



# Phenomenology of photons-enriched semi-visible jets

Cesare Cazzaniga<sup>1,a</sup> , Alessandro Russo<sup>2,b</sup> , Emre Sitti<sup>1,c</sup> , Annapaola de Cosa<sup>1,d</sup> 

<sup>1</sup> ETH Zürich, Institute for Particle Physics and Astrophysics, 8093 Zurich, Switzerland

<sup>2</sup> Department of Physics, Stanford University, Stanford, CA 94305, USA

Received: 11 July 2024 / Accepted: 6 November 2024  
© The Author(s) 2024

**Abstract** This Letter proposes a new signature for confining dark sectors at the LHC. Under the assumption of a QCD-like hidden sector, hadronic jets containing stable dark bound states originating from hidden strong dynamics, known as semi-visible jets, could manifest in proton-proton collisions. In the proposed simplified model, a heavy  $Z'$  mediator coupling to SM quarks allows the resonant production of dark quarks, subsequently hadronizing in stable and unstable dark bound states. The unstable dark bound states can then decay back to SM quarks via the same  $Z'$  portal or photons via a lighter pseudo-scalar portal (such as an axion-like particle). This mechanism creates a new signature where semi-visible jets are enriched in non-isolated photons. We show that these exotic jets evade the phase space probed by current LHC searches exploiting jets or photons due to the expected high jet neutral electromagnetic fraction and photons candidates non-isolation, respectively. In the proposed analysis strategy to tackle such signature, we exploit jets as final state objects to represent the underlying QCD-like hidden sector. We show that, by removing any selection on the neutral electromagnetic fraction from the jet identification criteria, higher signal efficiency can be reached. To enhance the signal-to-background discrimination, we train a deep neural network as a jet tagger that exploits differences in the substructure of signal and background jets. We estimate that with the available triggers for Run 2 and this new strategy, a high mass search can claim a  $5\sigma$  discovery (exclusion) of the  $Z'$  boson with a mass up to 5 TeV (5 TeV) with the full Run 2 data of the LHC when the fraction of unstable dark hadrons decaying to photons pairs is around 30%, and with a coupling of the  $Z'$  to SM quarks of 0.25.

## 1 Introduction

The presence of dark matter (DM) in the universe [1,2] is clear evidence of limitations of the Standard Model (SM) framework, invoking new physics to find a full explanation of the phenomenon. The lack of experimental confirmation of the most popular hypotheses investigated so far [3,4] connects to the idea of new physics belonging to complex sectors of matter barely accessible to SM particles, *Hidden sectors*. Hidden sectors are introduced in the Hidden Valley models [5] as extensions of the SM, and connected to the SM via a heavy mediator or a weak coupling. These models can rather easily accommodate DM candidates, identifying them with stable particles within the dark sector [6–12]. Moreover, if realised in nature, the Hidden Valley scenario may result in unusual and little-studied phenomena at the LHC. A notable example is *semi-visible jets (SVJ)* signatures appearing in strongly coupled Hidden Valley scenarios, where dark QCD exists and behaves in a QCD-like way. Previous works [10–17], have introduced such class of experimental signatures, composed of a mixture of stable and unstable dark bound states originating from the hadronization of dark quarks at colliders. SVJs appear as partially visible jets of particles in the detectors, with missing transverse momentum from stable dark mesons ( $\cancel{E}_T$ ) aligned to the direction of one of the jets. Fully hadronic SVJ signatures have been proposed in [11,13–15] and searched for by the CMS and ATLAS experiments in s- and t-channel production modes [18,19], respectively. Novel signatures with SVJ enriched in leptons from dark photon decays have been recently proposed in [16], along with signatures enriched in  $\tau$  leptons [17] proposed by models assuming non-democratic decays to leptons). In this Letter we present a simplified model for SVJs enriched in photons. The model has been developed following [16,17,20,21]. In the proposed models, a  $Z'$  mediator acts as a messenger between the visible and dark sectors via coupling to SM quarks, thus allowing the resonant pro-

<sup>a</sup> e-mail: cesare.cazzaniga@cern.ch (corresponding author)

<sup>b</sup> e-mail: arusso00@stanford.edu

<sup>c</sup> e-mail: esitti@ethz.ch

<sup>d</sup> e-mail: adecosa@phys.ethz.ch

duction of dark quarks in proton-proton collisions and the decays of the unstable dark pions to SM quarks. In such way the SVJs can be resonantly produced at the LHC. The decay of the dark pions to photons is guaranteed via a pseudo-scalar (such as an axion-like particle, ALP) portal coupling to dark quarks and SM photons. This mechanism allows the production of non-isolated photon pairs inside semi-visible jets (SVJ $\gamma$  signature). As an extreme case of this signature, we also show that within our model, it is possible to obtain jets comprised exclusively of photons (photon-jets [22]). We find that present searches have limited sensitivity to these novel signatures since SVJ $\gamma$  objects do not pass standard identification criteria both for jets and photons. We propose a dedicated strategy to identify such objects with respect to jets faking photons and prompt photon production.

### 2 Model setup

The model is primarily based on the frameworks introduced in [16,17,20,21], allowing the dark sector to communicate with the SM. As mentioned above, it employs a  $Z'$  mediator and a pseudo-scalar ALP  $a$ . A simple description and summary of the physical processes are shown in Fig. 1, and the main interactions are governed by the following Lagrangians:

$$\begin{aligned} \mathcal{L}_{Z'} \supset & -Z'_\mu \bar{q}_i \gamma^\mu (g_{ij}^R P_R + g_{ij}^L P_L) q_j \\ & -Z'_\mu \bar{\psi}_i \gamma^\mu (y_{ij}^R P_R + y_{ij}^L P_L) \psi_j, \end{aligned} \tag{1}$$

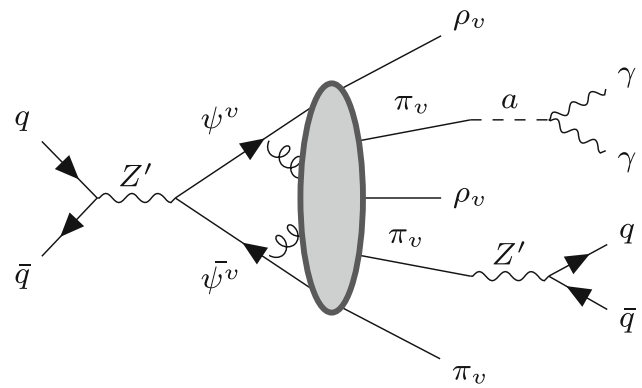
for  $Z'$ , and

$$\begin{aligned} \mathcal{L}_a \supset & (D_\mu a)^\dagger (D^\mu a) - \frac{1}{2} m_a^2 a^2 \\ & - i \lambda^L a \bar{L} \gamma_5 L - i \lambda_{ij}^\psi a \bar{\psi}_i^v \gamma_5 \psi_j^v + \text{h.c.} \end{aligned} \tag{2}$$

for  $a$ . Here,  $q_i$  are the SM quarks of flavor  $i$ , while  $\psi_i^v$  denotes the dark quarks. In what follows, we will omit the dark sector flavor indices to simplify the notation; however, this does not imply that there is only one flavor. We also assume the couplings of the  $Z'$  to leptons to be negligible compared to quarks as suggested by the constraints from the high mass di-lepton searches [23,24]. Additionally, a vector-like lepton ( $L$ ) [21] mediates the decay of the dark bound state into two photons, resulting in the following effective interaction:

$$\mathcal{L}_{\text{eff}} \supset \frac{\alpha_{EM} \lambda_L \lambda_\psi \Lambda_v F_{\pi_v}}{4\pi} \frac{\pi_v}{m_a^2 M_L} F_{\mu\nu} \tilde{F}^{\mu\nu}, \tag{3}$$

where  $\alpha_{EM}$  is the QED coupling,  $M_L$  is the vector-like lepton mass,  $\Lambda_v$  is the dark sector confinement scale,  $F_{\pi_v}$  is the dark (pseudo-scalar) meson decay constant,  $\pi_v$  is the lightest dark bound state and  $F^{\mu\nu}$  is the electromagnetic field tensor. In



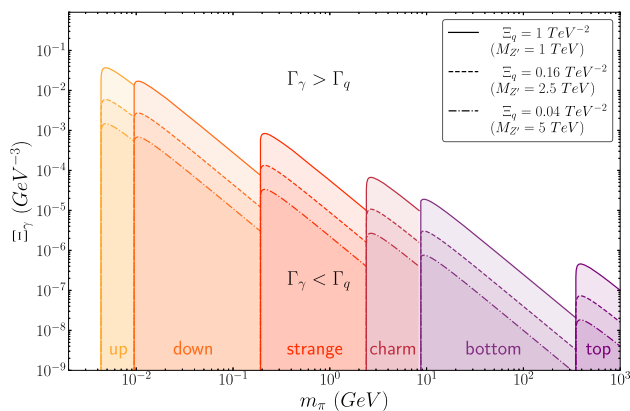
**Fig. 1** Feynman diagram illustrating the main processes in a collider experiment. The gray ellipses represent the hadronization part

addition, as discussed in [21], it is also possible to implement a gluon portal for the ALP, similar to the one for Eq. (3), but employing a vector-like quark  $Q$  (of mass  $M_Q$ ) instead of  $L$ . This new portal will, however, be suppressed by a factor of  $(1/M_Q)^2$  in the decay width. Vector-like quarks have been extensively studied [25] and are more constrained on the full range of masses with respect to vector-like leptons [26]. For this reason, this portal is neglected in favour of the photon channel. In a collider experiment, as shown in Fig. 1, the dark sector is accessed through the  $Z'$ , which is initially produced by a parton interaction. Then the dark quarks hadronize, producing multiple hadrons. These states can decay into stable DM or back into the SM using the portals described in Eqs. (1) and (2).

