AXION-LIKE PARTICLE SEARCH USING MACHINE LEARNING FOR THE SIGNAL SENSITIVITY OPTIMIZATION WITH RUN 2 LHC DATA RECORDED BY THE ATLAS EXPERIMENT*

Ondřej Matoušek

Czech Technical University in Prague, Prague, Czech Republic

Received 14 September 2025, accepted 10 October 2025, published online 19 December 2025

The neutral Standard Model Higgs boson was discovered in 2012 at CERN, and the search for further particles of extended models continues, in particular, the search for an axion-like particle (ALP). Using machine learning technologies, this analysis addresses the separation of ALP production from unwanted background reactions. In this project, the Run 2 data from the ATLAS detector are used and the efficiency as well as the significance of the machine learning algorithm are optimized as a function of the theoretical ALP mass.

DOI:10.5506/APhysPolBSupp.18.6-A24

1. Introduction

Axion-like particles (ALPs) are heavy particles predicted by some extensions of the Standard Model (SM). In this analysis, the production of ALPs that couple to two photons is studied [1–3]. Such a process can be identified as a resonant peak at the ALP invariant-mass value in the $\gamma\gamma \to \gamma\gamma$ interaction channel, which is known as light-by-light scattering (LbL). Apart from ALP production, the LbL process can also occur through an intermediate fermion or W boson. LbL has been measured at the LHC in nucleus–nucleus collisions using lead-ion beams (Pb–Pb collisions) [4–7]. Here, the LbL cross section was enhanced due to the high nuclear charge. These analyses searched for ALPs with the theoretical ALP invariant mass up to 100 GeV. At higher invariant diphoton masses, the effective luminosity of $\gamma\gamma$ scattering in pp collisions surpasses that of Pb–Pb collisions [8], even though the scattering cross section is smaller at lower invariant masses. In the ATLAS experiment, the diphoton signature can be detected in the

* Presented at the V4-HEP 1 — Theory and Experiment in High Energy Physics Workshop, Bratislava, Slovakia, 26–28 July, 2023.

central detector around the pp beam interaction point. As the interacting protons travel close to their original beam direction, the ATLAS forward-proton detectors (AFPs) can be used for their tagging [2]. The production of lepton pairs by $\gamma\gamma$ interactions, $pp \to (\gamma\gamma \to \ell^+\ell^-)p^{(*)}$, has been measured by ATLAS and CMS using AFPs [9, 10]. Here, $p^{(*)}$ denotes a dissociating proton. This analysis serves as a revision to the original ALP search in pp interactions in ATLAS [11], by incorporating Artificial Neural Networks (ANNs) into the selection process [12]. Using the ATLAS detector, pp collisions are studied in the target invariant mass range of 150–1600 GeV for 13 discrete values with varying step-size. In total, three possibilities for the reaction are considered. As depicted in figure 1, in the exclusive process, both protons stay intact. In the single- and double-dissociative processes, one or both protons dissociate while radiating a virtual photon. It is possible to tag the undissociated proton. The dissociated proton, however, is in practice unmeasurable. This search uses 14.6 fb^{-1} of 13 TeV pp collision data and requires at least one tagged proton. Building on the previous search [13, 14], the focus is placed on the single-dissociative process. The reason behind this is that in the double-dissociative and exclusive processes, the sensitivity has been saturated, and the use of ANNs would not increase it.

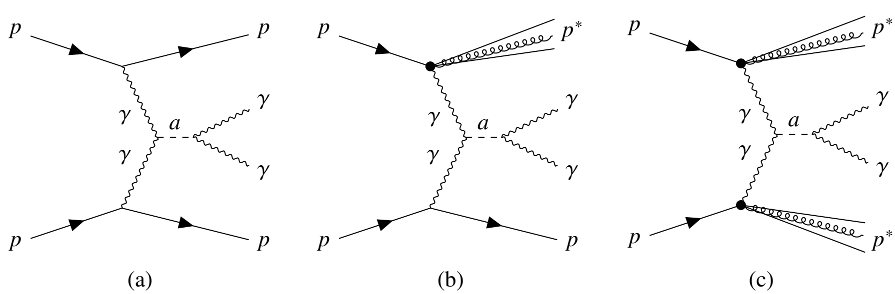

Fig. 1. Feynman diagrams for (a) exclusive, (b) single-dissociative, and (c) double-dissociative light-by-light scattering with outgoing photon pairs mediated by an ALP denoted by a [13].

