ATLAS CONF Note

ATLAS-CONF-2023-022

May 19, 2023

Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

Searches for new resonances in two-body invariant mass distributions are performed using an unsupervised anomaly detection technique in events produced in pp collisions at a center-of-mass energy of 13 TeV recorded by the ATLAS detector at the LHC. Studies are conducted in data containing at least one isolated lepton. An autoencoder network is trained with 1% randomly selected collision events and anomalous regions are then defined which contain events with high reconstruction losses from the decoder. Nine invariant mass distributions are inspected which contain pairs of one light jet (or one b-jet) and one lepton (e, μ), photon, or a second light jet (b-jet). No significant deviation from the background-only hypothesis is observed after applying the event-based anomaly detection technique. The 95% confidence level upper limits on contributions from generic Gaussian signals are reported for the studied invariant mass distributions. The widths of the signals range between 0% and 15% of the resonance mass and masses range from 0.3 TeV to 7 TeV. The obtained model-independent limits are shown to have a strong potential to exclude generic heavy states with complex decays.

ATLAS-CONF-2023-022
19 May 2023

Searches for new physics phenomena beyond those described by the Standard Model (SM) require advanced techniques to devise selections that involve a large number of variables characterizing collision events. Furthermore, limited understanding of how new physics manifests itself has been an inspiration for designing model-independent searches [1]. In traditional methods, event selections are optimized to target specific signal signatures beyond the SM (BSM) and to maximize their separation from SM background processes. Alternatively, event selection criteria can be relaxed to target wider signatures, but this reduces the ability to suppress background.

Machine learning (ML) anomaly-detection methods [2–12] open a new way to study the collision events. One such approach uses an autoencoder (AE) [13–16], a neural network architecture that is commonly used with unsupervised learning. The AE is trained using predominantly SM background events and is applied to identify collision events that display kinematic properties different from those of SM events. Previously, ATLAS used a weakly supervised learning technique for massive dijet final states [17] and unsupervised machine learning method to identify anomalous jets to search for BSM resonances decaying into a Higgs boson and a generic new boson [18].

This note presents a generic search for resonances in various two-body final states using an anomaly detection method on the event topology. Events are triggered by an isolated lepton to reduce contamination from QCD multi-jet events. The two-body final states consist of jet+Y, where the jet can be a light jet or a b-jet while Y can be a lepton (electron or muon), a photon, or another jet or b-jet. The highest-p_T candidate for each kind of final-state object is used, except for the case of final states with two light jets (b-jets), where both the leading and subleading light jets (b-jets) are considered. Triggering on the lepton allows to study invariant mass distributions lower than 1 TeV, which would be otherwise difficult to model due to trigger threshold effects if the jet triggers were used [19].

The data corresponds to 140 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV collected by the ATLAS experiment at the Large Hadron Collider (LHC). The ATLAS detector [20] is a multipurpose particle detector with cylindrical geometry.[1] It consists of an inner tracking detector (ID) surrounded by a superconducting solenoid, sampling electromagnetic (EM) and hadronic calorimeters, and a muon spectrometer with three toroidal superconducting magnets, providing a near 4π coverage in solid angle. A two-level trigger system is used to select events for storage. Events used in this analysis were selected online by single-electron or single-muon triggers [21–23]. An extensive software suite [24] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

The interaction vertex with the highest sum of the squared transverse momenta (p_{T}^2) of all associated tracks with $p_{\text{T}} > 500$ MeV is selected as the primary vertex [25]. Electrons are reconstructed from energy clusters in the EM calorimeter that match a reconstructed track [26]. Muons are reconstructed by combining a track from the ID and one from the muon spectrometer [27]. Electrons (muons) must have $E_{\text{T}} > 20$ GeV ($p_{\text{T}} > 20$ GeV) and pseudorapidity $|\eta| < 2.47$ ($|\eta| < 2.5$), excluding the region of $1.37 < |\eta| < 1.52$ for electrons. To ensure that selected leptons originate from the primary vertex, their tracks must have $|d_0/\sigma(d_0)| < 5(3)$ for electrons (muons) and $|z_0/\sin(\theta)| < 0.5$ mm for both lepton flavors. Here d_0 and $\sigma(d_0)$ are the transverse impact parameter and its uncertainty, and z_0 is the longitudinal impact parameter along the beam line. Electrons must pass the "Tight" likelihood-based identification criterion defined in

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector and the z-axis along the beam pipe. The x-axis points to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z–axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. The distance between two objects in η–ϕ space is $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Transverse momentum is defined as $p_{\text{T}} = p \cdot \sin\theta$.

Ref. [26], and the "FCTight" isolation requirement defined in Ref. [22]. Muons must pass the "Medium" cut-based identification criterion and the particle-flow-based "PflowTight" isolation requirement [27]. Photons are reconstructed from energy deposits in the central EM calorimeter [28]. They must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and pseudorapidity $|\eta| < 2.37$, excluding photons within $1.37 < |\eta| < 1.52$. The "tight" identification requirement and the "tight" isolation requirement are applied, both defined in Ref. [28]. Jets are reconstructed using the anti-k_t algorithm [29, 30] with a radius parameter of 0.4, applied to tracks in the ID and topological clusters [31] processed using a particle-flow algorithm [32]. To reduce the effect from additional collisions in the same and nearby bunch crossings (pileup) [33], jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ must pass a jet-vertex-tagger requirement that corresponds to a selection efficiency for hard scattered jets of 96% [34]. Jets containing a b-hadron (b-jets) are identified using the 77% efficiency working point of the DL1r b-tagging algorithm [35, 36]. Jets that fail the b-tagging criteria are identified as light jets. An overlap-removal procedure detailed in Ref. [37] is applied to the selected objects.

The missing transverse momentum, with magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$, is calculated as the negative of the vector sum of the transverse momenta of all reconstructed objects associated with the primary vertex. To account for the soft hadronic activity, a term including tracks associated with the primary vertex, but not with any of the reconstructed objects, is included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation [38].

Following the aforementioned object reconstruction, events are further selected by requiring at least one lepton (e or μ, denoted as ℓ) with $p_{\mathrm{T}}^{\ell} > 60\,\mathrm{GeV}$, and at least one jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. A trigger matching requirement [21] is applied in which the reconstructed lepton must lie within the vicinity of the corresponding trigger-level object. Jets and b-jets are required to have a pseudorapidity of $|\eta| < 2.4$. This step is denoted as *preselection* hereinafter. Nine invariant mass distributions are studied in this analysis. The invariant mass distributions of m_{jj} and m_{bb} are reconstructed from the leading and sub-leading (b-)jets in each event; m_{jb} is reconstructed from the leading jet and leading b-jet; m_{je} (m_{be}), $m_{j\mu}$ ($m_{b\mu}$), and $m_{j\gamma}$ ($m_{b\gamma}$) are reconstructed from the leading (b-)jet and leading electron, muon, or photon, respectively. An event that contains multiple types of objects can contribute to multiple mass distributions.

Several Monte Carlo (MC) samples are simulated to verify the analysis procedure, although the anomaly detection only relies on data samples. All SM samples, which are used to establish the background modeling, are simulated using a detailed ATLAS detector simulation [39] based on GEANT4 [40]. Samples for the benchmark BSM signal processes are simulated using the ATLAS fast simulation framework [39]. Additional pp collisions generated using PYTHIA with the A3 set of tuned parameters [41] are overlaid to simulate the effects of pileup that matches the multiplicity of additional collisions in data.

Kinematic features of the final state objects in the preselected events are structured in a matrix called rapidity-mass matrix (RMM) which is proposed as an input for machine learning [42]. The RMM was tested for an anomaly detection method using MC event generators and demonstrated to produce a more robust AE training than using four-momenta as an input [43]. The RMM is a square matrix employed to represent a comprehensive picture of the event based on $E_{\mathrm{T}}^{\mathrm{miss}}$, transverse energies, transverse masses, Lorentz factors, two-particle invariant masses, and two-particle rapidity differences. In this analysis, the final-state objects are reconstructed light jets, b-jets, muons, electrons, and photons. The maximum number of light jets or b-jets considered is 10, while the maximum number of electrons, muons, and photons considered is 5. The objects within each type are ordered by their transverse energy in descending order. If the number of available objects for a particular type is less than the maximum limit, the remaining elements in the corresponding row and column of the RMM are filled with zeros. This ensures that the input size is the same for every event.

