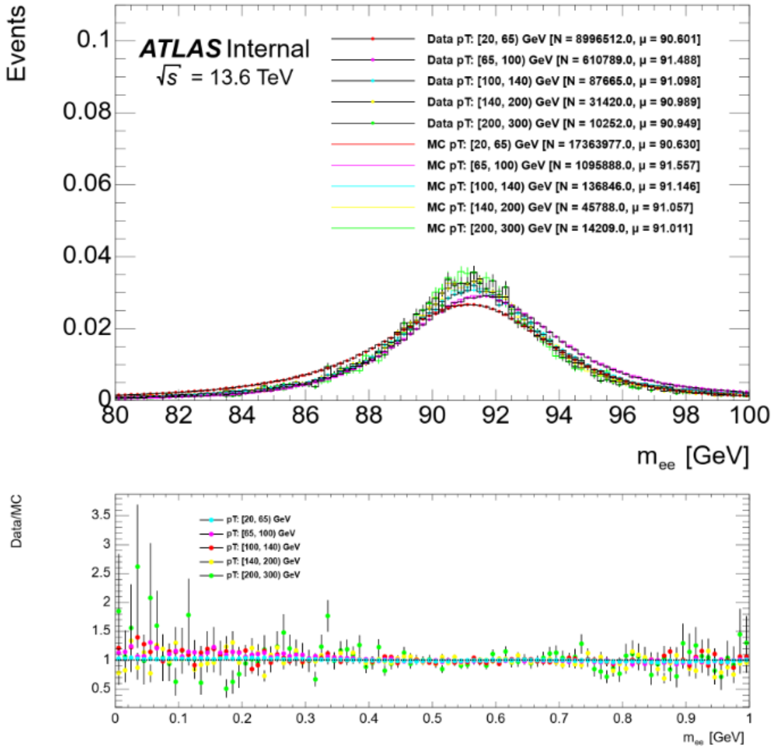


Figure 14: *The p_T binning of the electron parent mass is shown. Data distributions are plotted using black lines with colored markers, and simulation distributions are lined in color.*

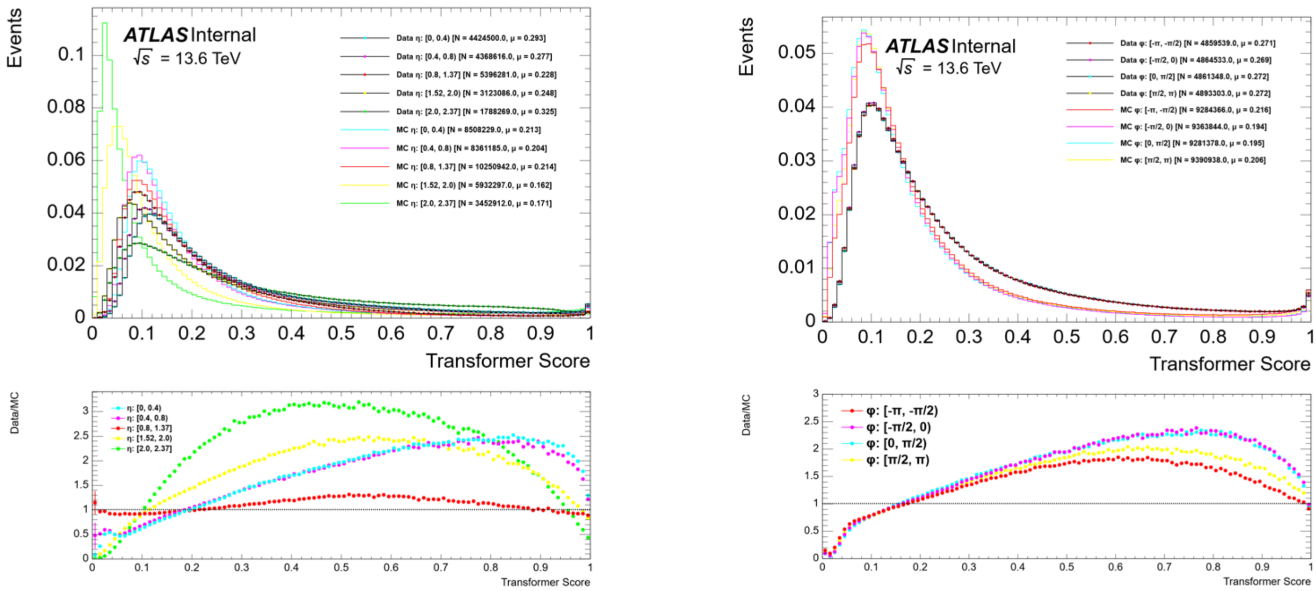
3.3.2 Variation Across Spatial Coordinates