In the context of our model, there are two isospin triplets for the dark hadrons:  $\rho_v$  and  $\pi_v$ , the latter being the lightest. Our simplified model does not include any isospin-singlet mesons. Each of these states can have three different configurations depending on the dark sector flavors:  $(\pi_v/\rho_v)^+$ ,  $(\pi_v/\rho_v)^0$ , and  $(\pi_v/\rho_v)^-$ . This should not be understood as an electromagnetic charge, as it arises from an accidental symmetry (isospin) which can be easily broken. For this reason, we will omit this notation and simply refer to the bound states as  $\pi_v$  and  $\rho_v$ . Depending on the model's symmetry, the latter will decay into its lighter counterpart ( $\pi_v$ ), which can then decay into either two SM quarks via the  $Z'$  or two photons via the pseudo-scalar  $a$ . These are the “visible” decays. However, it is also possible for the light mesons (or a fraction of them) to remain stable, making them potential DM candidates. The decay widths for the visible decays are as follows:

$$\Gamma(\pi_v \rightarrow \bar{q}_i q_i) = \Gamma_q = \frac{N_c}{32\pi} m_{q_i}^2 m_{\pi_v}^3 \sqrt{1 - \frac{4m_{q_i}^2}{m_{\pi_v}^2}} \mathcal{E}_q^2, \tag{4}$$

$$\Gamma(\pi_v \rightarrow \gamma\gamma) = \Gamma_\gamma = \frac{\alpha_{EM}^2}{64\pi^2} m_{\pi_v}^7 \mathcal{E}_\gamma^2. \tag{5}$$



**Fig. 2** The plot shows seven different regions in parameter space ( $\mathcal{E}_\gamma$  vs  $m_{\pi_v}$ ) where different decay modes of the bound states are dominant. In the shaded regions below the lines,  $\pi_v$  will predominantly decay into two SM quarks: up quarks in the yellow region, down quarks in the orange region, strange quarks in the red region, charm quarks in the cardinal region, bottom quarks in the violet region, and top quarks in the purple region. In the white region above the lines, the dominant decay mode of  $\pi_v$  is into two photons. The lines themselves represent the case where the branching ratio between  $\pi_v \rightarrow \gamma\gamma$  and  $\pi_v \rightarrow \bar{q}q$  is exactly 50%. This analysis has been performed for different values of  $\mathcal{E}_q$

where  $N_c$  is the number of SM colors,  $m_{q_i}$  is the SM quark mass and  $m_{\pi_v}$  is the dark pion mass. Here  $\mathcal{E}_q$  and  $\mathcal{E}_\gamma$  functions are effective couplings for the theory, keeping within themselves some of the free degrees of freedom. They are defined as:

$$\mathcal{E}_q := \frac{\Delta_i^q \Delta^\psi}{M_{Z'}^2} \quad \text{and} \quad \mathcal{E}_\gamma := \frac{\lambda_L \lambda_\psi}{m_a^2 M_L}, \tag{6}$$

where  $\Delta_i^q$  and  $\Delta^\psi$  are, respectively,  $g_{ii}^R - g_{ii}^L$  and  $y^R - y^L$ , and the assumption  $F_{\pi_v} \approx \Lambda_v \approx m_{\pi_v}$  was made. Depending on the values of  $\mathcal{E}_q$  and  $\mathcal{E}_\gamma$ , the bound states will decay back into photons or quarks. By studying the branching fraction in the latter case, one can observe that only one flavor of quark becomes relevant for the same value of  $\mathcal{E}_q$  and given  $m_{\pi_v}$ . Specifically, if we order the quarks by their masses ( $i < j \Rightarrow m_{q_i} < m_{q_j}$ ) and we have that  $m_{q_i} < m_{\pi_v} < m_{q_{i+1}}$ , then the flavor  $i + 1$  is forbidden, and the flavor  $i$  dominates over  $i - 1$  by a factor of  $(m_{q_i} / m_{q_{i-1}})^2$ .

It is also possible to tune  $\mathcal{E}_q$  and  $\mathcal{E}_\gamma$  to achieve a specific value of  $R_{\pi_v} := \Gamma_q / \Gamma_\gamma$ . By adjusting this ratio, the number of photons or quarks coming from the bound states in the SVJ can be controlled. Figure 2 shows the boundary between two scenarios as a function of the bound state mass and  $\mathcal{E}_\gamma$ . The shaded region below the lines represents the case where the bound states decay predominantly into quarks over photons while, above the lines, the opposite case occurs. This analysis is conducted for different values of  $\mathcal{E}'_q$  (or  $M_{Z'}$  by keeping the  $\Delta$ s unitary). It is worth mentioning that in the specific case

where  $\Delta_i^q = 0$ , so for a pure  $Z'$  vector portal, the decays of the dark pions back to SM quarks via the  $Z'$  are prohibited. In this specific scenario, the pseudo-scalar can still decay to photons via the ALP, leading to sprays of collimated photons resembling the photon-jets signature [22], but with a potentially much larger expected number of photons pairs per jet. In this study, we only consider dark-bound states to be heavy enough to decay promptly. However, longer lifetimes can be obtained, allowing for signatures with additional displaced tracks and vertices inside the signal jets and non-prompt, non-isolated photons.

### 3 Monte Carlo simulations

Samples containing signal events, including initial- and final-state radiation and underlying event, have been produced with `Pythia8.307` [27]. The Hidden Valley module [28,29] in `Pythia8` simulated the dark sector’s showering, hadronization, and decays. Detector effects are simulated using `Delphes3` [30]. Following previous studies [18], anti- $k_T$  jets [31,32] with  $R = 0.8$  (AK8 jets) are clustered requiring a minimum  $p_T$  of 200 GeV. All the samples have been normalised to the LO cross-section prediction computed with the `MadGraph5_aMC@NLO` [33] event generator using the NN23LO1 parton distribution functions [34] from the `Lhapdf` [35] repository. As a reference, we have normalized the signals to the expected cross-section for  $pp \rightarrow Z' \rightarrow q_v \bar{q}_v$  as in [18], thus assuming a coupling of the  $Z'$  to SM quarks equal to 0.25, and its coupling to dark quarks equal to 0.4, following the most recent recommendations from the LHC DM Working Group [36].

For the SVJ  $s$ -channel signal process,  $20 \cdot 10^3$  events have been generated scanning a mass range for  $M_{Z'}$  between 1.5 and 5 TeV. The number of stable and unstable dark hadrons produced in the dark hadronization process can vary according to the details of the dark sector. In previous literature [13,14], an effective invisible fraction parameter has been defined as  $r_{\text{inv}} = \langle N_{\text{stable}} / (N_{\text{stable}} + N_{\text{unstable}}) \rangle$ , where  $N_{\text{stable}}$  is the number of stable dark hadrons, and  $N_{\text{unstable}}$  is the number of those that are unstable, decaying back to SM quarks and photons. This invisible fraction allows the capture of variations in the details of the hidden sector and defines the parameter space of a SVJ signature in terms of the amount of missing transverse momentum  $\cancel{E}_T$ . The quantity  $\cancel{E}_T$  is computed from the transverse component of the sum of the momenta of all detected particles in an event [30], and can be used to quantify the amount of transverse momentum missing due to invisible particles. The higher  $r_{\text{inv}}$ , the more the final state topology will resemble a mono-jet signature, where  $\cancel{E}_T$  recoils against initial-state radiation jets. For our studies, we choose a benchmark value  $r_{\text{inv}} = 0.3$ , representing an intermediate regime between di-jet and mono-jet topologies.