By construction, all elements of the RMM are defined between 0 and 1, and most variables are Lorentz invariant under boosts along the longitudinal axis. To reduce biases on the shapes of the jet+Y invariant mass spectra, the nine invariant mass variables are excluded from the RMM. The resulting input dimension is $36^2 - 9 = 1287$. The RMM matrix is then flattened to a 1-dimensional input vector before being fed into the AE.

The AE is implemented using `TensorFlow` [44]. It is comprised of two sections, an encoder and a decoder. The encoder compresses the input to a latent dimensional space, whereas the decoder takes the data in the latent layer and decompresses it back to its original size. The network architecture for the encoder contains two hidden layers, with 800 and 400 neurons respectively, and a latent layer of 200 neurons. The decoder reverses the structure of the encoder, namely 400 and 800 neurons for the two hidden layers, with 1287 neurons for the output layer. The Leaky ReLU [45] activation function is applied to the output in all hidden and output layers. Artificial anomalous events with different characteristics were created and used as input to test the separation power of this architecture. The default AE has shown to outperform (convolutional) variational AEs. Furthermore, the selected architecture has shown better separation between anomalous and non-anomalous events than AE architectures with other numbers of neurons. To form the training and validation datasets, 1% collision events are randomly selected after the preselection. They offer both sufficient statistics and a good representation of typical collision events. The chance to find BSM signal events in these datasets is considered to be negligible. Even if the training dataset is influenced by BSM physics, such contributions would alter the distribution of the reconstruction loss of the AE and thus the shape of the mass spectra but would not produce discernible bumps [43]. Therefore, the search for localized enhancements in the invariant mass distributions should not be significantly affected.

The training and validation sets are randomly split using a 7:3 ratio. The optimizer Adam [46] is used to train the network by minimizing the reconstruction loss of the training set, which is defined as the mean squared error between the input and reconstructed values of the training set. The batch size is 100 and all events in the training set are reshuffled in the beginning of every epoch. Training of the AE is monitored on the reconstruction loss of the validation set, and terminated if this value does not decrease within 30 epochs. To avoid selecting the over-trained model, the training is repeated 50 times with different network initialization and different random splits of the training and validation sets. The performance is found to be stable and the model that gives median validation loss among these trials is used. The logarithm of the reconstruction loss, log(Loss), is defined as the anomaly score for each event.

The distribution of the anomaly scores of the collision data is shown in Fig. 1. A peak is shown near log(Loss) = −10.8 from the total of 166 055 597 events after preselection. As a comparison, several benchmark BSM models, studied in Ref. [47, 48], are overlaid. These representative models are characterized by the presence of an isolated lepton in the final states. A list of the models and their parameters is the following: (1) charged Higgs production in association with a top quark, tbH^+, with the mass of H^+ between 0.4 TeV and 2 TeV with a varying step size [49]; (2) a Kaluza–Klein gauge boson, W_{kk}, with the SM W boson and a radion that decays into two gluons, with the mass of W_{kk} running from 0.5 TeV to 6 TeV while the mass difference of W_{kk} and the radion being 0.25 TeV [50]; (3) a model with composite $\mathrm{SU(2)_L}$ fermion doublets that breaks lepton-flavor universality ("composite lepton"), with the mass of the parent Z' boson between 0.5–4 TeV, and various mass hypotheses of the composite lepton [51]; (4) the sequential standard model (SSM) $W' \to WZ' \to \ell\nu q\bar{q}$, with the mass of the W' boson ranging from 0.7 TeV to 6.2 TeV with a varying step size and the mass difference of the W' and Z' bosons being 0.25 TeV [52]; (5) a simplified dark-matter (DM) model, $Z' \to q\bar{q}$ with an axial-vector mediator Z' boson whose mass ranges between 0.5–6 TeV with a varying step size, where one of the quarks radiates a W boson which decays to $\ell\nu$ [53]. As can be seen in Fig. 1, the anomaly scores of the BSM processes tend to be larger than those of the collision

events, which are predominantly produced by the SM processes. The SSM and DM models tend to have similar characteristics as the SM background under the event selection of this analysis, therefore yielding lower anomaly scores. Although only one hypothetical mass is shown for each type of BSM models, it was found that events with larger hypothetical masses lead to larger anomaly scores. Furthermore, three anomaly regions (ARs) are chosen to maintain sensitivity to different BSM models. They are defined by $\log(\text{Loss}) > -9.1$, > -8.0, and > -6.5, respectively, as indicated by the vertical lines in Fig. 1. The labels of the three ARs indicate the visible cross section of hypothetical processes yielding a number of events equal to what observed in the data for an integrated luminosity of 140 fb^{-1}. It was checked that the anomaly score distributions in each data-taking year are consistent, indicating that the AE training is robust against different pileup conditions and triggering criteria.

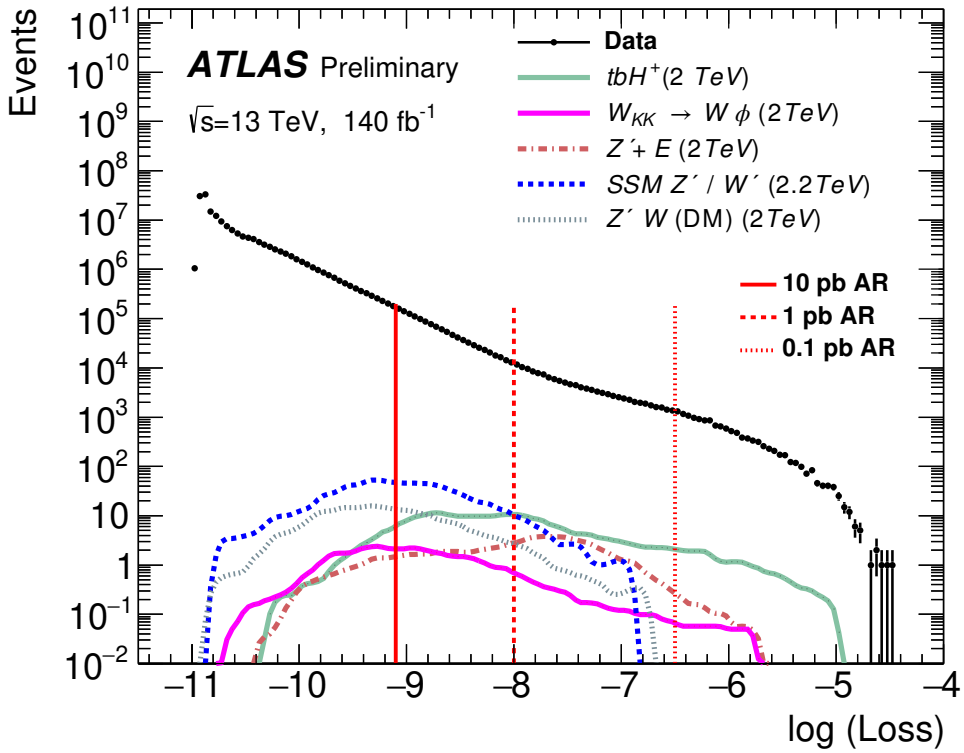

Figure 1: Distributions of the anomaly score from the AE for data and five benchmark BSM models. Their legends, from top to bottom, are: (1) charged Higgs production tbH^+ in association with a top quark; (2) a Kaluza–Klein gauge boson, W_{kk}, with the SM W boson and a radion ϕ; (3) a Z' boson decaying to a composite lepton E with a mass of 0.5 TeV; (4) the sequential standard model $W' \to WZ' \to \ell\nu q\bar{q}$; (5) a simplified dark-matter model with an axial-vector mediator Z' boson. The BSM predictions represent the expected number of events at 140 fb^{-1} at around 2 TeV of the heavy particles (H^+, W_{kk}, Z', W' and Z', respectively). The histograms of the BSM models are smoothed to avoid fluctuations due to low MC statistics. The vertical lines indicate the three anomaly regions (ARs). The labels of the three ARs indicate the visible cross section of hypothetical processes yielding a number of events equal to what observed in the data for an integrated luminosity of 140 fb^{-1}. The AE is applied on preselected events without any requirements on invariant mass distributions.

