PROCEEDINGS OF SCIENCE

Anomaly Detection in CMS

Amandeep Kaur on behalf of the CMS Collaboration*
Department of Physics and Astronomy, Purdue University,
525 Northwestern Avenue, West Lafayette, IN, 47907
E-mail: amandeep.kaur@cern.ch

With current advancements in computational resources and algorithmic developments, the CMS experiment at the LHC has been incorporating machine learning (ML) techniques to further enhance its physics potential. While ML offers powerful computational tools, the foundational building blocks remain rooted in the underlying physics phenomena. These advancements have significantly improved the search for new physics, allowing physicists to conduct more effective searches and measurements while enabling innovative approaches to data analysis. In these proceedings, we present our recent advancements in ML techniques applied to anomaly detection within the CMS experiment, focusing on both dijet final states and enhancements at the Level-1 (L1) trigger. Our work includes novel methods for identifying anomalous jet substructures, enhancing the discovery potential of new physics signatures that were previously unexplored. Additionally, we discuss the implementation of ML for anomaly detection at the L1 trigger, underscoring its potential to improve the early selection of interesting events. Our findings illustrate the efficacy of these approaches in maximizing sensitivity to rare events, contributing to the ongoing efforts to unravel the mysteries of the universe.

12th Large Hadron Collider Physics Conference (LHCP2024)
3-7 June 2024
Boston, USA

*Speaker

PoS(LHCP2024)017

1. Introduction

The search for new physics beyond the Standard Model (SM) remains a central focus of high-energy particle physics. With the Large Hadron Collider (LHC) at CERN providing unprecedented collision energy and luminosity, researchers are continuously developing innovative methodologies to probe unexplored territories of particle interactions. One promising avenue of investigation is the search for dijet resonances, particularly in the context of anomalous jet substructure, which can serve as a sensitive indicator of new physics phenomena.

Machine learning (ML) has been a vital tool in this field for decades, with applications evolving from early neural networks used for charged particle tracking in the 1980s [2] to recent advances that contributed to the discovery of the Higgs boson in 2012 [3]. Today, ML not only enhances data analysis but also plays a crucial role in data quality monitoring within the experiment. Building on this history, our work benefits from the advancements in computational resources and machine learning (ML) techniques to enhance the physics potential of the CMS experiment [1]. While ML methods provide powerful computational tools for data analysis, they remain grounded in the fundamental physics phenomena underlying particle collisions. This duality is crucial, as it ensures that the search strategies employed remain robust and relevant to the complexities of the underlying processes.

In these proceedings, we present a model-agnostic search strategy for dijet resonances, utilizing a range of multivariate ML algorithms to identify jets exhibiting anomalous substructure. By placing minimal assumptions on signal hypotheses, we aim to maximize sensitivity to a wide variety of potential new physics signatures. Our analysis is based on data collected by the CMS experiment during proton-proton collisions at a center-of-mass energy of 13 TeV, encompassing an integrated luminosity of 138 fb^{-1}. This paper also explores the application of ML techniques for anomaly detection at the Level-1 (L1) trigger [14], enhancing the ability to identify interesting events in real-time.

2. Dijet resonance anomaly search

The dijet resonance anomaly search focuses on identifying new physics by searching for narrow-width heavy resonance with TeV-scale mass decaying into two other resonances, which decay hadronically with high Lorentz boost, such that their decay products are contained in two large-radius jets. This search is designed to be model-agnostic, meaning it does not assume a specific signal hypothesis beyond the presence of a narrow dijet resonance with a mass between 1.8 and 6.0 TeV. By minimizing assumptions about the new physics signals, the search maximizes its sensitivity to a wide range of potential signatures, including those that have not been considered in previous analyses [4–8].

To explore possible new physics signal, we rely on proton-proton collision data collected by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The goal is to identify any deviations from SM predictions in the dijet mass spectrum, particularly in events where the jet substructure appears anomalous. A key challenge in designing such a search is the need to balance sensitivity to potential new physics while suppressing the overwhelming Quantum Chromodynamics (QCD) background, which dominates dijet production at the LHC. Signal models with A→BC topology are used in this search which are categorized based on the number of hard prongs in the B and C jets such as 1+2, 2+2, 3+3, 2+4, 5+5, and 6+6 prong topology. We consider benchmark models of narrow resonances, such as excited quarks and heavy vector bosons.