Simulations scanning values setting  $r_{\text{inv}} = 0.5, 0.7$  have been produced to assess the impact on the performance of the jet tagger proposed in Sect. 5. The dark pseudo-scalar ( $\pi_v$ ) and vector ( $\rho_v$ ) meson masses can differ according to the non-perturbative dynamics of the hidden sector, even for mass-degenerate dark quarks. Following a similar approach as in [16, 17], we employed the lattice QCD fits in [37] to predict the masses of dark vector mesons  $m_{\rho_v}$  from the input ratio  $m_{\pi_v}/\Lambda_v$ , where  $\Lambda_v$  (HiddenValley:Lambda) is the dark confinement scale, which fixes the overall mass scale for the dark bound states. As a benchmark,  $\Lambda_v$  has been set to 20 GeV and the ratio  $m_{\pi_v}/\Lambda_v = 1.0$ , therefore fixing  $m_{\pi_v}$  (4900111:m0, 4900211:m0) at 20 GeV and  $m_{\rho_v}$  (4900113:m0, 4900213:m0) at  $\sim 54$  GeV. With these dark QCD parameter settings, the  $\rho_v \rightarrow \pi_v \pi_v$  decay is open, and we assume 100% branching ratio for this internal decay in the dark sector. Additional simulations of the signal model scanning values of  $\Lambda_v = 15, 30, 50$ , with fixed  $m_{\pi_v}/\Lambda_v = 1.0$ , has been produced for additional studies on the performance of the jet tagging technique proposed in Sect. 5. Consequently, only unstable pseudo-scalar dark mesons are allowed to decay back to the SM within the chosen parameter space. As discussed in the previous section, the unstable bound states can decay via the  $Z'$  portal to SM quarks and via the ALP to photons. The fraction of decays of the dark pseudo-scalars to photons can be tuned via the effective parameter  $\text{BR}_\gamma$ . This parameter can assume values between 0 and 1, and can be tuned by changing  $R_{\pi_v}$  in the simplified model introduced in the previous section.

In summary, there are four parameters sensitive to the details of the dark sector: the confinement scale  $\Lambda_v$ , the pseudo-scalar mass ratio  $m_{\pi_v}/\Lambda_v$ , the invisible fraction  $r_{\text{inv}}$ , and the branching ratio to photons  $\text{BR}_\gamma$ . Moreover, there are three portal parameters related to the production of  $\text{SVJ}\gamma$  in proton-proton collisions: the  $Z'$  pole mass  $M_{Z'}$ , the coupling of the  $Z'$  to dark quarks  $g_{q_v}$  and the coupling to SM quarks  $g_q$ . The portal parameters related to the VLL and ALP are entered in the  $\text{BR}_\gamma$  definition and are only relevant to control the bound states' decays to photons. In order to probe the impact of enhanced photon content, we scan over the branching  $\text{BR}_\gamma$  (0.3, 0.5, 0.7, 1) which is equivalent to varying the parameters  $\mathcal{E}_q$  and  $\mathcal{E}_p$ . Similar to hadronic and leptonic SVJ signatures, SM QCD interactions, top pair productions, and electroweak processes are background sources for  $\text{SVJ}\gamma$ . Moreover, prompt photons production ( $\gamma + \text{jets}$ ) can also emulate signal jets. We, therefore, include this additional process in our studies. All the background samples have been generated at LO with the MadGraph5\_aMC@NLO event generator using the NN23LO1 parton distribution functions from the Lhapdf repository. The evolution of the parton-level events and hadronization are performed with Pythia8. The QCD sample ( $4.5 \cdot 10^7$  events) has been produced requiring a generator level cut on the leading parton

jet transverse momentum  $p_T > 500$  GeV. The QCD background is particularly relevant due to the large cross-section and the possibility of mis-reconstruction of the jet momentum, leading to additional missing momentum aligned with the jet axis. Moreover, QCD jets can contain photons from neutral SM pions decays as well as final state QED radiation, which might resemble the  $\text{SVJ}\gamma$  signal. The  $t\bar{t} + \text{jets}$  inclusive sample ( $5 \cdot 10^7$  events) has been generated with up to two additional partons. This background mainly becomes relevant when the top quarks are highly boosted, and therefore, the  $W$  boson decay and the  $b$ -initiated jets are merged into a larger jet. The electroweak inclusive backgrounds  $Z(\nu\bar{\nu}) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  ( $2.5 \cdot 10^7$  events each) have been produced with a generator level cut  $H_T > 100$  GeV and including up to three additional partons in the matrix element. Finally, the  $\gamma + \text{jets}$  inclusive sample ( $2 \cdot 10^7$  events) has been generated with up to two additional partons and requiring a generator level cut on the leading parton jet transverse momentum  $p_T > 500$  GeV. This background is especially relevant when the expected fraction of photons in the signal jets is particularly high, and thus, the signal jet becomes particularly hard to distinguish from a prompt photon.

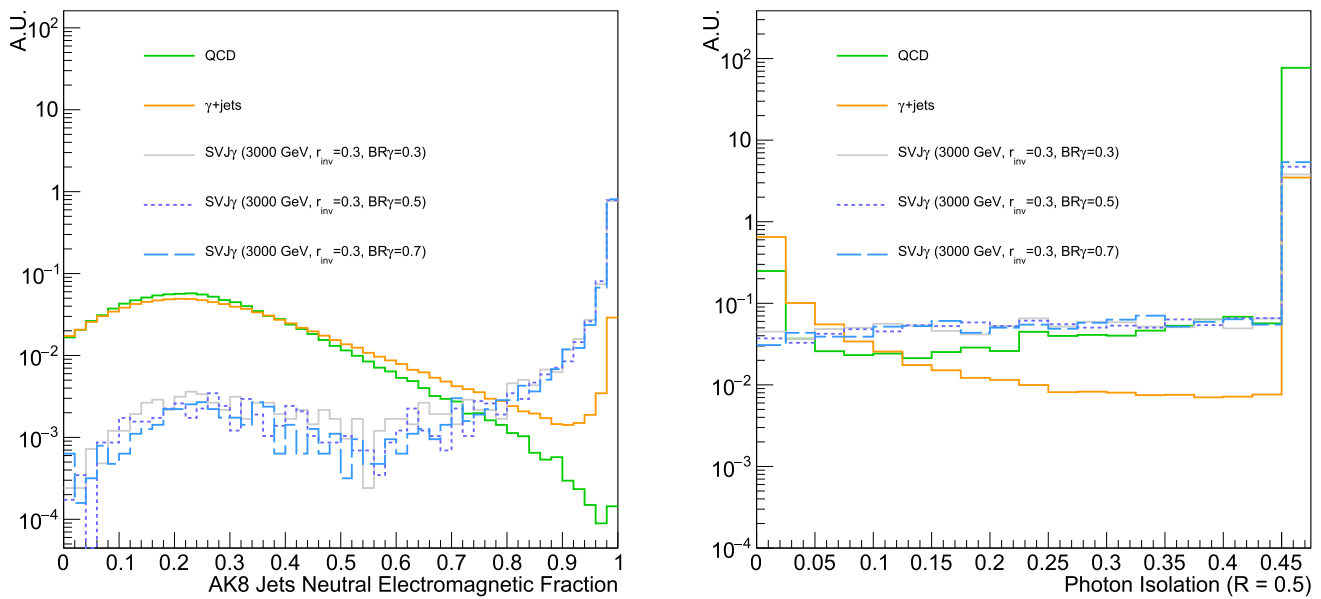
#### 4 Search strategies

In this section, we show how the analysis strategy proposed in published experimental results from resonant hadronic  $\text{SVJ}$  searches [18] is not sensitive to the signature proposed in this Letter. Indeed, one would expect major constraints on  $\text{SVJ}\gamma$  signature from such results. In the following, we report the object and event selections of [18], applied to  $\text{SVJ}\gamma$  events to verify sensitivity of available searches:

- Number of good jets:  $N(j^{\text{AK8}}) \geq 2$ ,  $p_T(j_{1,2}^{\text{AK8}}) > 200$  GeV,  $|\eta(j_{1,2}^{\text{AK8}})| < 2.4$  and passing identification criteria (JetID)
- Jet separation:  $\Delta\eta(j_1^{\text{AK8}}, j_2^{\text{AK8}}) < 1.5$
- Transverse mass selection,<sup>1</sup>:  $M_T(j_1^{\text{AK8}}, j_2^{\text{AK8}}) > 1.5$  TeV
- Jet- $\cancel{E}_T$  separation:  $\Delta\phi_{\min}(\cancel{E}_T, j_{1,2}^{\text{AK8}}) < 0.8$

The jet identification criteria (JetID) are usually applied to select good jets that can be used for analysis [38]. It consists of a set of cuts which aims to reject fake, badly reconstructed and noise jets while retaining a large part of the real jets (around 98 to 99%). The JetID employed by ATLAS and CMS experiments generally makes use of jet energy fractions

<sup>1</sup> The di-jet transverse mass is computed from the 4-vector of the di-jet system  $(E_{T,jj}, \mathbf{p}_{T,jj})$  and  $(\cancel{E}_T, \cancel{E}_T)$  [13]:  $M_T^2 = (E_{T,jj} + \cancel{E}_T)^2 - (\mathbf{p}_{T,jj} + \cancel{E}_T)^2$ . The requirement on the transverse mass allows to select events for which the Run2 jet HT-based triggers employed in [18] are fully efficient.