The nine invariant mass spectra in each anomaly region are examined to search for any localized excesses above the background hypothesis. The distributions are binned with increasing bin widths from 16 GeV to 150 GeV over the range of 0.3 to 8 TeV to reflect the jet energy resolution of the ATLAS detector. Similar

to the previous ATLAS searches based on the single-lepton requirement [47, 48], the following analytic function is used to describe the shape of the smoothly falling background:

$$f(x) = p_1(1-x)^{p_2}x^{p_3+p_4\ln x+p_5\ln^2 x}, \tag{1}$$

where $x \equiv m/\sqrt{s}$ and p_i are free parameters.

To avoid any possible bias, the analysis was developed without looking at the signal mass distributions in the data (blinded). To verify that Eq. (1) can describe the background shape separately for each of the nine invariant mass distributions, fits are performed on the SM background. This background is estimated from the MC simulation samples, which are composed of W+jets, $t\bar{t}$ and single-top processes, and a "loose electron" control sample from data that represents the QCD multi-jet events. Each event in the control sample must contain a leading electron that satisfies the "Loose" identification criteria [26] and fails the "Tight" identification criteria while also passing the anomaly region selections. This ensures the orthogonality of the control sample with the signal sample. The fit results show that Eq. (1) can describe the shape of the nine invariant mass distributions in the anomaly regions, with the χ^2/ndf values ranging between 0.7 and 1.9. The distributions of the fit residuals are consistent with the normal distribution. More complex fit functions do not improve this description. These studies do not take into account systematic variations in the SM simulations.

The established functional form of the background hypothesis is then applied to data in the anomaly regions. The likelihood fits of the data with Eq. (1) are performed to determine the free parameters and are shown in Figure 2 for the 10 pb AR. The data are well described by the fit with the χ^2/ndf values ranging between 0.64 and 1.19 for all nine mass spectra. The lower panels show the fit significances, calculated as $(d_i - f_i)/\delta_i$, where d_i is the data yield, f_i is the fit value, and δ_i is the uncertainty of data in the i-th bin. The distributions of the significances are found to be consistent with the normal distributions.

Each invariant mass spectrum is independently analyzed using the BumpHunter algorithm [54] to look for localized excesses without assumptions on BSM signal shapes. The uncertainty of the background shape is taken into account by using an alternative function that replaces the term $p_5\ln^2 x$ with $p_5/\sqrt{x}$ in Eq. (1). The algorithm searches for statistically significant deviations after taking into account the look-elsewhere effect [55] and the uncertainty of the background shape. Among the nine invariant mass distributions in the 10 pb AR, BumpHunter reports the largest deviation in the $m_{j\mu}$ spectrum near 4.8 TeV and the second largest deviation near 1.2 TeV, as shown in Fig. 2. Their statistical significances with systematic uncertainties will be discussed later.

The searches for localized excesses are also performed in the nine invariant mass distributions for 1 pb and 0.1 pb ARs. No significant excesses were found. The 4.8 TeV region of the $m_{j\mu}$ mass does not show the deviation from the background hypothesis of the same strength as that observed for the 10 pb AR.

The anomaly region selection is expected to improve the discovery potentials for BSM models. Figure 3 shows the improvements of the discovery sensitivity in the five benchmark BSM models at the 10 pb AR. The improvement is defined as $\Delta Z = ((Z_{\mathrm{AE}}/Z) - 1) \times 100\%$, where Z_{AE} is the discovery sensitivity in the anomaly region and Z is that before any cut on the anomaly score. The discovery sensitivity is defined as $Z = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}$ [56], where s is the number of signal events and b the number of background events calculated from the fit of Eq. (1) to the data. Both are counted under the bulk of the reconstructed signal shape within ±1 Root Mean Square around the mean value. The results show that the application of the AE trained on data improves the discovery sensitivity for most of the benchmark BSM models and the mass points. The BSM hypotheses that do not show improvements in the sensitivity

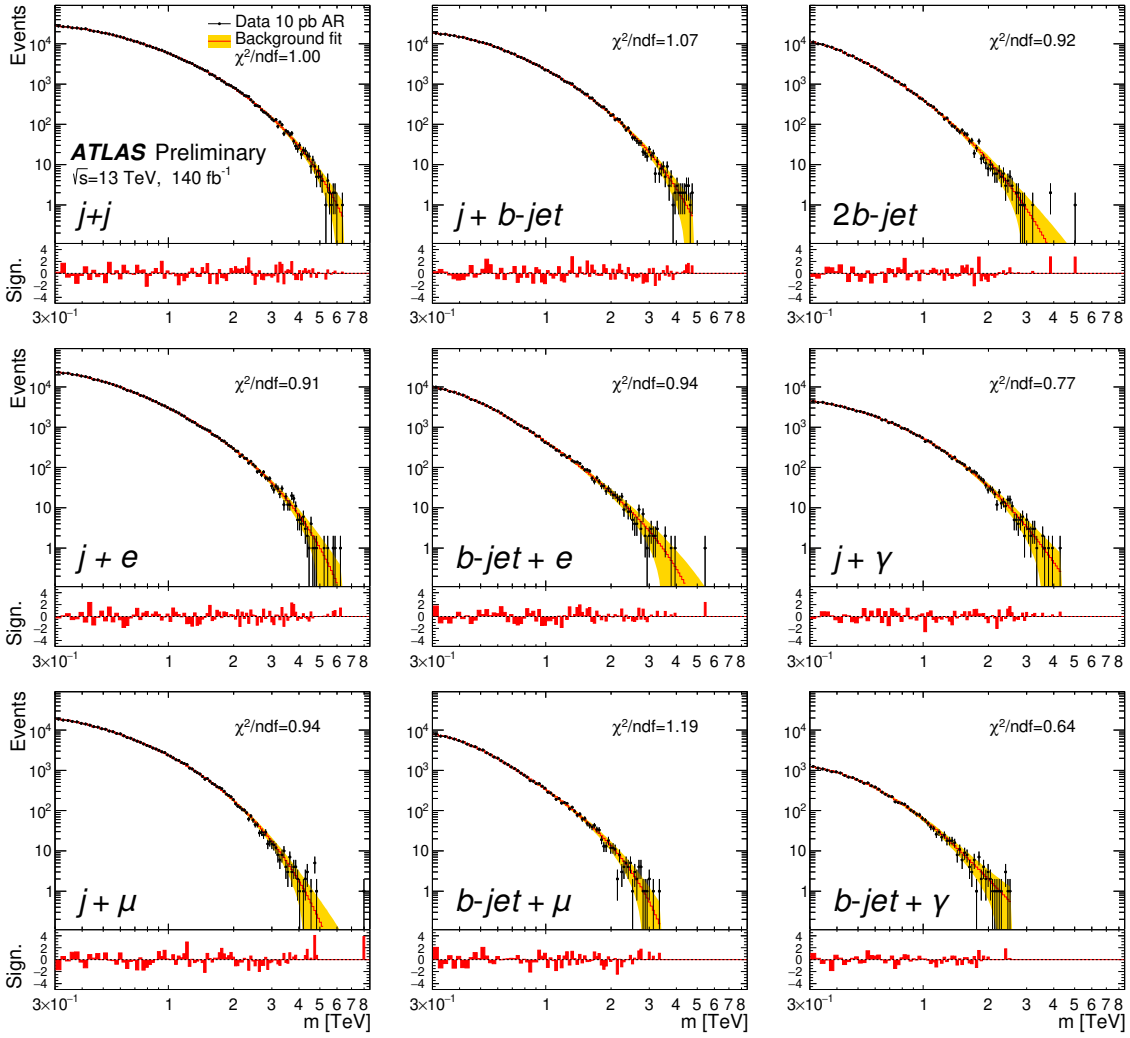

Figure 2: Invariant mass distributions of $j + Y$ for $m_{jY} > 0.3$ TeV in the 10 pb AR along with the fit from Eq. (1). The fits are represented by the lines, while the associated statistical uncertainties are indicated by the shaded bands. The lower panels show the bin-by-bin significances of deviations from the fit, calculated as $(d_i - f_i)/\delta_i$, where d_i is the data yield, f_i is the fit value, and δ_i is the uncertainty of data in the i-th bin.

correspond to models with a low mass of hypothetical particles or models with the number of reconstructed objects and the kinematic variables in the RMM that are closer to the SM events.