The second approach taken is an examination of the behavior of the transformer score across spatial coordinates. The transformer score is first investigated with respect to η and ϕ individually, and then the two are combined to search the detector space for spatial correlations in the transformer score.

Bins in the η dimension are assigned based on detector geometry. The LAr calorimeter is composed of two major regions: the barrel, at $|\eta| < 1.37$, and the endcap, at $1.52 < |\eta| < 2.5$. The chosen bins adhere to this geometry, with bins over the intervals $|\eta| = [0.0, 0.4), [0.4, 0.8), [0.8, 1.37]$ corresponding to the barrel region and bins over the intervals $|\eta| = [1.52, 2.0), [2.0, 2.37]$ corresponding to the endcap region. The final bin ends at $|\eta| = 2.37$ to ensure that any showers captured near the edge of the region are recorded in good quality ($|\eta| < 2.5$) on all sides.

As with p_T, the transformer score is plotted for each η bin to search for any dependency between the score and the η dimension (Fig. 15a). The η-binning of the transformer score again reveals only moderate variation across the bins. There is some correlation in the level of agreement with the detector geometry; for example, the agreement between data and simulation in the two innermost η bins is very similar, and it is distinct from the level of disagreement observed to be somewhat common in the two outermost η bins. Notably, the agreement between data and simulation in the middle η bin, between $|\eta| = 0.8$ and $|\eta| = 1.37$, is significantly stronger than has been seen in any presentation of the transformer score distribution so far.

The transformer score is further examined across bins in ϕ (Fig. 15b). The choice of these bins is arbitrary; since both the detector and the distribution of events are reasonably invariant across ϕ (see Fig. 9d), this dimension is simply sectioned into intervals of $\frac{\pi}{2}$ ranging from $-\pi$ to π. The transformer score shows little variation across ϕ bins, and the disagreement between the data and simulation score distributions remains evident in all cases.



(a) *η-binned Transformer Score* ***(b)*** *ϕ-binned Transformer Score*

Figure 15: *Each color in the plot corresponds to a different η or ϕ bin. Data distributions are plotted using black lines with colored markers, and simulation distributions are lined in color.*

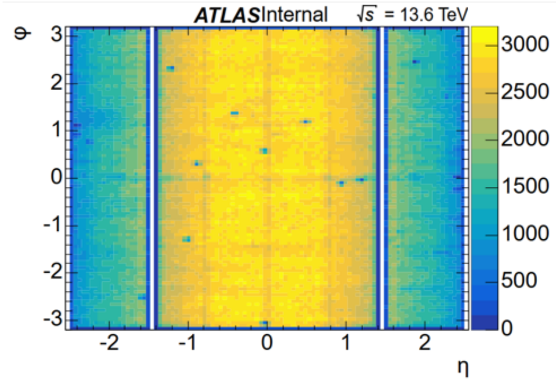
When plotted along perpendicular axes, the η and ϕ dimensions create a plane that represents the cylindrical inner surface of the detector as if it were unrolled along the ϕ dimension. When this plane is shown perpendicular to an axis showing event count (Fig. 16a, 16b), it is clear that the distribution of events in the detector is relatively uniform, and a comparison between data and simulation reveals no significant disagreement. However, choosing the perpendicular axis to be the average transformer score in each bin (Fig. 16c, 16d) allows for an examination of spatial correlation in the transformer score.