The event selection criteria are carefully designed to isolate potential new physics signals while reducing the contribution of QCD background. Events are required to have at least two jets with $p_{\mathrm{T}} > 300\,\mathrm{GeV}$, with the leading-$p_{\mathrm{T}}$ jets satisfying invariant mass m_{jj} >1455 GeV threshold to ensure sensitivity to high-mass resonances. Since the resonant particle A is produced via the s-channel, the t-channel dominated QCD background is reduced relative to the signal by requiring the two jets to have a rapidity difference of $|\Delta\eta_{jj}| < 1.3$. The dijet invariant mass is a key observable, and events are selected within a mass window spanning from 1.8 to 6.0 TeV. In addition to kinematic selections, jet substructure variables such as the jet mass and N-subjettiness are utilized to distinguish potential signal jets from those produced by QCD processes. These substructure variables are crucial for identifying jets with anomalous internal structure, which could indicate the presence of new physics.

A range of multivariate machine learning methods is employed in the analysis to further refine the event selection and enhance the sensitivity to signals with non-standard jet substructure. These techniques allow for a more detailed examination of the internal structure of jets, enabling the identification of anomalies that may not be apparent through traditional cut-based methods. By combining information about jet kinematics, substructure, and the dijet invariant mass, we can more effectively isolate events that could correspond to new physics signatures.

3. Anomaly detection methods

Several advanced anomaly detection techniques are devised to identify potential new physics decaying to jets with anomalous substructure. The Variational Autoencoder (VAE), an unsupervised learning method, utilizes the kinematic information such as the x, y, and z component of momentum of each particle to learn the behavior of regular jets and detect outliers. In weakly-supervised training methods, Classification Without Labels (CWoLa), alongside Tag N' Train (TNT), uses features such as the softdrop mass (m_{SD}) [9], n-subjettiness variables (τ_{21}, τ_{32}, τ_{43}) [10], the number of PF candidates inside the jet, lepton subjet fraction [11], and the maximum b-tagging score from the DeepCSV algorithm [12] to classify signal versus background on data without explicit labels in order to scan over the full dijet mass spectrum. Classifying anomalies through outer density estimation (CATHODE), which also focuses on a signal-rich dijet mass window, uses a different approach for the background sample. It first learns the conditional probability density of background events as a function of invariant mass using a normalizing flow generative model. This model is trained on events outside the signal window, then interpolated to generate synthetic background events for the signal region. A weakly supervised classifier is trained to distinguish data events in the signal window from these synthetic background samples. Lastly, Quasi Anomalous Knowledge (QUAK) encodes prior knowledge of potential new physics signals by utilizing features like the ratio of the jet mass and transverse momentum and modified substructure variables, making it the most model-dependent approach in the analysis.

4. Results and interpretation

Following the validation in simulation, the anomaly detection methods are applied to the data in the signal region. The fitting procedure is performed on the dijet invariant mass spectra after a selection from each of the anomaly detection method and is shown in Fig. 1(left) for VAE method. No significant excess is found over smoothly falling QCD background [13]. The largest excesses seen by the CATHODE, CATHODE-b, QUAK, and VAE-QR methods has local significance of 2.2, 2.9, 2.6, and 2.3 standard deviations (σ) at resonance masses of 2.3, 2.3, 4.7, and 4.9 TeV respectively. The TNT and CWoLa Hunting methods did not report any excesses larger than 1.5σ.

Having seen no significant excesses in the data, studies were performed to evaluate the sensitivity of the search procedure on a subset of the signal models. These benchmark signal models cover several different combinations of the B and C particle substructure. The discovery sensitivity of each method for the benchmark signals was then estimated. For each method, the signal cross sections that would have led to an expected 3σ and 5σ excess were determined. The results, shown in Fig. 1(right), indicate that the anomaly detection methods outperform the traditional cut-based selections for all the benchmark signals considered, significantly improving upon the discovery sensitivities of the inclusive search. The largest improvements are seen for the 6+6 pronged signal, where anomaly detection methods reduce the cross section needed for a 5σ discovery by a factor of 7 (3.3) compared to the inclusive (three-pronged) selection.