**Fig. 3** (Left) **a** Distribution of the neutral electromagnetic fraction (NEMF) for the two highest  $p_T$  AK8 jets for different signal benchmarks and QCD,  $\gamma$  + jets backgrounds. (Right) **b** Distribution of photon isolation for  $R_{iso} = 0.5$  for different signal benchmarks and QCD,

$\gamma$  + jets backgrounds. The last bin represents the overflow. The signal distributions are referred to a benchmark mass point  $M_{Z'} = 3$  TeV,  $r_{inv} = 0.3$  and  $BR_\gamma = 0.3, 0.5, 0.7$

and the number of jet constituents to select high-quality jets. Namely, to reject photons resembling jets, typically the fraction of energy in a jet carried by photons (neutral electromagnetic fraction, NEMF) is required to be less than 0.9 (0.99) for  $|\eta| < 2.6$  ( $2.6 < |\eta| < 2.7$ ) [38]. Very high values of NEMF point to energy deposits from photons. Figure 3a shows the expected NEMF distribution for different signals and backgrounds. A clear limitation of analysis employing jets as final state objects in terms of sensitivity to the SVJ $\gamma$  signature comes from such a requirement on the NEMF, which rejects most of the signal. Considering a reference mass point at  $M_{Z'} = 3$  TeV, the signal efficiency decreases from around 14 to 20% (see Table 2, Appendix A) down to 0.4% when applying the NEMF selection (see Table 2, Appendix A). Removing the requirement on the NEMF from the JetID implies higher signal efficiencies while keeping the background efficiencies the same (see Tables 3 and 4). Thus a signal significance improvement is expected (look at Sect. 6).

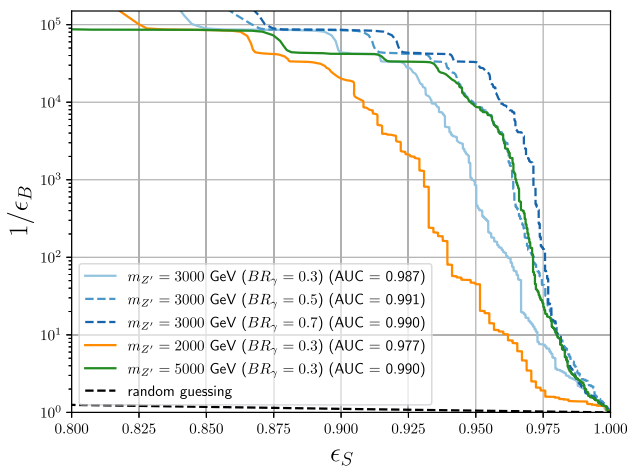
New physics searches looking for jets final states, like the CMS resonant SVJ search, are not the only ones that might be sensitive to SVJ $\gamma$ . Analyses focusing on photons final states, such as the high-mass di-photon searches [39], might show some sensitivity. However, the efficiency of SM photon selections to the SVJ $\gamma$  signature is expected to be very limited. Indeed, photons in SVJ $\gamma$  are expected to be non-isolated due to nearby hadrons coming from the unstable dark bound states decays. Figure 3b shows how the majority of the signal photons lie at high values of the energy deposits around

the photon (*isolation*), while most of the background from prompt photons remains at very low values. Typically, analysis exploiting photons requires isolation (or equivalently, the ratio between the energy deposited in the hadronic and electromagnetic calorimeters) to be small, thus removing the majority of our signal.

From our studies is clear how SM jets and photon identification criteria are suboptimal for identifying SVJ $\gamma$ , pointing to the need for a dedicated identification strategy.

## 5 SVJ $\gamma$ identification

The selection requirements employed by the CMS SVJs search [18] rely on event-level variables (including missing transverse momentum) or basic kinematic properties of the reconstructed jets and missing transverse momentum. Even if this selection rejects the vast majority of the SM background events (Table 3, Appendix A), it does not capture the peculiarities of the SVJ $\gamma$  signature since the expected enrichment in photons is not used anywhere to discriminate against the background. Moreover, removing the NEMF criteria from the JetID, even if keeping similar expected background rates, results in retaining more jets which are faked by a photon, especially  $\gamma$  + jets background, as can be seen from the peak near 1 from Fig. 3a. In order to reduce the QCD and  $\gamma$  + jets backgrounds further and enhance the sensitivity of the search, we employ the tools developed in the



**Fig. 4** Performance of the DNN-based jet tagger scanning on  $M_{Z'}$   $\in$  [2, 3, 5] TeV with fixed  $BR_\gamma = 0.3$  (solid lines), and scanning over  $BR_\gamma \in [0.3, 0.5, 0.7]$  with fixed  $M_{Z'} = 3$  TeV. For all signals we consider a benchmark with  $r_{\text{inv}} = 0.3$ . The signal efficiency is reported on the x axis, while the reciprocal of the background efficiency is reported on the y axis. The areas under the receiver operating characteristic curves (curves in  $\epsilon_B - \epsilon_S$  plane) are reported in the legend. The dashed black line represents random guessing, and all the performance curves are far above this line, thus showing high signal-over-background discrimination

context of jet substructure physics [40–44] to design a novel jet tagging algorithm leveraging the specific features of signal jets. In our jet tagger, we exploit 3 main categories of jet substructure variables: generalized angularities [45] and jet shape observables [46], N-subjettiness [47] and energy fractions. The generalized angularities describe the spatial distribution of the constituents in a jet as well as their momentum flow, and in a similar way, the jet shape observables encode the information about the spatial distribution of the constituents. The N-subjettiness provides a simple way to effectively count the number of subjets inside a given jet. It captures whether the energy flow inside a jet deviates from the one-prong configuration expected to characterize a typical QCD-jet. Finally, the third class of variables includes the energy fractions, which describe the energy carried by specific final states objects (such as electrons, photons, muons and charged hadrons) inside the jet. The list of all 17 input variables used with their definition can be found in Appendix D.

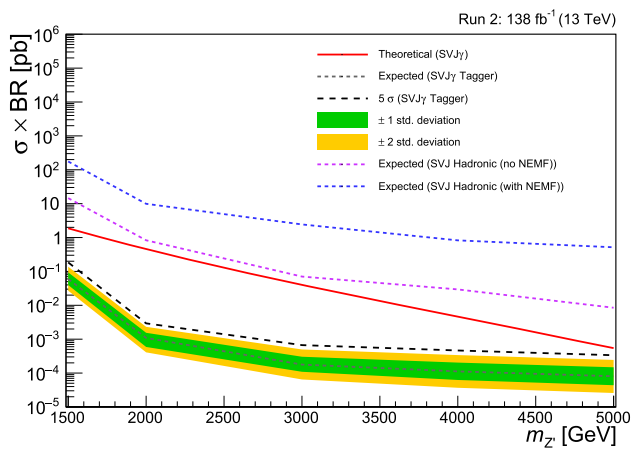
These variables are used as input to a Deep Neural Network (DNN). The DNN is trained using the Pytorch package [48]. The inputs are the two highest  $p_T$  jets from simulated signal and background samples, with the variables described above computed for each jet. The DNN is trained on a mixture of signal models scanning a grid of parameters with  $M_{Z'} = \{1.5, 2, 3, 4, 5\}$  TeV and  $BR_\gamma = \{0.3, 0.5, 0.7\}$ . The other parameters have been fixed to  $\Lambda_v = 20$  GeV and  $r_{\text{inv}} = 0.3$ . Each signal sample is weighted equally and considering the different backgrounds with the same proportions

with respect to one another. More details on the training of the network, the choice of the hyperparameters, and the architecture can be found in Appendix C.

The DNN exhibits strong rejection of jets from the SM background processes. The performance curves of the tagger as a function of model parameters ( $M_{Z'}$ ,  $BR_\gamma$ ) versus all background jets are shown in Fig. 4. As shown in Figs. 7 and 8 in Appendix, the performance of the tagger has been studied for values of  $r_{\text{inv}}$  and  $\Lambda_v$  on which the network was not trained. As expected the performance of the jet tagger diminishes for points of the parameter space on which the network was not trained on, even if the network still have discrimination power also in those cases. Thus, in a real analysis performed at the LHC, exposing the network during the training to a larger signal model parameters grid is advisable and it is expected to improve both the performance and the resilience of the network with respect to multiple signal model variations. Other possible improvements in the direction of being robust against the signal model variation can be achieved by using parametric learning [49]. All this is beyond the scope of this phenomenological study, and it is not further investigated in this Letter.

