Machine Learning: Science and Technology

PAPER • **OPEN ACCESS**

A deep neural network to search for new long-lived particles decaying to jets

To cite this article: The CMS Collaboration 2020 *Mach. Learn.: Sci. Technol.* **1** 035012

IOP Publishing *Mach. Learn.: Sci. Technol.* **1** (2020) 035012 https://doi.org/10.1088/2632-2153/ab9023

MACHINE
LEARNING
Science and Technology

RECEIVED
22 January 2020

REVISED
25 March 2020

ACCEPTED FOR PUBLICATION
5 May 2020

PUBLISHED
18 August 2020

A deep neural network to search for new long-lived particles decaying to jets

The CMS Collaboration
CERN, Geneva, Switzerland

E-mail: cms-publication-committee-chair@cern.ch

Keywords: CMS, split SUSY, long-lived particles, deep neural network

Abstract

A tagging algorithm to identify jets that are significantly displaced from the proton-proton (pp) collision region in the CMS detector at the LHC is presented. Displaced jets can arise from the decays of long-lived particles (LLPs), which are predicted by several theoretical extensions of the standard model. The tagger is a multiclass classifier based on a deep neural network, which is parameterised according to the proper decay length $c\tau_0$ of the LLP. A novel scheme is defined to reliably label jets from LLP decays for supervised learning. Samples of pp collision data, recorded by the CMS detector at a centre-of-mass energy of 13 TeV, and simulated events are used to train the neural network. Domain adaptation by backward propagation is performed to improve the simulation modelling of the jet class probability distributions observed in pp collision data. The potential performance of the tagger is demonstrated with a search for long-lived gluinos, a manifestation of split supersymmetric models. The tagger provides a rejection factor of 10 000 for jets from standard model processes, while maintaining an LLP jet tagging efficiency of 30%–80% for gluinos with 1 mm$\leq c\tau_0 \leq$ 10 m. The expected coverage of the parameter space for split supersymmetry is presented.

1. Introduction

Machine-learned algorithms are routinely deployed to perform event reconstruction, particle identification, event classification, and other tasks [1] when analysing data samples recorded by experiments at the CERN LHC. Machine learning techniques have been widely adopted to classify a jet, a collimated spray of particles that originate from the hadronisation of a parton, according to the underlying flavour of the initial parton [2]. For example, jets that originate from the hadronisation of b quarks (b jets) exhibit characteristic experimental signatures that can be exploited by dedicated algorithms to identify b jets, in a procedure known as b tagging. The b hadrons, with proper lifetimes of $\mathcal{O}(10^{-12}$ s), typically travel distances of approximately 1–10 mm, depending on their momenta, before decaying. As a result, charged particle tracks in jets can originate from one or more common vertices that may be displaced with respect to the proton-proton (pp) collision region. Furthermore, the impact parameter of each track, defined as the spatial distance between the originating pp collision and the track at its point of closest approach, can have a significant nonzero value. The ATLAS [3] and CMS [4] Collaborations have developed numerous algorithms based on boosted decision trees or neural networks to identify b jets [5, 6] using the aforementioned and other high-level engineered features. The latest b tagging algorithm developed by the CMS Collaboration is the DeepJet tagger [7, 8], which is a multiclass classifier that discriminates between jets originating from the hadronisation of heavy- (b or c) or light-flavour (u, d, and s) quarks or gluons (g) with unprecedented performance. The algorithm is based on a deep neural network (DNN) that exploits particle-level information, as well as jet-level engineered features used in preceding b tagging algorithms [5, 6].

Various theoretical extensions of the standard model (SM) [9–15] predict the existence of long-lived particles (LLPs) with a proper lifetime τ_0 that can be very different from those of known SM particles. Consequently, the production and decay of LLPs at the LHC could give rise to atypical experimental signatures. These possibilities have led to the development of a broad search programme at the LHC, based around LLP simplified models [16, 17] and novel reconstruction techniques. A comprehensive review of LLP searches at the LHC can be found in reference [18].

In this paper, we present the novel application of a DNN to tag (i.e. identify) a jet originating from the decay of an LLP (LLP jet), both in the presence or absence of a displaced vertex of charged particle tracks. The DNN is trained and evaluated using a range of signal hypotheses comprising simplified models of split supersymmetry (SUSY) [9, 10] with R parity [19] conservation. These models assume the production of gluino ($\tilde{\mathrm{g}}$) pairs. The gluino is a long-lived state that decays to a quark-antiquark ($\mathrm{q\bar{q}}$) pair and a weakly interacting and massive neutralino ($\tilde{\chi}_1^0$), which is the lightest SUSY particle and a dark matter candidate. Simplified models of SUSY that yield LLPs are widely used as a benchmark for searches in final states containing jets and an apparent imbalance in transverse momentum, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ [20–31]. The search in reference [27] does not explicitly target LLPs, but its inclusive approach provides sensitivity over a large range of lifetimes. This search is used later as a performance benchmark for the methods presented in this paper.