Due to the absence of any significant resonant signals, upper limits on the cross section times acceptance, efficiency, and branching ratio are set for Gaussian-shaped signals using the CL_s method [57]. The effects of systematic uncertainties associated with luminosity, jet energy scale, jet energy resolution, and object identification/reconstruction are included in the signal. These object related uncertainties are derived from the SSM simulated signals. The uncertainty stemming from the choice of background fit function is taken from the comparison of Eq. (1) with the alternative background shape. An uncertainty to estimate the fit bias, computed using the spurious-signal method described in Ref. [58], is included. The dominant source of systematic uncertainty is the spurious signal uncertainty which leads to a degradation of the upper limits by approximately 10%. The stochastic fluctuations in the AE training also cause an uncertainty if the AE

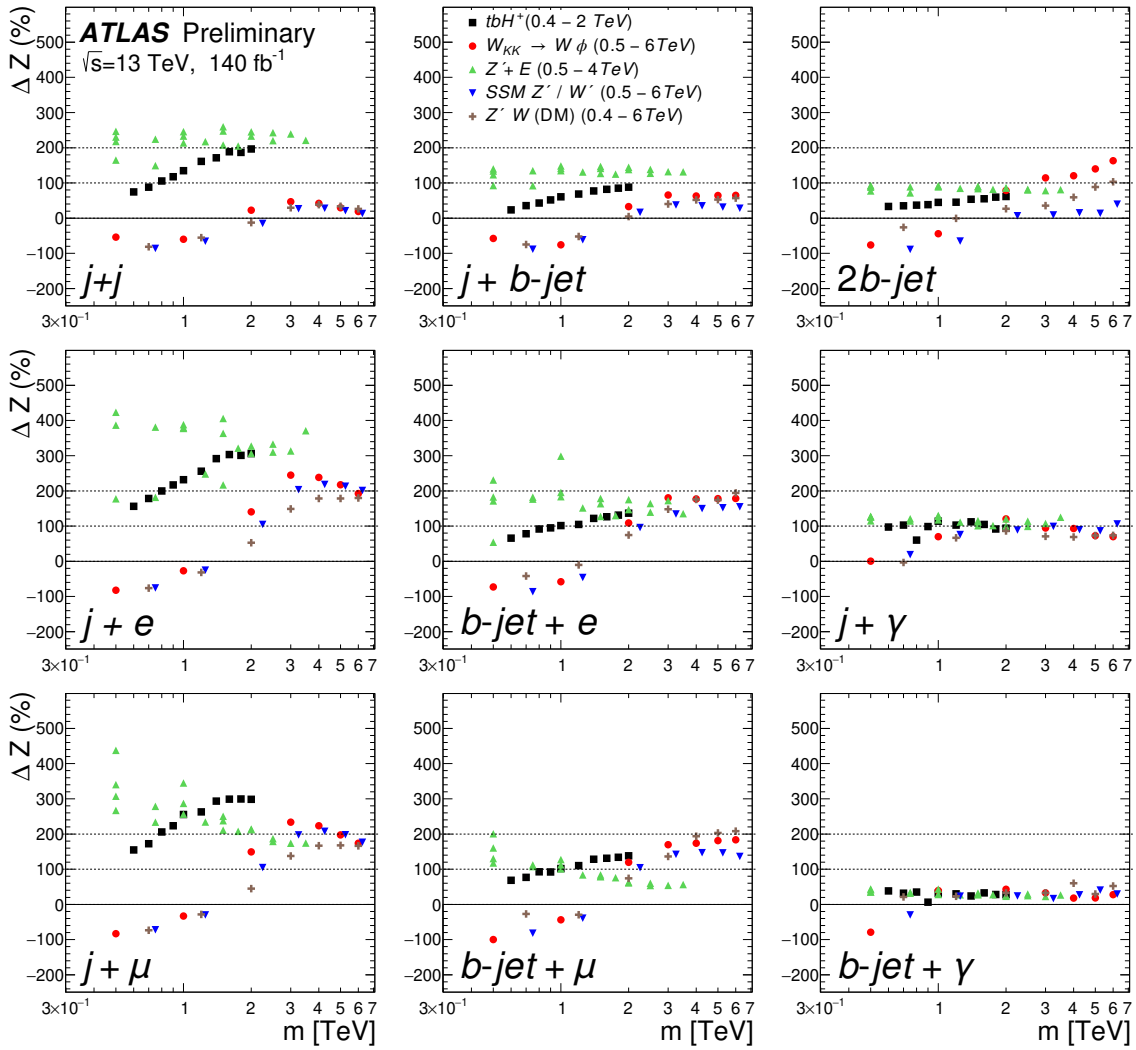

Figure 3: The values of ΔZ for the discovery sensitivity, as defined in the text, as a function of the invariant mass m. The nine invariant mass distributions are calculated in the 10 pb AR. The positive values in percentage indicate improvements in sensitivity. The two horizontal dashed lines at 100% and 200% are guides for the eye. The five benchmark BSM models are: (1) charged Higgs production tbH^+ in association with a top quark; (2) a Kaluza–Klein gauge boson, W_{kk}, with the SM W boson and a radion ϕ; (3) a Z' boson decaying to a composite lepton E. (4) the sequential standard model $W' \to WZ' \to \ell\nu q\bar{q}$; (5) a simplified dark-matter model with an axial-vector mediator Z' boson. The multiple markers shown for the composite-lepton model at fixed values of mass correspond to the variation of the composite lepton (E) mass between 0.25 and 3.5 TeV. The center positions of the markers are set to the masses of the corresponding heavy particles.

is retrained. It can cause an increase of the observed limits by 4% on average. This uncertainty is not included in the limits.

Upper limits at 95% confidence level (CL) on the production cross section of resonant signals with intrinsic width of 0% or 15% of the hypothetical mass in the 10 pb AR are presented in Figure 4. For the 0% width, the Gaussian signals have widths consistent with the experimental resolutions. The limits are weaker for the 15% resonance width as the signals are spread over more bins. The local significance of a signal with the

width of 0% for $m_{j\mu} = 4.8$ TeV, where the largest deviation was detected by BumpHunter, is found to be 2.9 standard deviations ($2.9\,\sigma$) using the asymptotic formulae. A more stringent ("HighPT") identification criterion on muons [27] removes the event near $m_{j\mu} = 8$ TeV, but it does not reduce the significance of the excess at 4.8 TeV. The local significance at around $m_{j\mu} = 1.2$ TeV is $2.8\,\sigma$.

The presented limits are more competitive than the limits on Gaussian signals presented in Ref. [47] which used the same preselected events. For example, the limits for the m_{jj} below 1 TeV reported in this analysis show about a factor of 2–3 improvement compared to the results in Ref. [47]. The search sensitivity is comparable at high masses. Many BSM models with complex final states are expected to have high AE acceptances. For example, the five benchmark BSM models discussed in Fig. 3 have an AE acceptance above 80% on average for the m_{jj} invariant mass distributions around 0.5 TeV for the 10 pb AR. The trained AE rejects about 90% of data for the same region of m_{jj}. This shows a potential to exclude heavy BSM particles with high AE acceptance.