The three most important variables exploited by the DNN to discriminate the signal from the total background, evaluated using SHapley Additive exPlanations (SHAP) [50,51] with the shap python package [52–55], have been found to be the transverse momentum dispersion ( $p_T D$ ), the photon energy fraction ( $f_\gamma$ ) and the 4-subjettiness ( $\tau_4$ ). The distributions of some of the most relevant DNN input features are shown in Fig. 9 (Appendix E). In particular, we observe that the signal usually reaches higher values for  $p_T D$  compared to QCD,  $t\bar{t}$  + jets and  $Z/W$  + jets, meaning that most of the momentum is carried by few of the signal jets constituents. While, for the  $\gamma$  + jets background, since some jets are just made up of one photon, the  $p_T D$  distribution is expected to peak at 1 and can still be separated from the signal which has a broader spectrum. Regarding the  $f_\gamma$  distribution, as mentioned in Sect. 4, the signal is expected to be distributed at larger values compared to QCD,  $t\bar{t}$  + jets and  $Z/W$  + jets because of the enriched content in photons, while for  $\gamma$  + jets the distribution is expected to peak at 1 for a fraction of the jets, however, the shape of the signal is expected to be different as shown in Fig. 3a. Finally, the signal is expected to reach smaller values for  $\tau_4$  due to the multi-prong substructure, which characterises it with respect to all SM backgrounds.

We choose a working point (WP) corresponding to a threshold of 0.8 on the discriminator output. Jets with a discriminator value higher than 0.8 are labeled as semi-visible. The distributions of the DNN score for signal and background with the chosen selection are reported in Fig. 6 (Appendix D). At this WP, the DNN rejects  $\sim 99.6\%$  of the simulated total background jets, while correctly classifying



**Fig. 5** Expected limits for Run 2 on  $\sigma \times BR(Z' \rightarrow q_v \bar{q}_v)$  for the signal benchmark with  $r_{inv} = 0.3$  and  $BR_\gamma = 0.3$ . The expected exclusion reach for the SVJ $\gamma$  signature from the fully hadronic SVJ inclusive analysis (applying the NEMF requirement in the JetID) (dashed blue line) is compared with both the cut-based strategy removing the NEMF selection (purple dashed line) and the SVJ $\gamma$  tagger-based analysis (grey dashed line), exploiting the jet tagger introduced in Sect. 5. The black dashed line corresponds to the expected 5 $\sigma$  discovery. The red line represents the theoretical prediction computed assuming the coupling of  $Z'$  to SM quarks to be 0.25 and the coupling to dark quarks to be 0.4

$\sim 95.2, 96.6, 97.3\%$  of jets from the benchmark signal models at  $M_{Z'} = 3$  TeV for  $BR_\gamma = 0.3, 0.5, 0.7$ , respectively. In particular, for the chosen WP, the DNN leaves 0.23% of QCD jets, 0.5% of  $t\bar{t}$  jets, 0.6% of jets from  $\gamma +$  jets and 0.9% of jets from electroweak processes.

### 6 Results

We report in this section the sensitivity and exclusion reach for the SVJ $\gamma$  signal. We have estimated the expected exclusion limit at 95% confidence level (CL) for  $\sigma \times BR$  for different  $M_{Z'}$  and  $BR_\tau$  hypotheses performing a binned likelihood template fit of the  $M_T$  spectrum and using the modified frequentist approach  $CL_s$  in the asymptotic approximation [56–58]. The sensitivity reach has been estimated for events selected according to the strategy defined in Sect. 5. We have combined three categories of events according to the jet-level decision of the tagger at chosen optimal WP (as introduced in Sect. 6): 0-tag category, where no jets are tagged as semi-visible, 1-tag category where one jet is tagged as semi-visible and 2-tag category where one jet is tagged as semi-visible.

We report the results for a benchmark point with  $r_{inv} = 0.3$  and  $BR_\gamma = 0.3$ . The LHC with full Run 2 collision data ( $\sqrt{s} = 13$  TeV,  $\mathcal{L}_{int} = 138 \text{ fb}^{-1}$ ) and following the tagger-based approach proposed here is expected to claim the discovery (exclusion) of a hypothetical  $Z'$  boson in the full mass range considered for a benchmark coupling of  $Z'$  to SM quarks equal to 0.25, and the coupling to dark quarks

equal to 0.4. More details are shown in Fig. 5. We observe that the sensitivity to the signal slightly increases with higher values of  $BR_\gamma$  due to the higher signal efficiency and tagger discrimination power. Furthermore, Fig. 5, shows how the inclusive CMS SVJ search [18] has no sensitivity to SVJ $\gamma$  signals, while the jet tagger proposed in this Letter allows probing values of the coupling of the  $Z'$  portal to SM quarks even smaller than the benchmark 0.25, for the entire  $M_{Z'}$  mass range tested.

### 7 Conclusions

This Letter proposes a novel experimental signature and relevant search strategy for the discovery of confining hidden sectors that give rise to SVJs with non-isolated prompt photon pairs. We have introduced a simplified model based on two messenger bosons, namely a  $Z'$  and an ALP, allowing the resonant production of SVJs enriched in photons. The phase space of the SVJ $\gamma$  signal is expected to be loosely constrained by the existing searches due to the tight NEMF and isolation requirements used in the standard jets and photons identification criteria. We have shown that the fully hadronic SVJ s-channel search has poor sensitivity to the SVJ $\gamma$  signature, and removing (or loosening) the NEMF requirement might not be enough to get the necessary sensitivity to discover the signal. Thus, we developed a DNN-based jet tagging algorithm based on jet substructure features to enhance the signal-to-background separation and boost the sensitivity of the proposed analysis strategy. The proposed search can claim a discovery (exclusion) of the  $Z'$  mediator in the full mass range considered with  $\mathcal{L}_{int} = 138 \text{ fb}^{-1}$ , if the jet tagging algorithm proposed here is applied.

Finally, the model proposed in this Letter also allows for displaced signatures in cases of light–dark-bound states. This would lead to signatures with additional displaced tracks and vertices inside the signal jets (similar to emerging jets [59]) and non-prompt, non-isolated photons. This signature requires a dedicated strategy that is not addressed in this Letter, but might be of interest for future works.

**Acknowledgements** C. Cazzaniga and A. de Cosa are supported by the Swiss National Science Foundation (SNFS) under the SNSF Eccellenza program (PCEFP2\_186878).

**Data Availability Statement** This manuscript has no associated data. [Authors’ comment: Data will not be published but can be shared under request to the authors.]

**Code Availability Statement** This manuscript has no associated code/software. [Authors’ comment: Code/Software will not be published but can be shared under request to the authors.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you

give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Funded by SCOAP<sup>3</sup>.

### Appendix A Event selection

The event selections used in this work are shown in the first columns of Tables 1 and 3, respectively for a benchmark signal and the backgrounds. This is based on a previous CMS study [18], and it is updated removing the requirement on the NEMF in the JetID. The last three columns show the signal efficiencies for different values of BR<sub>γ</sub> and a benchmark mass point M<sub>Z'</sub> = 3 TeV. In Table 2 for a benchmark signal and Table 4 for the backgrounds, the event selection including the NEMF requirement as JetID are shown.

### Appendix B DNN jet tagger input features

Here we report the definition of the input features used for the DNN training. All input features have been computed using AK8 jets and their constituents. The first category of variables we used are the generalized angularities, and they are defined from the constituents  $i \in \{1, \dots, N\}$  carrying

momentum fraction  $z_i$  inside a jet of cone size  $R$  as:

$$\lambda_{\beta}^{\kappa} = \sum_i z_i^{\kappa} \left( \frac{\Delta R_{i,\text{jet}}}{R} \right)^{\beta} \tag{B.1}$$

where the sum is over the jet constituents. From Eq. (B.1), the input features we used for the training of the network are the following:

- girth ( $\lambda_0^2$ )
- LHA ( $\lambda_1^1$ )
- thrust ( $\lambda_2^1$ )
- transverse momentum dispersion ( $p_T D = \lambda_0^2$ )
- multiplicity ( $\lambda_0^0$ )

The first three features describe the spatial distribution of the constituents in jets, while  $p_T D$  encodes how the momentum is distributed between the constituents of the jet.