The LLP jet tagger is inspired by the DeepJet approach, albeit with significant modifications. The DNN extends the multiclass classification scheme of the DeepJet algorithm to accommodate the LLP jet class. A procedure to reliably label LLP jets using generator-level information from the Monte Carlo (MC) programs is defined. The experimental signature for an LLP jet depends strongly on the proper decay length $c\tau_0$. Hence, a parameterised approach [32] is adopted by using $c\tau_0$ as an external parameter to the DNN, which permits hypothesis testing using a single network for models with values of $c\tau_0$ that span several orders of magnitude. Furthermore, the jet momenta depend strongly on the masses $m_{\tilde{\mathrm{g}}}$ and $m_{\tilde{\chi}_1^0}$, and particularly the mass difference $m_{\tilde{\mathrm{g}}} - m_{\tilde{\chi}_1^0}$. The DNN training is performed with simulated event samples drawn from the full mass parameter space of interest to ensure a broad generalisation and optimal performance over a range of jet momenta. The performance of the tagger is also quantified for simplified models of SUSY with gauge-mediated SUSY breaking (GMSB) [12] and weak R-parity violation (RPV) [13, 33]. Domain adaptation by backward propagation of errors [34] is incorporated into the network architecture to achieve similar classification performance in simulation and pp collision data, thus mitigating differences between the two domains of simulation and data. Differences can arise because of, for example, the limited precision of the simulation. The pp collision data were recorded by the CMS detector at a centre-of-mass energy of 13 TeV. The sample, recorded in 2016, corresponds to an integrated luminosity of 35.9 fb^{-1}.

This paper is organised as follows. section 2 describes the CMS detector and event reconstruction algorithms. sections 3 and 4 describe, respectively, the event samples and simulation software packages used in this study. section 5 describes the LLP jet tagger. sections 6 and 7 demonstrate, respectively, the validation of the tagger using control samples of pp collision data and its performance based on simulated samples. section 8 presents the expected performance of the tagger in a search for long-lived gluinos, as well as an in situ determination of a correction to the signal efficiency of the LLP jet tagger. section 9 provides a summary of this work.

2. The CMS detector and event reconstruction

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity (η) coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionisation chambers embedded in the steel flux-return yoke outside the solenoid. At $\eta = 0$, the outer radial dimension of the barrel section of the tracker, ECAL, HCAL, and muon subdetector is 1.3, 1.8, 3.0, and 7.4 m, respectively. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in reference [4].

The candidate vertex with the largest value of summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex, where p_{T} is the transverse momentum. Here, the physics objects are the jets, clustered using the jet finding algorithm [35, 36] with the tracks assigned to candidate vertices as inputs, and the associated $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, taken as the negative vector p_{T} sum of those jets.

The particle-flow (PF) algorithm [37] aims to reconstruct and identify each particle (PF candidate) in an event, with an optimised combination of all subdetector information. In this process, the identification of the particle type (photon, electron, muon, charged or neutral hadron) plays an important role in the determination of the particle direction and energy. Photons [38] are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL. Electrons [39] are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to the extrapolation of this track to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. Muons [40] are identified as tracks in the central tracker consistent with either a track or several hits in the muon system, and associated with calorimeter deposits compatible with the muon hypothesis. Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons. Finally, neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as a combined ECAL and HCAL energy excess with respect to the expected charged hadron energy deposit. The inclusive vertex finding algorithm [41] is used as the standard secondary vertex reconstruction algorithm, which uses all reconstructed tracks in the event with $p_{\mathrm{T}} > 0.8$ GeV and a longitudinal impact parameter of greater than 0.3 cm.

The PF algorithm is able to reconstruct reliably particle candidates with large displacements. In the inner tracker volume, charged particles are identified through the presence of an associated track. The later iterations of an iterative tracking procedure [42] explicitly target the most displaced tracks. At larger displacements, beyond the tracking volume, the energy and direction of displaced particles are solely determined from measurements in the calorimeter systems. In the case that an LLP decay occurs in the muon systems, a collimated spray of muon-track stubs is reconstructed. These stubs are ignored by the PF algorithm, as they cannot be reliably distinguished from hadronic showers that are not completely contained in the calorimeters (so-called hadronic punch-through).

Jets are clustered from PF candidates using the anti-k_{T} algorithm [35, 36] with a distance parameter of 0.4. Additional pp interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded [43] and an offset correction [44] is applied to correct for remaining contributions. In this study, jets are required to satisfy $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, and are subject to a set of loose identification criteria [45] to reject anomalous activity from instrumental sources, such as detector noise. These criteria ensure that each jet contains at least two PF candidates and at least one charged particle track, the energy fraction attributed to charged- and neutral-hadron PF candidates is nonzero, and the fraction of energy deposited in the ECAL attributed to charged and neutral PF candidates is less than unity.