2. Event samples and preselection

In the ATLAS detector, the forward protons are detected in the AFP spectrometer system [15, 16]. The AFP system consists of four tracking modules located at $z = \pm 205$ m and $z = \pm 217$ m, where z is the direction of the proton beam. The $\pm$ denotes the two locations of the AFP systems, denoted as the A(C) side for the $+z(-z)$ direction. Each module houses a silicon tracker consisting of four planes of silicon pixel sensors. A two-level trigger system was used in the initial event detection. The first-level trigger is hardware-implemented with a sampling rate below 100 kHz. The second-level trigger is software-based, and it has a reduced average sampling rate of 1 kHz. As AFP systems reconstruct proton tracks with 70% probability per bunch crossing, no AFP triggers were used in this analysis [17]. For this search, the background dataset used was collected in 2017 using pp collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV with an integrated luminosity of 14.6 fb^{-1}. A diphoton trigger implemented by two clusters of EM calorimeter cells inside the central ATLAS detector was used with transverse energies above 35 GeV and 25 GeV, respectively [18, 19], after which standard data-quality requirements were applied [20]. For an event to be measured, additionally, the AFP modules are required to have at least three operational silicon planes [21]. As in the previous analysis, the sensitivity to exclusive and double-dissociative events was already saturated to maximum, this search only covers and updates the search for SD ALP signatures.

The simulated SD signal events were produced using the SuperChic 4.14 Monte Carlo (MC) generator [3, 22–24]. The simulated ALP events cover ALP invariant masses in the range of $m_X = 150$–1600 GeV. For each invariant mass, samples were generated with the ALP-to-diphoton coupling constant set to $f^{-1} = 0.05$ TeV^{-1} and natural width of the ALP being $\Gamma = \frac{m_X^3}{4\pi f^2}$. Generator-level preselections are applied to the dataset, requiring at least two photon candidates both to have transverse momentum $p_{\mathrm{T}} > 40$ GeV and the pseudorapidity $|\eta| < 2.37$ [13]. An additional preselection requirement is for the azimuthal misalignment between the pair of photons to be very small, as determined by the acoplanarity requirement $A_\phi^{\gamma\gamma} := 1 - \frac{|\Delta\phi_{\gamma\gamma}|}{\pi} < 0.01$.

3. Main ANN selection

In the previous search [13], the signal events were determined on the basis of the kinematics of the central photon pair. The fractional energy loss of the scattered proton was computed from the central detector and from the AFP stations. Due to the law of conservation of energy, it is required that both fractional energy losses be almost the same. The idea behind this search is to replace the main selection with an Artificial Neural Network that serves as a binary classifier. In practice, the use of an ANN involves several general steps. First, the input data has to be normalized. For that purpose, Z-score normalization was used [25]. Then, an ANN serving as a binary classifier is trained. In this search, an augmented version of a shallow multilayer perceptron (MLP) was used [12]. The training steps are performed by an optimizer. In this analysis, the Adam optimizer was used [26] (improved stochastic gradient descent). During training and validation, a total of 581216 signal and 68380 background events have been used with a conservative split of 80/20% for training/validation. To account for the class imbalance, a weighted cross-entropy loss function was used [27]. Additionally, in order to reduce the Internal Covariance Shift, the batch normalization (BN) technique was added to the model through specialized BN layers [28]. As the ANN MLP model was shallow, the use of residual connections to mitigate the vanishing gradient problem was not necessary [29]. A final optimization was performed to enhance the more difficult events of a lower invariant mass. In total, eight times more weight was applied during the training on the 150 GeV ALP samples in comparison to all other weights, which led to the model being superior to the cut-based preselection method on every invariant mass point.