We supplement such set of features with further jet shape variables known as axes. The shape of the jet can be approximated by an ellipse in the  $\eta - \phi$  plane. The major and minor axes are the two principal components of this ellipse and are defined from the following symmetric matrix  $M$ :

$$M = \begin{bmatrix} \sum_i p_{T,i}^2 (\Delta\eta_i)^2 & -\sum_i p_{T,i}^2 \Delta\eta_i \Delta\phi_i \\ -\sum_i p_{T,i}^2 \Delta\eta_i \Delta\phi_i & \sum_i p_{T,i}^2 (\Delta\phi_i)^2 \end{bmatrix} \tag{B.2}$$

where the sum runs over all constituents of the jet and  $\Delta\eta$ ,  $\Delta\phi$  are the differences in  $\eta$  and  $\phi$  with respect to the jet axis. The major and minor axes are defined from the eigenvalues

**Table 1** Event selections (without the NEMF requirement in the JetID for AK8 jets) applied to SVJ<sub>γ</sub> signal for a benchmark mass point M<sub>Z'</sub> = 3 TeV and varying the branching into photons BR<sub>γ</sub>

Selection	Signal efficiency (%)		
BR <sub>γ</sub> :	0.7	0.5	0.3
$N(\text{AK8 jets with NEMF}) \geq 2$	78.00	69.34	52.71
$\Delta\eta(j_1, j_2) < 1.5$	52.70	47.74	37.59
$M_T > 1500 \text{ GeV}$	42.66	35.62	23.84
$R_T > 0.15$	23.38	22.12	16.79
$\Delta\phi_{\min}(\mathcal{E}_T, j) < 0.8$	21.04	19.43	13.98

**Table 2** Event selections (including the NEMF requirement in the JetID for AK8 jets) applied to SVJ<sub>γ</sub> signal for a benchmark mass point M<sub>Z'</sub> = 3 TeV and varying the branching into photons BR<sub>γ</sub>. The JetID applied here consists only of a requirement on the NEMF [NEMF is required to be less than 0.9 (0.99) for  $|\eta| < 2.6$  ( $2.6 < |\eta| < 2.7$ )]

Selection	Signal efficiency (%)		
BR <sub>γ</sub> :	0.7	0.5	0.3
$N(\text{AK8 jets without NEMF}) \geq 2$	1.60	1.51	1.54
$\Delta\eta(j_1, j_2) < 1.5$	1.00	1.05	1.08
$M_T > 1500 \text{ GeV}$	0.74	0.74	0.73
$R_T > 0.15$	0.47	0.50	0.60
$\Delta\phi_{\min}(\mathcal{E}_T, j) < 0.8$	0.39	0.38	0.38



**Table 3** Event selections (without the NEMF requirement in the JetID for AK8 jets) applied to the backgrounds

Selection	Background efficiency (%)				
	QCD	$t\bar{t}$ + jets	$Z(\nu\bar{\nu})$ + jets	$W(\ell\nu)$ + jets	$\gamma$ + jets
$N(\text{AK8 jets with NEMF}) \geq 2$	98.17	7.19	1.04	1.59	98.55
$\Delta\eta(j_1, j_2) < 1.5$	66.53	5.33	0.67	1.11	68.40
$M_T > 1500 \text{ GeV}$	14.98	0.15	0.04	0.03	17.37
$R_T > 0.15$	0.69	0.03	0.02	0.01	0.81
$\Delta\phi_{\min}(\cancel{E}_T, j) < 0.8$	0.68	0.03	0.01	0.01	0.79

**Table 4** Event selections (including the NEMF requirement in the JetID for AK8 jets) applied to the backgrounds. The JetID applied here consists only of a requirement on the NEMF [NEMF is required to be less than 0.9 (0.99) for  $\eta < 2.6$  ( $2.6 < \eta < 2.7$ )]

Selection	Background efficiency (%)				
	QCD	$t\bar{t}$ + jets	$Z(\nu\bar{\nu})$ + jets	$W(\ell\nu)$ + jets	$\gamma$ + jets
$N(\text{AK8 jets with NEMF}) \geq 2$	98.03	7.19	1.04	1.57	89.16
$\Delta\eta(j_1, j_2) < 1.5$	66.44	5.33	0.67	1.10	60.86
$M_T > 1500 \text{ GeV}$	14.96	0.15	0.03	0.03	15.68
$R_T > 0.15$	0.69	0.03	0.02	0.01	0.76
$\Delta\phi_{\min}(\cancel{E}_T, j) < 0.8$	0.68	0.03	0.01	0.01	0.75

$\lambda_1$  and  $\lambda_2$  of  $M$  as:

$$\sigma_{\text{minor,major}} = \sqrt{\frac{\lambda_{1,2}}{\sum_i p_{T,i}^2}} \tag{B.3}$$

The average axis can be defined from  $\sigma_{\text{minor,major}}$  as:

$$\sigma_{\text{avg}} = \sqrt{\sigma_{\text{minor}}^2 + \sigma_{\text{major}}^2} \tag{B.4}$$

The second set of features we feed into the DNN are the N-subjettiness [47]. The N-subjettiness  $\tau_N^\beta$  are designed to count the number of subjects inside a jet. In specific, N-subjettiness is defined as:

$$\tau_N^\beta = \sum_i p_{T,i} \min(R_{1,i}^\beta, R_{2,i}^\beta, \dots, R_{N,i}^\beta) \tag{B.5}$$

where the sum is over the jet constituents, and  $R_{N,i}^\beta$  is the distance between the  $N$ th subjet and the  $i$ th constituent of the jet.  $\tau_N^\beta$  measures departure from  $N$ -parton energy flow: if a jet has  $N$  subjects,  $\tau_N^\beta$  should be much larger than  $\tau_N^\beta$ . For our DNN training we use N-subjettiness with  $N \in \{1, \dots, 5\}$  and  $\beta = 1$ .

Finally, the third set of variables we used are energy fractions. These quantities can be defined for different jet constituents, such as electrons ( $e$ ), muons ( $\mu$ ), photons ( $\gamma$ ) and charged hadrons ( $h_{ch}$ ). Fixed a specific final state object  $\tilde{c}$  among those listed above, the energy fraction relative to that object is defined as:

$$f_{\tilde{c}} = \sum_i \frac{E_{\tilde{c},i}}{E_i} \quad \tilde{c} \in \{e, \mu, \gamma, h_{ch}\} \tag{B.6}$$

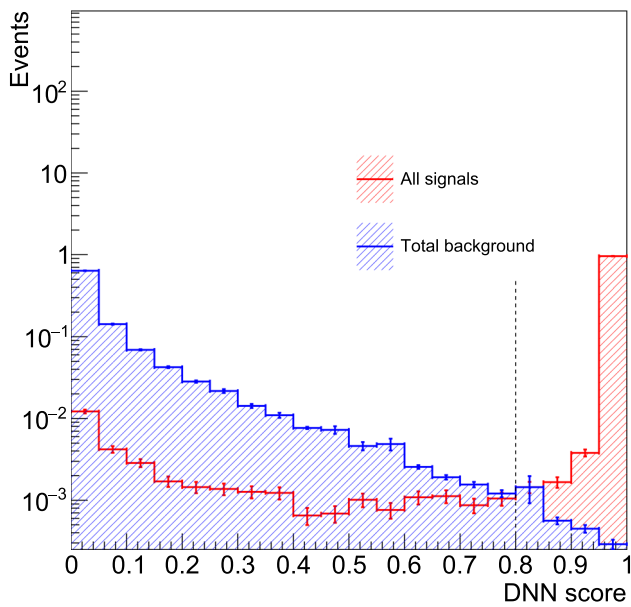
For the training of the DNN tagger, we use as input features the following energy fractions:  $f_e, f_\mu, f_\gamma, f_{h_{ch}}$ .

### Appendix C DNN architecture and training

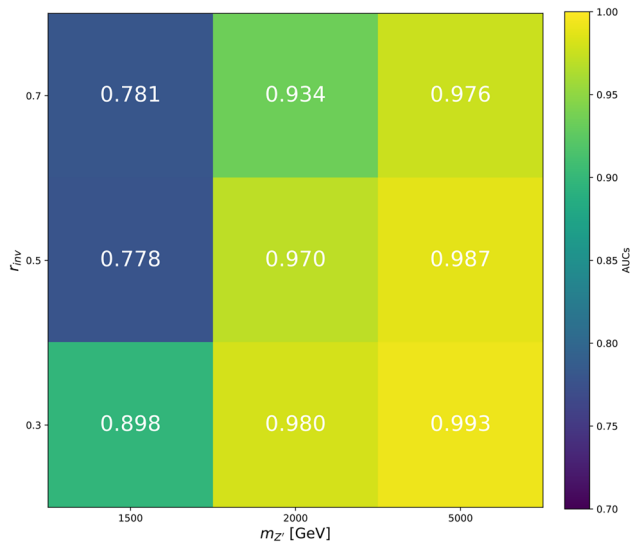
The Deep Neural Network has been implemented using the Pytorch package [48]. The DNN takes 17 input features, and consists of 2 hidden linear layers with 34 nodes each. Each linear layer is followed by a batch normalization layer and an activation function. The `relu` activation function has been chosen for the two hidden layers, and the `sigmoid` for the output layer of the network. For the output layer, between the batch normalization layer and the sigmoid activation function, a `dropout` layer has been added in order to prevent overfitting. The value of the dropout has been chosen to be 0.2. The DNN has been trained to reach converge for 300 epochs with a batch size of 256 and a learning rate of 0.001. The Adam optimizer has been employed for the gradient descent.