In conclusion, model-independent searches for new phenomena in two-body invariant mass distributions are performed using an unsupervised anomaly detection technique with 140 fb^{-1} of pp collision events at $\sqrt{s} = 13$ TeV recorded by the ATLAS detector. Events are preseleted to contain at least one lepton with $p_{\mathrm{T}}^{\ell} > 60$ GeV and at least one jet with $p_{\mathrm{T}} > 30$ GeV. An unsupervised machine learning algorithm based on the AE architecture is trained with randomly selected 1% preselected collision events, without targeting any specific BSM signal. Three anomaly regions are defined using the reconstruction loss of the AE. Nine two-body invariant mass spectra, m_{jY}, are studied, where j stands for a light jet or a b-jet, and Y is a second light jet or b-jet, a lepton, or a photon. No evidence of significant excesses is observed. The maximum excess reported by BumpHunter is near 4.8 TeV in the $m_{j\mu}$ distribution for the 10 pb AR. Assuming a resonance with a width of 0%, the local significance of this excess is $2.9\,\sigma$.

The discovery sensitivity shows large improvement after the anomaly region selection, which is illustrated using several benchmark BSM models in the 10 pb AR. The analysis sets 95% CL upper limits on contributions from generic Gaussian signals to the studied invariant-masses distributions in this AR. The reported model-independent limits have a stronger potential to exclude generic heavy states with complex decaying final states, compared to the limits without anomaly region selections.

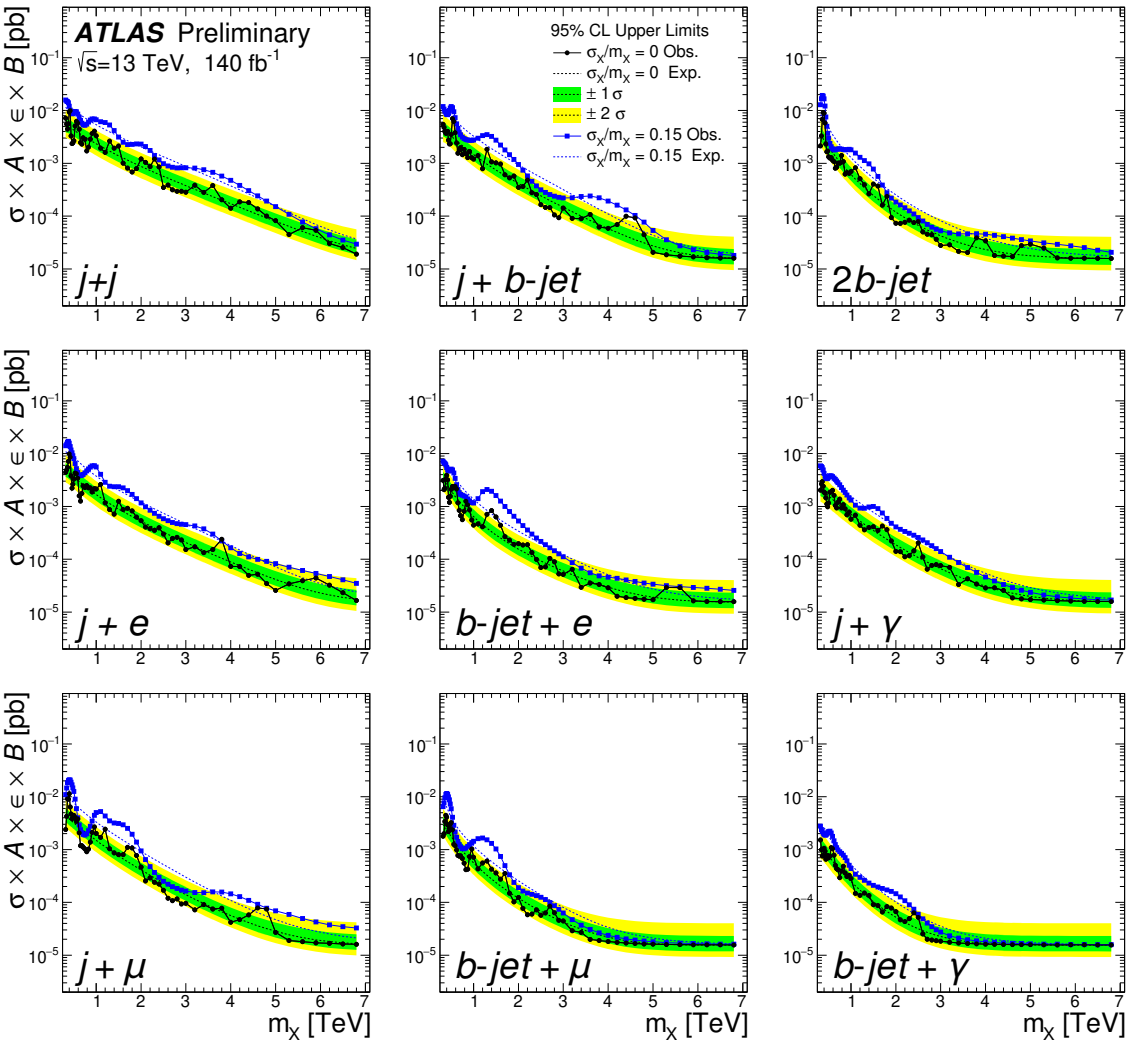

Figure 4: The 95% CL upper limits on the cross section times acceptance (A), efficiency (ϵ), and branching ratio (B) for Gaussian-shaped signals with various signal widths. The limits are calculated for events with at least one lepton with $p_{\mathrm{T}}^{\ell} > 60$ GeV in the 10 pb AR anomaly region. Two width hypotheses are shown: $\sigma_X/m_X = 0$ and $\sigma_X/m_X = 0.15$. In both cases, the detector resolution on jets is included in the simulation of signal samples. The $\pm 1\sigma$ and $\pm 2\sigma$ bands are shown for $\sigma_X/m_X = 0$ signals. Mass points are spaced 5% apart from their preceding point, starting from 0.3 TeV.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and

MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d'Avenir Labex, Investissements d'Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and MINERVA, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [59].

Appendix

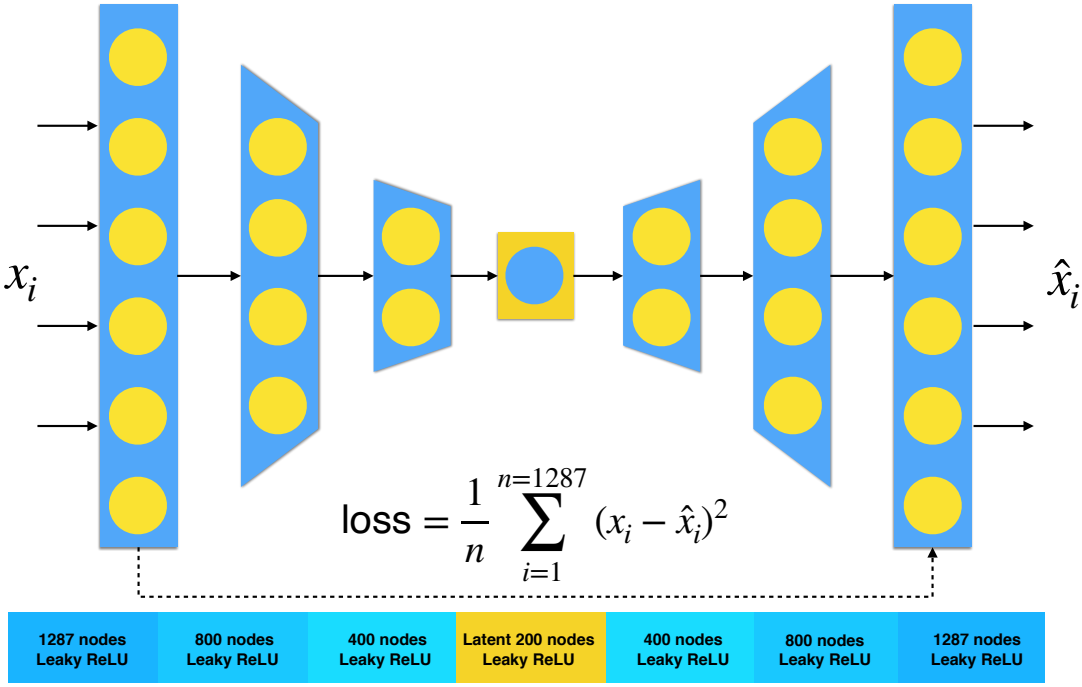

Figure 5: A schematic representation of the autoencoder architecture used for training and selection of the three anomaly regions. The input x_i and the output $\hat{x}_i$ are RMM values.