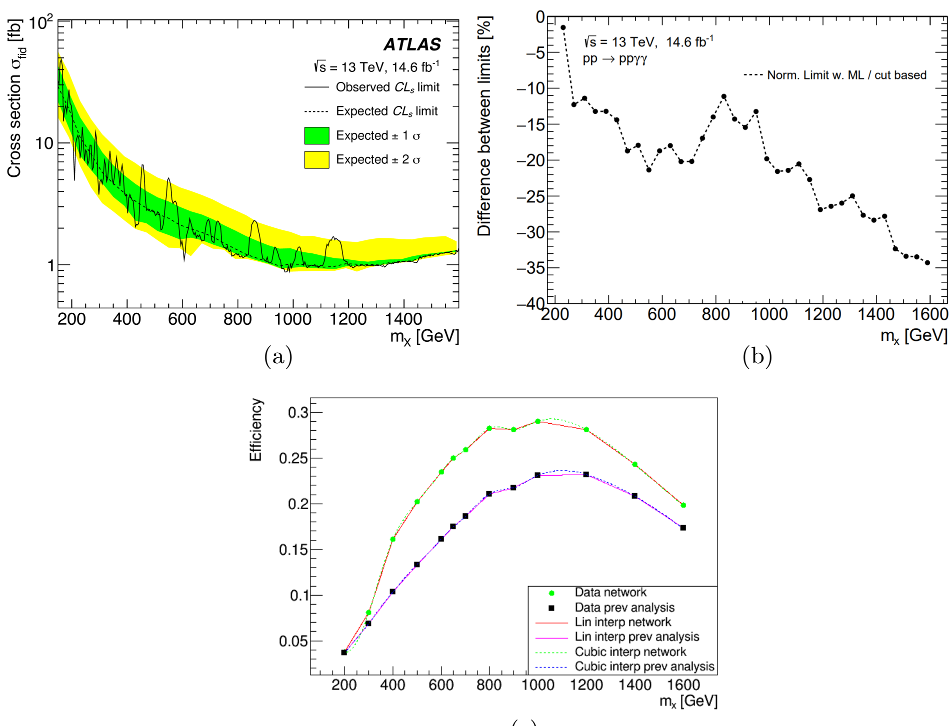

Fig. 2. (a) Expected and observed 95% C.L. upper limits on the signal cross section using the cut-based selection, assuming 100% branching ratio for ALP decay into two photons, as functions of the hypothetical ALP invariant mass m_X. (b) Normalized ratio of the expected 95% C.L. upper limits on the signal cross section obtained using the ANN method relative to those from the cut-based selection [13, 31]. (c) Signal selection efficiency as a function of ALP invariant mass m_X for the SD process [13, 31].

4. Results

During the final prediction of the ANN signal/background, the network predicts normalized probabilities [30]. These are then thresholded to produce Boolean decision variables. This threshold is selected so that the number of signal events detected is the same as in the previous analysis — 441. This ensures that the comparison is fair as we now directly study the efficiency of the selection process. For the original pp analysis, the expected, as well as observed limits at the 95% C.L. are shown in figure 2 (a). The normalized comparison between the expected limits of both approaches in % is visualized in figure 2 (b). Furthermore, for additional comparison, the signal selection efficiencies for the two approaches are shown in figure 2 (c). Further details are given in reference [31].

5. Conclusions

A search for an axion-like particle (ALP) was performed with the ATLAS experiment using 14.6 fb^{-1} of $\sqrt{s} = 13$ TeV proton–proton collision data. This analysis focused on the single-dissociative light-by-light scattering process, $pp \to p(\gamma\gamma \to \gamma\gamma)p^{(*)}$, where an ALP resonance would appear as a peak in the invariant mass spectrum. An artificial neural network (ANN) was employed as the primary selection tool in place of traditional cut-based kinematic selection.