The most accurate estimator of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative vector p_{T} sum of all the PF candidates in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [46]. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is modified to account for corrections to the energy scale [44] of the reconstructed jets in the event. Anomalous high-$p_{\mathrm{T}}^{\mathrm{miss}}$ events can be due to a variety of reconstruction failures, detector malfunctions, or noncollision backgrounds. Such events are rejected by event filters that are designed to identify more than 85%–90% of the spurious high-$p_{\mathrm{T}}^{\mathrm{miss}}$ events with a mistagging rate of less than 0.1% for genuine events [46].

Events of interest are selected using a two-tiered trigger system [47]. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors alone, whereas a version of the full event reconstruction software optimised for fast processing is performed at the second level, which runs on a farm of processors.

3. Event selection and sample composition

Split SUSY, as characterised by the simplified models discussed in this paper, would reveal itself in events containing jets, significant $p_{\mathrm{T}}^{\mathrm{miss}}$ from undetected neutralinos, and an absence of photons and leptons. Candidate signal events are required to satisfy a set of selection requirements that define a signal region (SR). Conversely, event samples that are enriched in the same background processes that populate the SR, while being depleted in contributions from SUSY processes, are identified as control regions (CRs).

In this analysis, CRs are used to assess the residual differences in tagger performance between data and simulation. The CRs are chosen to have an SM background composition similar to that of the SR. This permits a validation that the simulated event samples can accurately predict the SM background yields without significant systematic bias.

Candidate signal events in the SR are required to satisfy the following set of selection requirements. Events are required to contain at least three jets, as defined in section 2. Events are vetoed if they contain at least one electron (muon), isolated from other activity in the event, that satisfies $p_{\mathrm{T}} > 15(10)$ GeV, $|\eta| < 2.4$, and loose identification criteria [39, 48]. The mass scale of each event is estimated from the scalar p_{T} sum of the jets, $H_{\mathrm{T}} = \sum_i^{\mathrm{jets}} p_{\mathrm{T}}^i$, which is required to be larger than 300 GeV. An estimator for $p_{\mathrm{T}}^{\mathrm{miss}}$ is given by the magnitude of the vector $\vec{p}_{\mathrm{T}}$ sum of the jets, $H_{\mathrm{T}}^{\mathrm{miss}} = \left|\sum_i^{\mathrm{jets}} \vec{p}_{\mathrm{T}}^{\,i}\right|$, which is required to be larger than 300 GeV. Events in the SR can be efficiently recorded with a trigger condition that requires the presence of a single jet with $p_{\mathrm{T}} > 80$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV, and $p_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV.

Following these selections, the dominant contribution to the SM background comprises multijet events produced via the strong interaction, a manifestation of quantum chromodynamics (QCD). The multijet contribution is reduced to a negligible level using the following three criteria. Events containing at least one jet that satisfies $p_T > 50$ GeV and $2.4 < |\eta| < 5$ are vetoed to ensure an adequate resolution performance for the H_T^{miss} variable. Events are required to satisfy $H_T^{miss}/p_T^{miss} < 1.25$, which mitigates the rare circumstance in which several jets with p_T below the aforementioned 30 GeV threshold and collinear in ϕ lead to large values of H_T^{miss} relative to p_T^{miss}. The minimum azimuthal separation between each jet and the vector p_T sum of all other jets in the event, denoted $\Delta\phi^{\star}_{min}$ [49], is required to be greater than 0.2 radians.

The remaining background events in the SR are dominated by contributions from processes that involve the production of high-p_T neutrinos in the final state, such as the associated production of jets and a Z boson that decays to $\nu\bar{\nu}$. A further significant background contribution arises from events that contain a W boson that undergoes a leptonic decay, W($\rightarrow\ell\nu$)+jets, where the charged lepton ($\ell \equiv$ e, μ or τ) is outside the experimental acceptance, or is not identified, or is not isolated. Hence, a substantial contribution is also expected from the production of single top quarks and top quark-antiquark pairs ($t\bar{t}$), both of which can lead to a final state containing one leptonically decaying W boson and at least one b jet. Residual contributions from rare SM processes, such as diboson production or the associated production of $t\bar{t}$ and a vector or scalar boson, are not considered in this study.