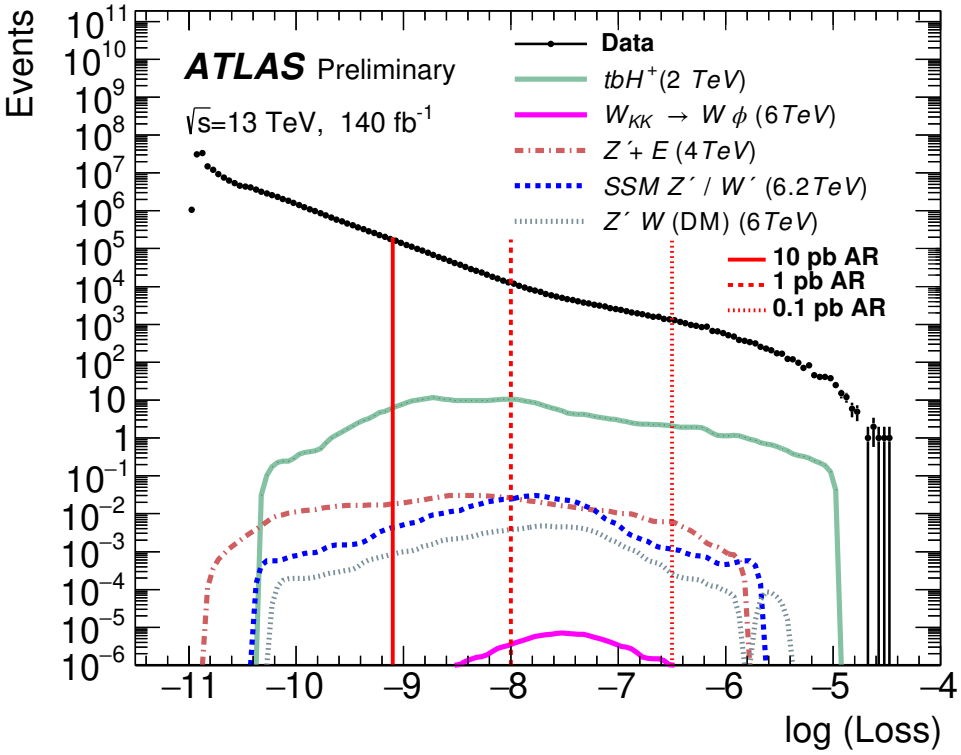

Figure 6: Distributions of the anomaly score from the AE for data and five benchmark BSM models. Their legends, from top to bottom, are: (1) charged Higgs production tbH^+ in association with a top quark, with the H^+ mass of 2 TeV; (2) a 6 TeV Kaluza–Klein gauge boson, W_{kk}, with the SM W boson and a radion ϕ; (3) a 4 TeV Z' boson decaying to a composite lepton E with a mass of 0.5 TeV; (4) the sequential standard model $W' \to WZ' \to \ell\nu q\bar{q}$, with the W' mass of 6.2 TeV and the Z' mass of 6 TeV; (5) a simplified dark-matter model with an axial-vector mediator Z' boson with a mass of 6 TeV. The histograms of the BSM models are smoothed to avoid fluctuations due to low MC statistics. The vertical lines indicate the three anomaly regions (ARs). The labels of the three ARs indicate the visible cross section of hypothetical processes yielding a number of events equal to what observed in the data for an integrated luminosity of 140 fb^{-1}.

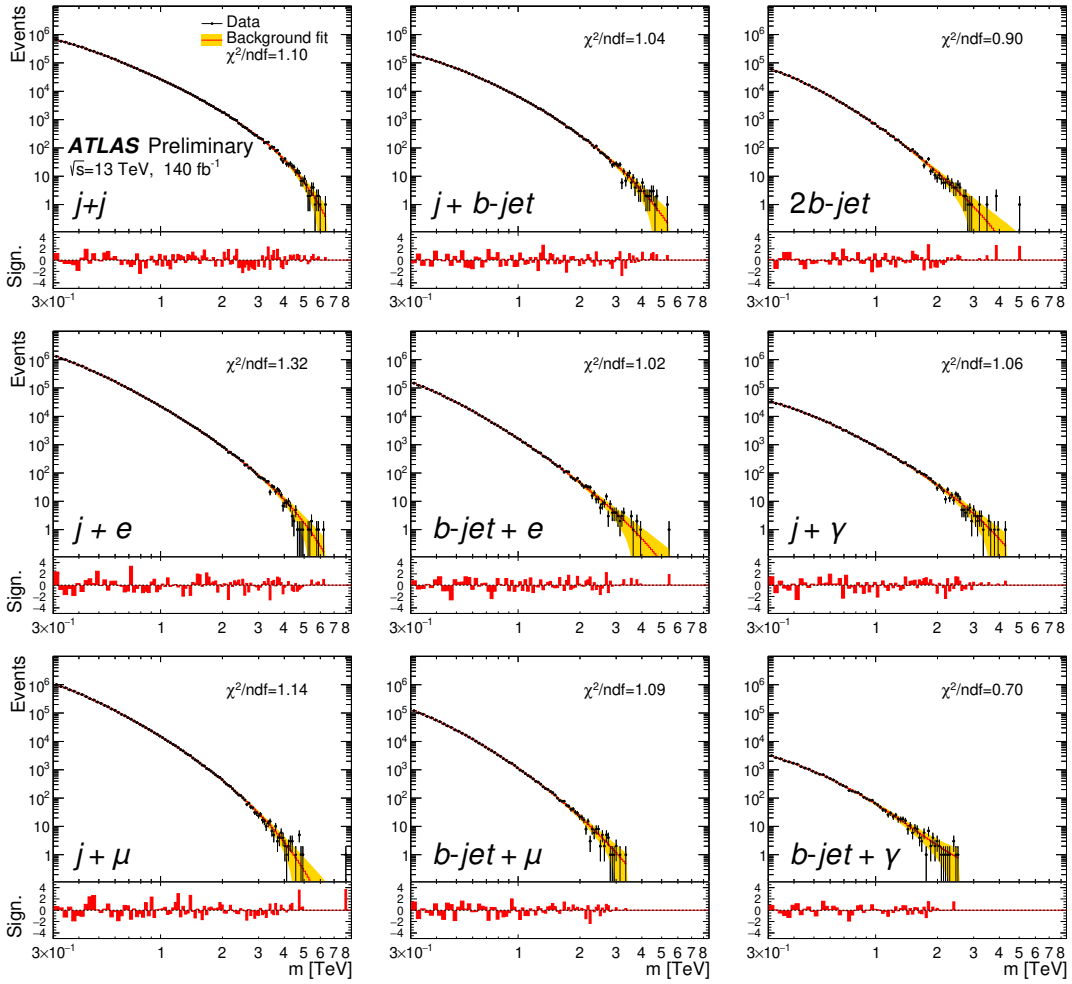

Figure 7: Invariant mass distributions of $j + Y$ for $m_{jY} > 0.3$ TeV after preselection along with the fit from Eq. (1). The fit is represented by red lines, and the associated uncertainties are indicated by yellow bands. The lower panels show the bin-by-bin significances of deviations from the fit, calculated as $(d_i - f_i)/\delta_i$, where d_i is the data yield, f_i is the fit value, and δ_i is the uncertainty of data in the i-th bin.