The ANN-based analysis achieved an improved signal selection efficiency across the considered invariant mass range of $m_X = [150, 1600]$ GeV, and resulted in stronger expected exclusion limits on the ALP expected cross section at the 95% confidence level. The selection efficiencies were higher in comparison to the cut-based approach. Compared to the previous analysis, the use of machine learning increased the sensitivity of the search. These results demonstrate the potential of ANN-based methods to enhance the sensitivity of New Physics searches at the LHC. Such methods can be applied in future searches.

REFERENCES

[1] S. Knapen, T. Lin, H.K. Lou, T. Melia, *Phys. Rev. Lett.* **118**, 171801 (2017).

[2] C. Baldenegro, S. Fichet, G. von Gersdorff, C. Royon, *J. High Energy Phys.* **2018**, 131 (2018).

[3] L.A. Harland-Lang, M. Tasevsky, *Phys. Rev. D* **107**, 033001 (2023).

[4] ATLAS Collaboration, *Nature Phys.* **13**, 852 (2017).

[5] CMS Collaboration, *Phys. Lett. B* **797**, 134826 (2019).

[6] ATLAS Collaboration (G. Aad *et al.*), *Phys. Rev. Lett.* **123**, 052001 (2019).

[7] ATLAS Collaboration (G. Aad *et al.*), *J. High Energy Phys.* **2021**, 243 (2021).

[8] R. Bruce *et al.*, *J. Phys. G: Nucl. Part. Phys.* **47**, 060501 (2020).

[9] ATLAS Collaboration (G. Aad *et al.*), *Phys. Rev. Lett.* **125**, 261801 (2020).

[10] CMS Collaboration (A.M. Sirunyan *et al.*), *J. High Energy Phys.* **2018**, 153 (2018).

[11] ATLAS Collaboration (G. Tateno), *Nuovo Cim. B* **47**, 8 (2024).

[12] D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Nature* **323**, 533 (1986).

[13] ATLAS Collaboration (G. Aad *et al.*), *J. High Energy Phys.* **2023**, 234 (2023).

[14] G. Tateno, Ph.D. Thesis, CERN, 2023, CERN-THESIS-2023-006.

[15] ATLAS Collaboration, 2015, CERN-LHCC-2015-009.

[16] ATLAS Collaboration (B. Olaizola *et al.*), *Phys. Rev. C* **94**, 054318 (2016).

[17] G. Tateno, 2023, ATL-PHYS-PROC-2023-051.

[18] ATLAS Collaboration, 2018, ATL-DAQ-PUB-2018-002.

[19] ATLAS Collaboration (G. Aad *et al.*), *Eur. Phys. J. C* **80**, 47 (2020).

[20] ATLAS Collaboration (G. Aad *et al.*), *J. Instrum.* **15**, P04003 (2020).

[21] M. Kocian, *J. Instrum.* **12**, C01077 (2017).

[22] L.A. Harland-Lang, V.A. Khoze, M.G. Ryskin, *Eur. Phys. J. C* **76**, 9 (2016).

[23] L.A. Harland-Lang, V.A. Khoze, M.G. Ryskin, *Eur. Phys. J. C* **79**, 39 (2019).

[24] L.A. Harland-Lang, M. Tasevsky, V.A. Khoze, M.G. Ryskin, *Eur. Phys. J. C* **80**, 925 (2020).

[25] G. James, D. Witten, T. Hastie, R. Tibshirani, «An Introduction to Statistical Learning», 1st edition, *Springer*, New York 2013.

[26] D.P. Kingma, J. Ba, `arXiv:1412.6980 [cs.LG]`.

[27] T.M. Cover, J.A. Thomas, «Elements of Information Theory», *Wiley-Interscience*, 1999.

[28] S. Ioffe, C. Szegedy, `arXiv:1502.03167 [cs.LG]`.

[29] K. He, X. Zhang, S. Ren, J. Sun, in: «Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)», Las Vegas, NV, USA, 27–30 June, 2016, pp. 770–778.

[30] V.N. Vapnik, «Statistical Learning Theory», *Wiley*, 1999.

[31] O. Matousek, Bachelor Thesis, Czech Technical University in Prague, CERN-THESIS-2023-075, 2023