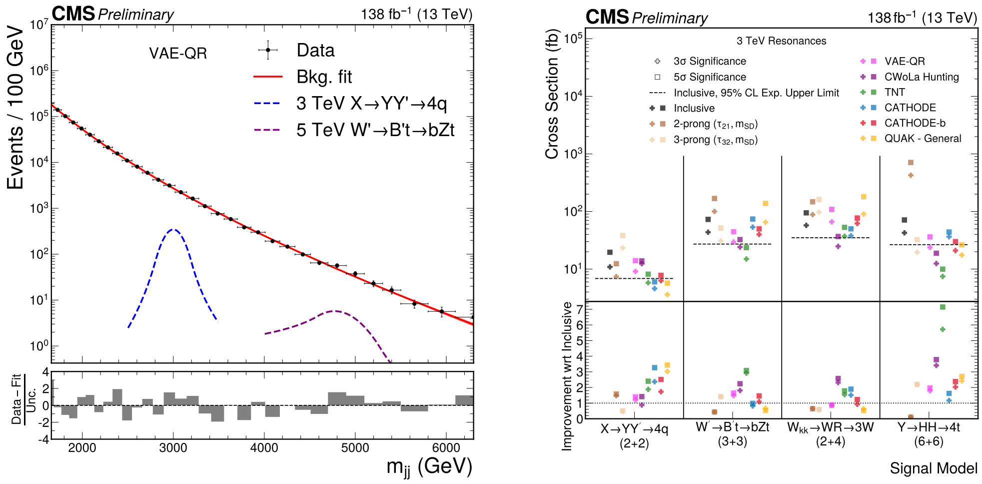

Figure 1: The dijet invariant mass spectrum and resulting background fit to the data for VAE-QR (left), The discovery sensitivity for the process A → BC, using the anomaly detection methods, and a comparison to sensitivity of the inclusive search. In all signal processes, the mass of the heavy resonance was set to m(A) = 3 TeV (right).

5. Anomaly detection in real-time

The Level-1 (L1) trigger plays a crucial role in high-energy physics experiments by rapidly deciding which events should be recorded for further analysis, based on muon-system and calorimeter data. However, the traditional L1 trigger relies on predefined criteria that may miss rare or unexpected SM and Beyond Standard Model (BSM) events. To address this limitation, new algorithms based on anomaly detection techniques are developed to implement directly at the L1 trigger level, offering an unbiased approach with the potential to enhance sensitivity to exotic or rare physics that might otherwise evade conventional triggers.

One of the major innovations in this field is the development of AXOL1TL [16], an anomaly detection algorithm specifically designed for the CMS L1 global trigger. AXOL1TL is based on a variational autoencoder (VAE) trained directly on zero-bias data events using a range of kinematic features from up to 10 jets, 4 muons, 4 electrons, and the missing transverse momentum (MET) in each event. The unsupervised nature of this VAE allows it to detect unusual event topologies without relying on explicit assumptions about the signal model. Once trained, the encoder is compiled into firmware using the hls4ml package [15] and deployed on Field Programmable Gate Arrays (FPGAs), allowing it to compute anomaly scores in real time. The algorithm operates within the strict timing constraints of the L1 trigger, making decisions within hundreds of nanoseconds. The anomaly score, derived from the latent representation of the events, determines whether an event exhibits characteristics that deviate from typical SM background processes. Early results from the 2024 CMS data collection indicate that AXOL1TL is capable of selecting unique, anomalous events that are missed by existing L1 triggers. Specifically, the anomaly score distribution shows a strong ability to highlight events with complex kinematic features. However, it was observed that AXOL1TL exhibits a preference for high-multiplicity events, which introduces some sensitivity to pileup. Mitigating this pileup dependence is an area of active development.

In addition to AXOL1TL, another powerful anomaly detection algorithm known as CICADA (Calorimeter Image Convolutional Anomaly Detection Algorithm) [17] has been implemented at the L1 trigger level. CICADA is designed to use fast, unsupervised machine learning techniques to identify anomalous events based on calorimeter data. The algorithm treats calorimeter information as image-like data, using convolutional neural networks (CNNs) to process the event topologies and identify patterns that deviate from expected background behavior. CICADA offers a complementary approach to AXOL1TL by focusing on calorimeter-based event features, allowing it to capture a different range of anomalous signals. Its ability to run in real-time within the L1 trigger system makes it a promising tool for enhancing event selection efficiency, particularly in searches for rare or unexpected phenomena.

The integration of anomaly detection algorithms like AXOL1TL and CICADA into the L1 trigger system represents a significant step forward in the use of machine learning to enhance the physics potential of the CMS experiment. By applying these algorithms directly at the trigger level, the experiment gains the ability to capture rare, unusual events that might otherwise be missed, improving sensitivity to both SM and BSM processes. Moreover, these algorithms are trained directly on data, allowing them to remain unbiased and adaptable to a wide range of potential new physics signatures. Their deployment on FPGAs ensures that the real-time constraints of the L1 trigger are met, making these techniques both powerful and efficient.