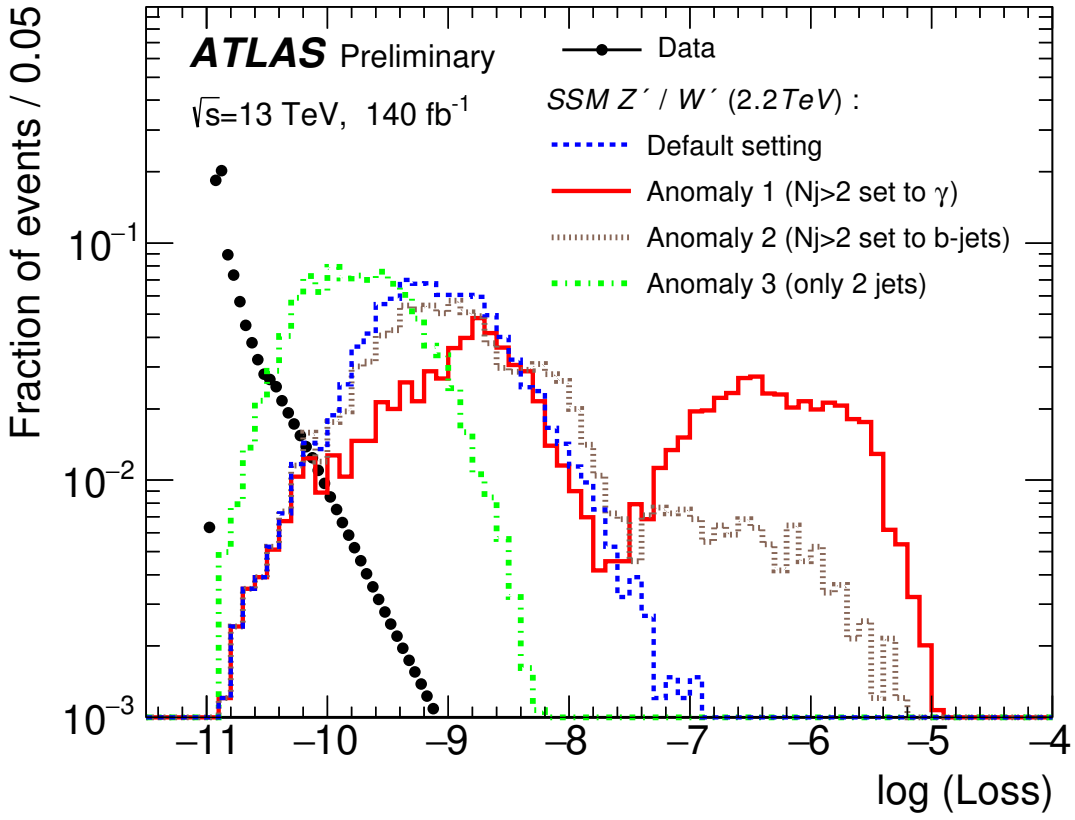

Figure 8: Distributions of the anomaly score for data and several anomaly scenarios. The example BSM model (shown with the dashed blue lines) is the sequential standard model $W' \to WZ' \to \ell\nu q\bar{q}$. The mass of W' is set to 2.2 TeV and the mass of the Z' is set to 2 TeV. This model leads to the final state of one lepton, two jets, and small missing transverse energy that is similar to the SM backgrounds. The other histograms represent events from the same model with some modifications to represent anomalous events: (1) "Anomaly 1" is the case where all jets beyond the second jet are artificially replaced with photons; (2) "Anomaly 2" is the case where all jets beyond the second jet are artificially replaced with b-jets; (3) "Anomaly 3" is the case of low-multiplicity events where all jets beyond the second jet are artificially removed. The histograms are normalized to unit. The left peak near log(Loss) = −9 visible in Anomaly 1 and 2 is from the events without a third jet.

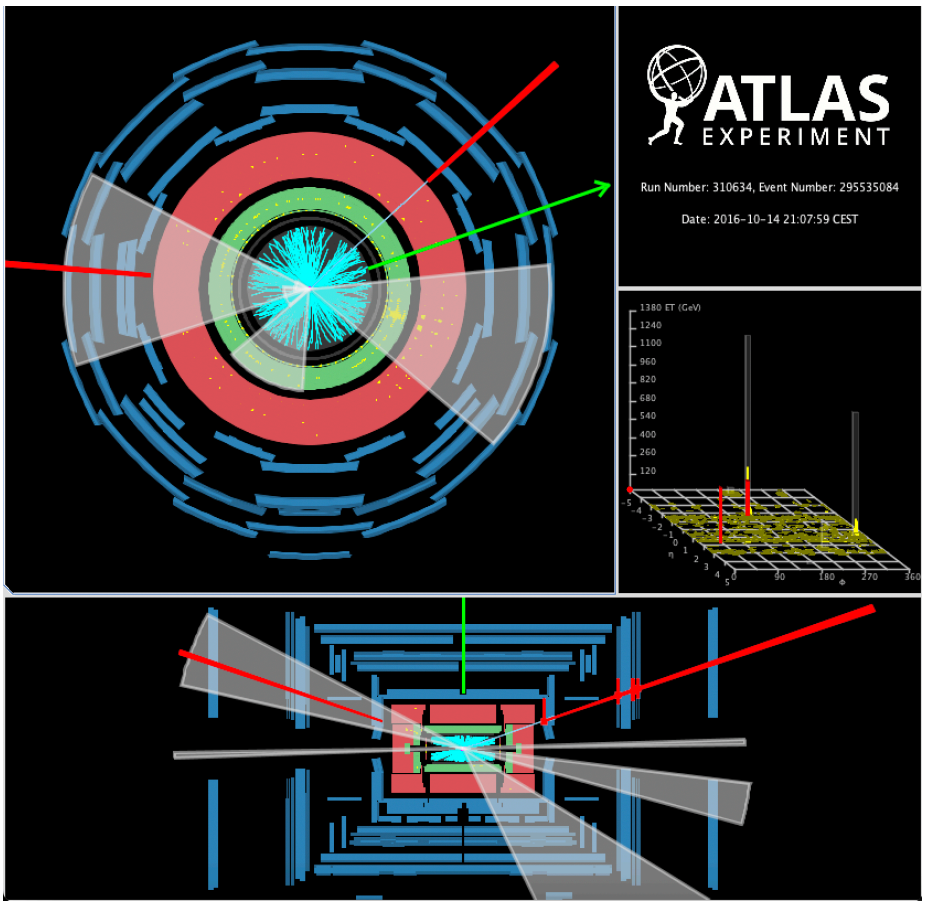

Figure 9: A display of an event with the reconstructed invariant mass of $m_{j\mu}$ = 4.72 TeV (Run Number: 310634, Event Number 295535084). It is recorded by ATLAS on 14th October 2016 with LHC stable beams at a pp collision energy of 13 TeV. This event belongs to the group of eight events with $m_{j\mu} > 4.7$ TeV contributing to the largest deviation from the background hypothesis in the 10 pb AR. The top-left panel presents a transverse (x–y plane) view of the event. The gray cones represent jets and the red lines represent muons. The green arrow indicates a missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) of 256 GeV . The red line closest to MET represents a muon with p_{T} = 430 GeV. The jet shown on the left side of the display forms the mass of $m_{j\mu}$ = 4.72 TeV. Other jets with $p_{\mathrm{T}} > 20$ GeV and muons with $p_{\mathrm{T}}^{\mu} > 60$ GeV are also displayed. The top-right panel shows the distribution of energy deposited in the calorimeters in the η–ϕ plane, where muons are marked in red and jets in yellow. The bottom panel displays the event in the longitudinal (z–y plane) perspective.

References

[1] ATLAS Collaboration, *A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment*, Eur. Phys. J. C **79** (2019) 120, arXiv: 1807.07447 [hep-ex] (cit. on p. 2).

[2] J. H. Collins, K. Howe, and B. Nachman, *Anomaly detection for resonant new physics with machine learning*, Phys. Rev. Lett. **121** (2018) 241803, arXiv: 1805.02664 [hep-ph] (cit. on p. 2).

[3] R. T. D'Agnolo and A. Wulzer, *Learning new physics from a machine*, Phys. Rev. D **99** (2019) 015014, arXiv: 1806.02350 [hep-ph] (cit. on p. 2).