### Appendix D DNN performance

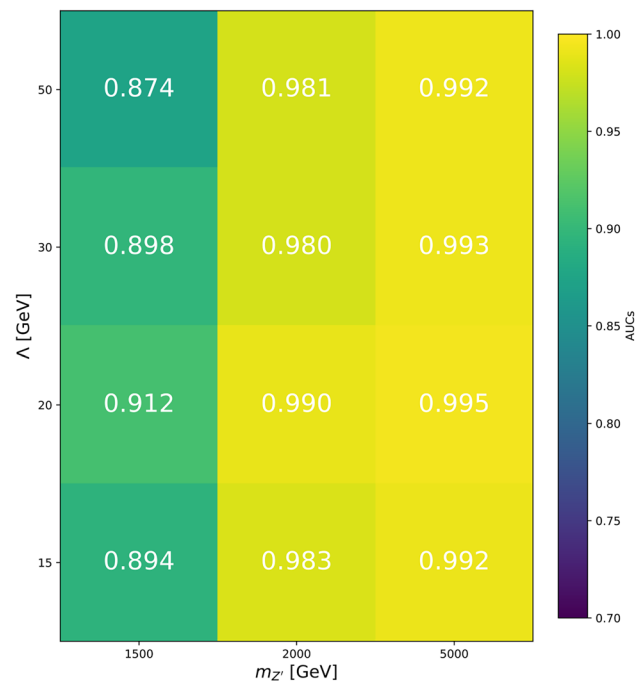
The distributions of the DNN scores for signal and background on the test dataset together with the working point chosen is shown in Fig. 6 (Figs. 7, 8).



**Fig. 6** DNN score distribution for signal (red) and background (blue) test datasets. The vertical black dashed line represents the cut value chosen to define a background jet (DNN score < 0.8) against a signal jet (DNN score > 0.8)



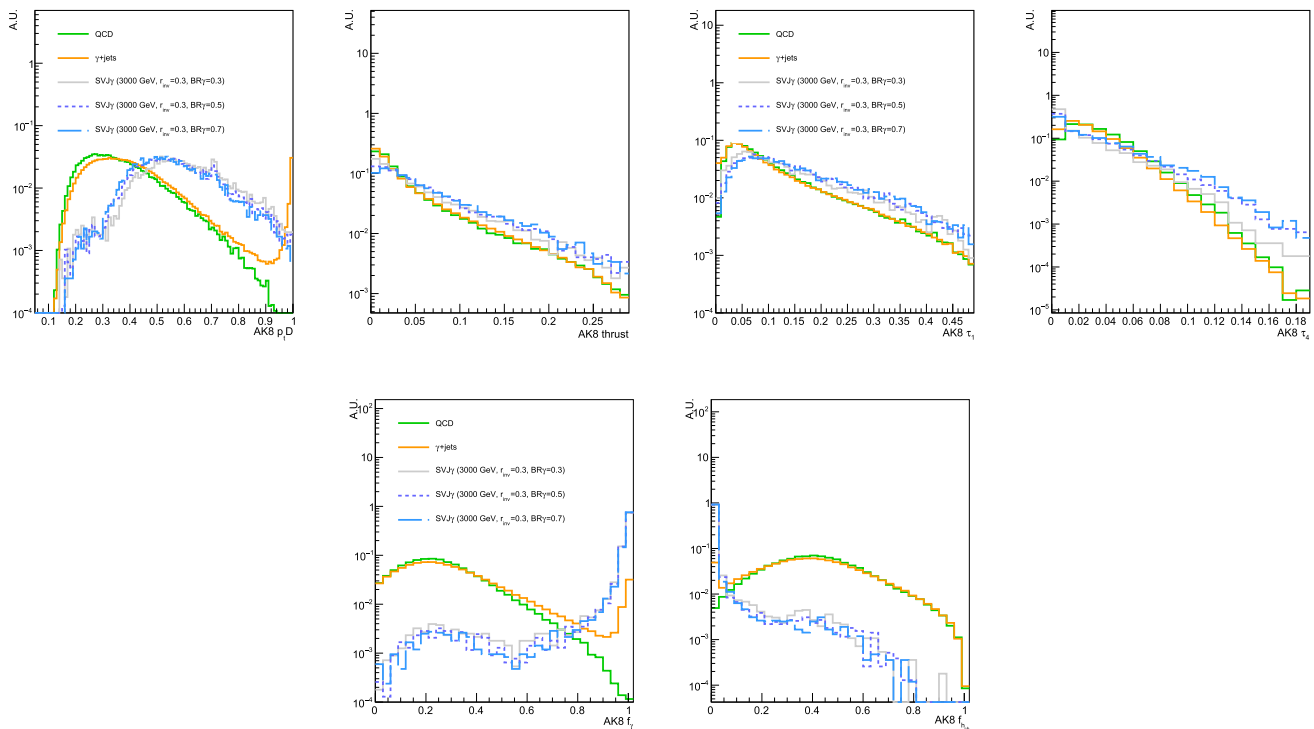
**Fig. 7** Performance of the jet tagger for different signal hypotheses changing the parameters  $M_{Z'}$  and  $r_{inv}$ . The value of  $BR_\gamma$  has been fixed to 0.5. The colour gradient and numbers represent the AUC of the ROC curve for the given signal hypotheses



**Fig. 8** Performance of the jet tagger for different signal hypotheses changing the parameters  $M_{Z'}$  and  $\Lambda_\nu$ . The value of  $BR_\gamma$  has been fixed to 0.5. The colour gradient and numbers represent the AUC of the ROC curve for the given signal hypothesis

### Appendix E DNN input features

The distributions of the two most discriminating input features based on SHAP used to train the DNN for each of the three classes are shown in Fig. 9.



**Fig. 9** (Top left) Distribution of the two most discriminating generalized angularities:  $p_T D$  and thrust. (Top right) Distribution of the two most discriminating N-subjettiness  $\tau_1$  and  $\tau_4$ . (Bottom) Distribution of the two most discriminating energy fractions:  $f_\gamma$  and  $f_{hch}$ . The signal

distributions are referred to a benchmark mass point  $M_{Z'} = 3 \text{ TeV}$ ,  $r_{inv} = 0.3$  and  $BR_\gamma = 0.3, 0.5, 0.7$ . The backgrounds plotted are  $\gamma + \text{jets}$  (orange line) and QCD (green line), while all the signals are merged. All distributions shown are normalised to one