6. Summary

In these proceedings, we present a model-agnostic search for new physics in the dijet final state, utilizing advanced anomaly detection methods to identify events with anomalous jet substructure. The analysis, based on 138 fb^{-1} of data collected by the CMS experiment at the LHC, explores a range of resonance masses from 1.8 to 6.0 TeV. Using unsupervised, weakly-supervised, and semi-supervised machine learning algorithms, no significant excesses over the Standard Model background are observed, allowing us to place stringent exclusion limits on various benchmark signal models. Additionally, we have discussed the novel application of anomaly detection techniques at the Level-1 trigger, using the AXOL1TL and CICADA algorithms to enhance the sensitivity of the CMS trigger system to rare or unexpected events. These developments underscore the growing role of machine learning in pushing the boundaries of high-energy physics, enabling more efficient event selection and opening new possibilities for discovering both Standard Model and Beyond Standard Model phenomena.

References

[1] CMS Collaboration,"The CMS Experiment at the CERN LHC," JINST **3**, S08004 (2008) doi:10.1088/1748-0221/3/08/S08004

[2] B. H. Denby,"Neural Networks and Cellular Automata in Experimental High-energy Physics," Comput. Phys. Commun. **49**, 429-448 (1988) doi:10.1016/0010-4655(88)90004-5.

[3] CMS Collaboration,"Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," Phys. Lett. B **716**, 30-61 (2012) doi:10.1016/j.physletb.2012.08.021.

[4] CMS Collaboration, "A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at $\sqrt{s}$ = 13 TeV," Eur. Phys. J. C **80**, no.3, 237 (2020) doi:10.1140/epjc/s10052-020-7773-5.

[5] CMS Collaboration, "Search for new heavy resonances decaying to WW, WZ, ZZ, WH, or ZH boson pairs in the all-jets final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV," Phys. Lett. B **844**, 137813 (2023) doi:10.1016/j.physletb.2023.137813.

[6] CMS Collaboration, "Search for a W' boson decaying to a vector-like quark and a top or bottom quark in the all-jets final state at $\sqrt{s}$ = 13 TeV," JHEP **09**, 088 (2022) doi:10.1007/JHEP09(2022)088.

[7] CMS Collaboration, "Search for resonances decaying to three W bosons in proton-proton collisions at $\sqrt{s}$ = 13 TeV," Phys. Rev. Lett. **129**, no.2, 021802 (2022) doi:10.1103/PhysRevLett.129.021802.

[8] CMS Collaboration, "Search for resonances decaying to three W bosons in the hadronic final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV," Phys. Rev. D **106**, no.1, 012002 (2022) doi:10.1103/PhysRevD.106.012002.

[9] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, "Soft Drop," JHEP **05**, 146 (2014) doi:10.1007/JHEP05(2014)146.

[10] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness," JHEP **03**, 015 (2011) doi:10.1007/JHEP03(2011)015.

PoS(LHCP2024)017

[11] C. Brust, P. Maksimovic, A. Sady, P. Saraswat, M. T. Walters and Y. Xin, "Identifying boosted new physics with non-isolated leptons," JHEP **04**, 079 (2015) doi:10.1007/JHEP04(2015)079.

[12] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at $\sqrt{s}$ = 13 TeV," JINST **13**, no.05, P05011 (2018) doi:10.1088/1748-0221/13/05/P05011.

[13] CMS Collaboration, "Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s}$ = 13 TeV," CMS-PAS-EXO-22-026.

[14] CMS Collaboration, "The Phase-2 Upgrade of the CMS Level-1 Trigger," CERN-LHCC-2020-004.

[15] J. Duarte, S. Han, P. Harris, S. Jindariani, E. Kreinar, B. Kreis, J. Ngadiuba, M. Pierini, R. Rivera and N. Tran, *et al.* "Fast inference of deep neural networks in FPGAs for particle physics," JINST **13**, no.07, P07027 (2018) doi:10.1088/1748-0221/13/07/P07027.

[16] CMS Collaboration, "Testing a Neural Network for Anomaly Detection in the CMS Global Trigger Test Crate during Run 3," JINST **19**, no.03, C03029 (2024) doi:10.1088/1748-0221/19/03/C03029.

[17] CMS Collaboration, "Level-1 Trigger Calorimeter Image Convolutional Anomaly Detection Algorithm," CMS Detector Performance Note, CMS-DP-2023-086, 2023.