The score distributions for the data and simulation samples are relatively comparable in the detector space. Across the majority of both distributions, especially in the barrel region, most bins show an average score of ~ 0.2, which is consistent with a background-like sample. However, there are some key differences between the two distributions. In the data distribution, more signal-like transformer scores are found at both extremes of the $|\eta|$ range, and these score fluctuations are approximately invariant in the ϕ dimension. In data, the lowest average scores are found around the transition region between the barrel and the calorimeter.

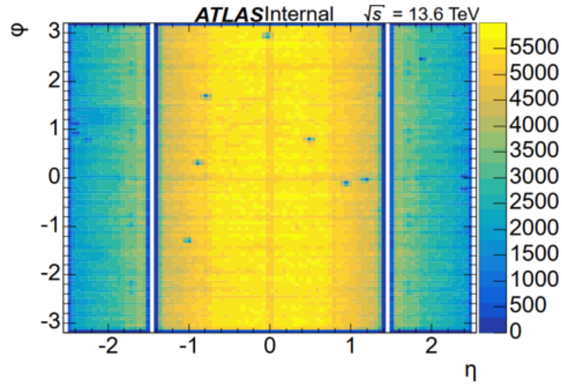
In contrast to the data sample, the simulation sample shows irregular variance across both the ϕ and η dimensions. Patches of signal-like scores are found in both the outer barrel and the endcaps, and where the score in the endcaps is not abnormally high, it is abnormally low. The ϕ-invariant qualities found in the data distribution are faintly mirrored by the simulation distribution in the endcaps near the transition region, but they are not found in the barrel region.

4 Summary and Conclusions

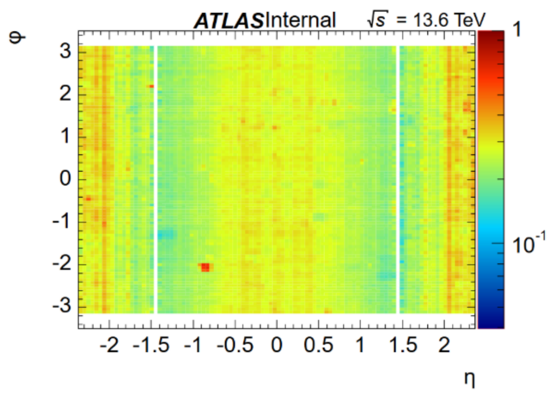
This examination of the $Z \to e^+e^-$ background sample and the transformer's performance across data and simulation sample sets has revealed a significant disagreement in the transformer's treatment of data and simulation. Despite observed agreement between the data and simulation sample sets in kinematics and SSVs, the transformer's score distributions showed a signal-like classification of background data in the high p_T range and spatial variations in transformer score that do not correspond with event distributions in the detector.



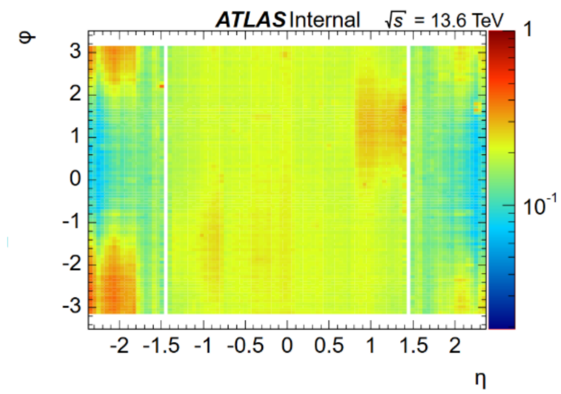
(a) Data Event Distribution



(b) Simulation Event Distribution



(c) Data Score Distribution



(d) Simulation Score Distribution

Figure 16: *Plots of the transformer across η-ϕ space show spatially-correlated variation in the transformer score that is not reflected in the distribution of events in the detector.*