## References

- G. Bertone, D. Hooper, J. Silk, Phys. Rep. **405**, 279 (2005). <https://doi.org/10.1016/j.physrep.2004.08.031>
- N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A.J. Banday, R.B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed et al., Astron. Astrophys. **641**, A6 (2020). <https://doi.org/10.1051/0004-6361/201833910>
- B.W. Lee, S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977). <https://doi.org/10.1103/PhysRevLett.39.165>
- G. Jungman, M. Kamionkowski, K. Griest, Phys. Rep. **267**, 195 (1996). [https://doi.org/10.1016/0370-1573\(95\)00058-5](https://doi.org/10.1016/0370-1573(95)00058-5)
- M.J. Strassler, K.M. Zurek, Phys. Lett. B **651**, 374 (2007). <https://doi.org/10.1016/j.physletb.2007.06.055>
- D. Spier Moreira Alves, S.R. Behbahani, P. Schuster, J.G. Wacker, JHEP **06**, 113 (2010). [https://doi.org/10.1007/JHEP06\(2010\)113](https://doi.org/10.1007/JHEP06(2010)113)
- H.M. Lee, M.S. Seo, Phys. Lett. B **748**, 316 (2015). <https://doi.org/10.1016/j.physletb.2015.07.013>
- S. Okawa, M. Tanabashi, M. Yamanaka, Phys. Rev. D **95**(2), 023006 (2017). <https://doi.org/10.1103/PhysRevD.95.023006>
- Y. Hochberg, E. Kuflik, H. Murayama, Phys. Rev. D **99**(1), 015005 (2019). <https://doi.org/10.1103/PhysRevD.99.015005>
- H. Beausnes, E. Bertuzzo, G. Grilli Di Cortona, JHEP **04**, 118 (2019). [https://doi.org/10.1007/JHEP04\(2019\)118](https://doi.org/10.1007/JHEP04(2019)118)
- E. Bernreuther, F. Kahlhoefer, M. Krämer, P. Tunney, JHEP **01**, 162 (2020). [https://doi.org/10.1007/JHEP01\(2020\)162](https://doi.org/10.1007/JHEP01(2020)162)
- H. Beausnes, G. Grilli di Cortona, JHEP **02**, 196 (2020). [https://doi.org/10.1007/JHEP02\(2020\)196](https://doi.org/10.1007/JHEP02(2020)196)
- T. Cohen, M. Lisanti, H.K. Lou, Phys. Rev. Lett. **115**(17), 171804 (2015). <https://doi.org/10.1103/PhysRevLett.115.171804>
- T. Cohen, M. Lisanti, H.K. Lou, S. Mishra-Sharma, JHEP **11**, 196 (2017). [https://doi.org/10.1007/JHEP11\(2017\)196](https://doi.org/10.1007/JHEP11(2017)196)
- H. Beausnes, E. Bertuzzo, G. Grilli Di Cortona, Z. Tabrizi, JHEP **08**, 030 (2018). [https://doi.org/10.1007/JHEP08\(2018\)030](https://doi.org/10.1007/JHEP08(2018)030)
- C. Cazzaniga, A. de Cosa, Eur. Phys. J. C **82**(9), 793 (2022). <https://doi.org/10.1140/epjc/s10052-022-10775-2>
- H. Beausnes, C. Cazzaniga, A. de Cosa, C. Doglioni, T. Fitschen, G.G. di Cortona, Z. Zhou, Eur. Phys. J. C (2023). <https://doi.org/10.1140/epjc/s10052-023-11775-6>
- A. Tumasyan et al., JHEP **06**, 156 (2022). [https://doi.org/10.1007/JHEP06\(2022\)156](https://doi.org/10.1007/JHEP06(2022)156)
- Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS. Tech. rep., CERN, Geneva (2022). <https://cds.cern.ch/record/2815284>
- P.J. Fox, J. Liu, D. Tucker-Smith, N. Weiner, Phys. Rev. D **84**, 115006 (2011). <https://doi.org/10.1103/PhysRevD.84.115006>
- S. Knapen, J. Shelton, D. Xu, Phys. Rev. D **103**(11), 115013 (2021). <https://doi.org/10.1103/PhysRevD.103.115013>
- S.D. Ellis, T.S. Roy, J. Scholtz, Phys. Rev. D **87**(1), 014015 (2013). <https://doi.org/10.1103/PhysRevD.87.014015>
- A.M. Sirunyan et al., JHEP **06**, 120 (2018). [https://doi.org/10.1007/JHEP06\(2018\)120](https://doi.org/10.1007/JHEP06(2018)120)
- G. Aad et al., Phys. Lett. B **796**, 68 (2019). <https://doi.org/10.1016/j.physletb.2019.07.016>
- A. Tumasyan et al., JHEP **07**, 020 (2023). [https://doi.org/10.1007/JHEP07\(2023\)020](https://doi.org/10.1007/JHEP07(2023)020)
- A.M. Sirunyan et al., Phys. Rev. D **100**(5), 052003 (2019). <https://doi.org/10.1103/PhysRevD.100.052003>
- C. Bierlich et al., (2022). <https://doi.org/10.48550/ARXIV.2203.11601>

28. L. Carloni, J. Rathsmann, T. Sjostrand, *JHEP* **04**, 091 (2011). [https://doi.org/10.1007/JHEP04\(2011\)091](https://doi.org/10.1007/JHEP04(2011)091)
29. L. Carloni, T. Sjostrand, *JHEP* **09**, 105 (2010). [https://doi.org/10.1007/JHEP09\(2010\)105](https://doi.org/10.1007/JHEP09(2010)105)
30. J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, M. Selvaggi, *JHEP* **02**, 057 (2014). [https://doi.org/10.1007/JHEP02\(2014\)057](https://doi.org/10.1007/JHEP02(2014)057)
31. M. Cacciari, G.P. Salam, G. Soyez, *JHEP* **04**, 063 (2008). <https://doi.org/10.1088/1126-6708/2008/04/063>
32. M. Cacciari, G.P. Salam, G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012). <https://doi.org/10.1140/epjc/s10052-012-1896-2>
33. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, *JHEP* **07**, 079 (2014). [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079)
34. R.D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N.P. Hartland, J. Rojo, *Nucl. Phys. B* **877**, 290 (2013). <https://doi.org/10.1016/j.nuclphysb.2013.10.010>
35. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, G. Watt, *Eur. Phys. J. C* **75**, 132 (2015). <https://doi.org/10.1140/epjc/s10052-015-3318-8>
36. A. Boveia et al., *Phys. Dark Univ.* **27**, 100365 (2020). <https://doi.org/10.1016/j.dark.2019.100365>
37. G. Albouy et al., *Eur. Phys. J. C* **82**(12), 1132 (2022). <https://doi.org/10.1140/epjc/s10052-022-11048-8>
38. (2024). <https://cds.cern.ch/record/2899451>
39. V. Khachatryan et al., *Phys. Rev. Lett.* **117**(5), 051802 (2016). <https://doi.org/10.1103/PhysRevLett.117.051802>
40. G. Brooijmans, High pT hadronic top quark identification. Tech. rep., CERN, Geneva (2008). <https://cds.cern.ch/record/1077731>
41. J.M. Butterworth, J.R. Ellis, A.R. Raklev, *JHEP* **05**, 033 (2007). <https://doi.org/10.1088/1126-6708/2007/05/033>
42. J.M. Butterworth, A.R. Davison, M. Rubin, G.P. Salam, *Phys. Rev. Lett.* **100**, 242001 (2008). <https://doi.org/10.1103/PhysRevLett.100.242001>
43. J. Thaler, K. Van Tilburg, *JHEP* **02**, 093 (2012). [https://doi.org/10.1007/JHEP02\(2012\)093](https://doi.org/10.1007/JHEP02(2012)093)
44. D.E. Kaplan, K. Rehermann, M.D. Schwartz, B. Tweedie, *Phys. Rev. Lett.* **101**, 142001 (2008). <https://doi.org/10.1103/PhysRevLett.101.142001>
45. A.J. Larkoski, J. Thaler, W.J. Waalewijn, *JHEP* **11**, 129 (2014). [https://doi.org/10.1007/JHEP11\(2014\)129](https://doi.org/10.1007/JHEP11(2014)129)
46. J. Gallicchio, M.D. Schwartz, *Phys. Rev. Lett.* **107**, 172001 (2011). <https://doi.org/10.1103/PhysRevLett.107.172001>
47. J. Thaler, K. Van Tilburg, *JHEP* **03**, 015 (2011). [https://doi.org/10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015)
48. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimselshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Rai-son, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, (2019). [arxiv:1912.01703](https://arxiv.org/abs/1912.01703)
49. P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, D. Whiteson, *Eur. Phys. J. C* **76**(5), 235 (2016). <https://doi.org/10.1140/epjc/s10052-016-4099-4>
50. E. Strumbelj, I. Kononenko, *Knowl. Inf. Syst.* **41**, 647 (2014). <https://api.semanticscholar.org/CorpusID:2449098>
51. S.M. Lundberg, S.I. Lee, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2017), NIPS'17, pp. 4768–4777
52. S.M. Lundberg, S.I. Lee, in *Advances in Neural Information Processing Systems 30*, ed. by I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Curran Associates, Inc., 2017), pp. 4765–4774. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
53. S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, *Nat. Mach. Intell.* **2**(1), 2522 (2020)
54. R. Mitchell, E. Frank, G. Holmes, GPUtreeShap: massively parallel exact calculation of Shap scores for tree ensembles. (2022). <https://arxiv.org/abs/2010.13972>
55. S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D.E. Liston, D.K.W. Low, S.F. Newman, J. Kim et al., *Nat. Biomed. Eng.* **2**(10), 749 (2018)
56. T. Junk, *Nucl. Instrum. Methods A* **434**, 435 (1999). [https://doi.org/10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2)
57. A.L. Read, *J. Phys. G* **28**, 2693 (2002). <https://doi.org/10.1088/0954-3899/28/10/313>
58. G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011). <https://doi.org/10.1140/epjc/s10052-011-1554-0>. [Erratum: *Eur. Phys. J. C* **73**, 2501 (2013)]
59. P. Schwaller, D. Stolarski, A. Weiler, *J. High Energy Phys.* (2015). [https://doi.org/10.1007/jhep05\(2015\)059](https://doi.org/10.1007/jhep05(2015)059)