[4] A. De Simone and T. Jacques, *Guiding new physics searches with unsupervised learning*, Eur. Phys. J. C **79** (2019) 289, arXiv: 1807.06038 [hep-ph] (cit. on p. 2).

[5] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, *Novelty Detection Meets Collider Physics*, Phys. Rev. D **101** (2020) 076015, arXiv: 1807.10261 [hep-ph] (cit. on p. 2).

[6] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, *QCD or What?* SciPost Phys. **6** (2019) 030, arXiv: 1808.08979 [hep-ph] (cit. on p. 2).

[7] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Variational autoencoders for new physics mining at the Large Hadron Collider*, JHEP **05** (2019) 036, arXiv: 1811.10276 [hep-ex] (cit. on p. 2).

[8] M. Farina, Y. Nakai, and D. Shih, *Searching for New Physics with Deep Autoencoders*, Phys. Rev. D **101** (2020) 075021, arXiv: 1808.08992 [hep-ph] (cit. on p. 2).

[9] A. Blance, M. Spannowsky, and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, JHEP **10** (2019) 047, arXiv: 1905.10384 [hep-ph] (cit. on p. 2).

[10] B. M. Dillon, D. A. Faroughy, J. F. Kamenik, and M. Szewc, *Learning the latent structure of collider events*, JHEP **10** (2020) 206, arXiv: 2005.12319 [hep-ph] (cit. on p. 2).

[11] T. Aarrestad et al., *The Dark Machines Anomaly Score Challenge: benchmark data and model independent event classification for the Large Hadron Collider*, SciPost Phys. **12** (2022) 043, arXiv: 2105.14027 [hep-ph] (cit. on p. 2).

[12] G. Kasieczka et al., *The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics*, Rept. Prog. Phys. **84** (2021) 124201, arXiv: 2101.08320 [hep-ph] (cit. on p. 2).

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533 (cit. on p. 2).

[14] H. Bourlard and Y. Kamp, *Auto-association by multilayer perceptrons and singular value decomposition*, Biological Cybernetics **59** (1988) 291 (cit. on p. 2).

[15] G. E. Hinton and R. S. Zemel, *Autoencoders, minimum description length, and Helmholtz free energy*, ed. by J. Cowan, G. Tesauro, and J. Alspector, NIPS'93: Proceedings of the 6th International Conference on Neural Information Processing Systems, 1994 3 (cit. on p. 2).

[16] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science **313** (2006) 504 (cit. on p. 2).

[17] ATLAS Collaboration, *Dijet Resonance Search with Weak Supervision Using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector*, Phys. Rev. Lett. **125** (2020) 131801, arXiv: 2005.02983 [hep-ex] (cit. on p. 2).

[18] ATLAS Collaboration, *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, ATLAS-CONF-2022-045, 2022, URL: https://cds.cern.ch/record/2816323 (cit. on p. 2).

[19] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs using $139\,fb^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **03** (2020) 145, arXiv: 1910.08447 [hep-ex] (cit. on p. 2).

[20] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 2).

[21] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on pp. 2, 3).

[22] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex] (cit. on pp. 2, 3).

[23] ATLAS Collaboration, *Trigger Menu in 2018*, ATL-DAQ-PUB-2019-001, 2019, URL: https://cds.cern.ch/record/2693402 (cit. on p. 2).

[24] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 2).

[25] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: https://cds.cern.ch/record/2037717 (cit. on p. 2).

[26] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 2, 3, 6).

[27] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 2, 3, 9).

[28] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on p. 3).

[29] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 3).

[30] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 3).

[31] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 3).

[32] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 3).

[33] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*,
ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on p. 3).

[34] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581,
arXiv: 1510.03823 [hep-ex] (cit. on p. 3).

[35] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 970,
arXiv: 1907.05120 [hep-ex] (cit. on p. 3).

[36] ATLAS Collaboration,
Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run,
ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on p. 3).

[37] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of* $t\bar{t}\gamma$ *production in leptonic final states at* $\sqrt{s} = 13$ *TeV in ATLAS*, Eur. Phys. J. C **79** (2019) 382,
arXiv: 1812.01697 [hep-ex] (cit. on p. 3).

[38] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78** (2018) 903,
arXiv: 1802.08168 [hep-ex] (cit. on p. 3).

[39] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823,
arXiv: 1005.4568 [physics.ins-det] (cit. on p. 3).

[40] GEANT4 Collaboration, S. Agostinelli, et al., *GEANT4 – a simulation toolkit*,
Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 3).

[41] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*,
ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965 (cit. on p. 3).

[42] S. V. Chekanov, *Imaging particle collision data for event classification using machine learning*,
Nucl. Instrum. Meth. A **931** (2019) 92, arXiv: 1805.11650 [hep-ph] (cit. on p. 3).

[43] S. V. Chekanov and W. Hopkins, *Event-based anomaly detection for searches for new physics*,
Universe **8** (2022) 494, arXiv: 2111.12119 [hep-ph] (cit. on pp. 3, 4).

[44] M. Abadi et al., *TensorFlow: A system for large-scale machine learning*,
Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation,
2016, URL: https://arxiv.org/abs/1605.08695 (cit. on p. 4).

[45] B. Xu, N. Wang, T. Chen, and M. Li,
Empirical evaluation of rectified activations in convolutional network, (2015),
arXiv: 1505.00853 [cs.LG] (cit. on p. 4).

[46] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, (2014),
arXiv: 1412.6980 [cs.LG] (cit. on p. 4).

[47] ATLAS Collaboration, *Search for dijet resonances in events with an isolated charged lepton using* $\sqrt{s} = 13$ *TeV proton–proton collision data collected by the ATLAS detector*, JHEP **06** (2020) 151,
arXiv: 2002.11325 [hep-ex] (cit. on pp. 4, 6, 9).

[48] ATLAS Collaboration,
Search for new phenomena in multi-body invariant masses in events with at least one isolated lepton and two jets using $\sqrt{s} = 13$ TeV proton-proton collision data collected by the ATLAS detector, (2022), arXiv: 2211.08945 [hep-ex] (cit. on pp. 4, 6).

[49] C. Degrande, M. Ubiali, M. Wiesemann, and M. Zaro,
Heavy charged Higgs boson production at the LHC, JHEP **10** (2015) 145,
arXiv: 1507.02549 [hep-ph] (cit. on p. 4).

[50] K. S. Agashe et al., *LHC signals from cascade decays of warped vector resonances*,
JHEP **05** (2017) 078, arXiv: 1612.00047 [hep-ph] (cit. on p. 4).

[51] M. Chala and M. Spannowsky,
Behavior of composite resonances breaking lepton flavor universality,
Phys. Rev. D **98** (2018) 035010, arXiv: 1803.02364 [hep-ph] (cit. on p. 4).

[52] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders*,
Zeit. Phys. C **45** (1989) 109, [Erratum: Zeit.Phys.C 47, 676 (1990)] (cit. on p. 4).

[53] D. Abercrombie et al., *Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum*, Phys. Dark Univ. **27** (2020) 100371,
arXiv: 1507.00966 [hep-ex] (cit. on p. 4).

[54] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter*, (2011),
arXiv: 1101.0390 [physics.data-an] (cit. on p. 6).

[55] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*,
Eur. Phys. J. C **70** (2010) 525, arXiv: 1005.1891 [physics.data-an] (cit. on p. 6).

[56] G. Cowan, *Discovery sensitivity for a counting experiment with background uncertainty*, (2012),
URL: https://www.pp.rhul.ac.uk/~cowan/stat/notes/medsigNote.pdf (cit. on p. 6).

[57] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693
(cit. on p. 7).

[58] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with $36\,fb^{-1}$ of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
Phys. Rev. D **98** (2018) 052005, arXiv: 1802.04146 [hep-ex] (cit. on p. 7).

[59] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, ATL-SOFT-PUB-2021-003, 2021,
URL: https://cds.cern.ch/record/2776662 (cit. on p. 11).