The next step is a re-training of the transformer. The model used in this project results only from the initial training of the architecture, and significant improvements stand to be gained by an updated training sample set. In the training sample, signal and background samples did not have equal quantities of events across the p_T range or the η range. A re-sampling of events is currently in progress to reduce this difference across the full range of both values. Re-training the model on this adjusted sample set will help ensure that the transformer learns only physical distinctions between signal and background rather than differences tied to event count and therefore unique to the training sample.

5 Acknowledgements

A special thank you goes to Professor John Parsons, Dr. Lauren Osojnak, Dr. Jonathan Long, and Gabriel Matos for amazing mentorship and support throughout this project. Thank you also to Professor Georgia Karagiorgi, Professor Reshmi Mukherjee, and Amy Garwood for all of the hard work that made this REU possible. Finally, thank you to my fellow REU students and everyone at Nevis Labs for making this summer so much fun!

This material is based upon work supported by the National Science Foundation under Grant No. PHY-2349438.

6 References

[1] The Standard Model. CERN. (n.d.). https://home.cern/science/physics/standard-model/

[2] Wikimedia Foundation. (2026, August 3). Standard model. Wikipedia. https://en.wikipedia.org/wiki/StandardModel

[3] The Large Hadron Collider. CERN. (n.d.). https://home.cern/science/accelerators/large-hadron-collider/

[4] Wikimedia Foundation. (2026, August 4). Large Hadron Collider. Wikipedia. https://en.wikipedia.org/wiki/LargeHadronCollider

[5] Detector and Technology. ATLAS Experiment at CERN. (n.d.). https://atlas.cern/Discover/Detector

[6] The ATLAS Collaboration., Aad, G., Abbott, B. et al. Readiness of the ATLAS liquid argon calorimeter for LHC collisions. Eur. Phys. J. C 70, 723–753 (2010). https://doi.org/10.1140/epjc/s10052-010-1354-y

[7] ATLAS LAr Phase-II Upgrade TDR. CERN Document Server. (2018, June 15). https://cds.cern.ch/record/2285582/files/ATLAS-TDR-027.pdf

[8] CP Violation. Wikipedia. (2026, July 17). https://en.wikipedia.org/wiki/CP violation

[9] Axion. Wikipedia. (2026, July 20). https://en.wikipedia.org/wiki/Axion

[10] Transformer-based machine learning using low-level calorimeter signals for collimated photon identification at collider experiments. arXiV. (2026, July 9). arXiv:2607.08175 [hep-ph]

[11] Search for short- and long-lived axion-like particles in H$\rightarrow$aa$\rightarrow$4$\gamma\gamma$ decays with the ATLAS experiment at the LHC. arXiv. (2024, Dec 17). arXiv:2312.03306 [hep-ex]

A Classifier Comparison Table

		Background Rejection ($1/\epsilon_{\mathrm{bkg}}$)		
Model	Overall AUC	90% ϵ_{sig}	95% ϵ_{sig}	99% ϵ_{sig}
SSV-Level				
BDT	0.92	3.2	1.8	1.2
DNN	0.92	3.1	1.8	1.2
Cell-Level				
CNN	0.95	6.4	2.6	1.3
PFN	0.97	62 ± 7	5.4 ± 0.2	1.48 ± 0.01
Transformer	**0.98**	$\mathbf{472 \pm 60}$	$\mathbf{9.0 \pm 0.4}$	$\mathbf{1.64 \pm 0.01}$
MLP Mixer	**0.93**	**4.3**	**2.1**	**1.2**

Figure 17: *The transformer model improves upon other existing classifier models through the use of cell-level information during training [10].*