Two CRs are used to assess differences in the performance of the tagger when using simulated events or pp collision data. The CRs are defined in terms of leptons and jets that satisfy $|\eta| < 2.4$ and the p_T requirements defined below. The single muon (μ+jets) CR is required to contain exactly one muon satisfying $p_T > 26$ GeV. The dimuon (μμ+jets) CR is required to contain a second muon that satisfies $p_T > 15$ GeV. The muons are required to be isolated from other activity in the event and satisfy identification criteria [48]. Events containing additional electrons (muons) with $p_T > 15(10)$ GeV and satisfying looser identification criteria, are vetoed. Both CRs must contain at least two jets satisfying $p_T > 30$ GeV. Events in the μ+jets and μμ+jets CRs are required, respectively, to satisfy $p_T^{miss} > 150$ GeV and $p_T(\mu\mu) > 100$ GeV, where the latter variable is the magnitude of the vector p_T sum of the two muons. The μ+jets CR comprises, in approximately equal measure, events from the associated production of jets and a W boson, and single top quark and $t\bar{t}$ production. The μμ+jets CR contains Drell–Yan ($q\bar{q} \rightarrow Z/\gamma^* \rightarrow \mu^{\pm}\mu^{\mp}$) events with subdominant contributions from $t\bar{t}$ and Wt-channel single top quark production. Events in both CRs are efficiently recorded with a trigger condition that requires the presence of a single isolated muon that satisfies $p_T > 24$ GeV.

4. Monte Carlo simulation

The DNN is trained to predict the jet class using supervised learning, which relies on generator-level information from MC programs. Various simulated event samples are also used during the evaluation of the DNN to benchmark the performance of the tagger.

Split SUSY predicts the unification of the gauge couplings at high energy [50–52] and a candidate dark matter particle, the neutralino. Apart from a low mass scalar Higgs boson, assumed in this model to be the state observed at 125 GeV [53, 54], only the fermionic SUSY particles may be kinematically accessible at the LHC. All other SUSY particles are assumed to be ultraheavy. The gluino is only able to decay through the highly virtual squark states. Hence, the gluino hadronises and forms a bound state with SM particles called an R hadron. The R hadron can travel a significant distance before the gluino undergoes a three-body decay to $q\bar{q}\chi_1^0$, depending on $c\tau_0$.

The split SUSY simplified models are defined by three parameters: $c\tau_0$ and the masses of the gluino $m_{\tilde{g}}$ and the neutralino $m_{\tilde{\chi}_1^0}$. The following model parameter space is considered by the search: $600 < m_{\tilde{g}} < 2400$ GeV, $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} > 100$ GeV, and $10\ \mu\text{m} < c\tau_0 < 10$ m. The lower and upper bounds of the $c\tau_0$ range are motivated by the $\mathcal{O}(10\ \mu\text{m})$ position resolution of the tracker subdetector [42] and the physical dimensions of the CMS detector, respectively. The tagger performance is also assessed using two split SUSY benchmark models that feature: an 'uncompressed' mass spectrum, with a large mass difference between the gluino and $\tilde{\chi}_1^0$, $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (2000, 0)$ GeV; or a 'compressed' spectrum that is nearly degenerate in mass, $(1600, 1400)$ GeV. The value of $c\tau_0$ for both models is defined in the text on a case-by-case basis.

Two benchmark models of GMSB and RPV SUSY are also considered to demonstrate the generalisation of the DNN. A GMSB-inspired model assumes a long-lived gluino, with a mass of 2500 GeV and $c\tau_0 = 1$ mm or 1 m, that decays to a gluon and a light gravitino of mass 1 keV. Again, all other SUSY particles are assumed to be ultraheavy and decoupled from the interaction. An RPV-inspired model assumes the production of top squark-antisquark pairs. The decay of the long-lived top squark, with a mass of 1200 GeV and $c\tau_0 = 1$ mm or 1 m, to a bottom quark and a charged lepton is suppressed through a small R parity violating coupling.

Samples of simulated events are produced with several MC generator programs. For the split SUSY simplified models, the associated production of gluino pairs and up to two additional partons are generated at leading-order (LO) precision in QCD with the MADGRAPH5_aMC©NLO 2.2.2 [55] program. The decay of the gluino is performed with the PYTHIA 8.205 [56] program. The RHADRONS package within the PYTHIA program, steered according to the default parameter settings, is used to describe the formation of R hadrons through the hadronisation of gluinos [57–59]. A similar treatment is performed for the GMSB- and RPV-inspired benchmark models.

The MADGRAPH5_aMC©NLO event generator is used to produce samples of $\mathrm{W}(\rightarrow\ell\nu)$+jets, $Z(\rightarrow\nu\bar{\nu})$+jets, and $\mathrm{Z}/\gamma^{*}(\rightarrow\mu\mu)$+jets events at next-to-leading-order (NLO) precision in QCD. The samples of $\mathrm{W}(\rightarrow\ell\nu)$+jets events are generated with up to two additional partons at the matrix element level and are merged with jets from the subsequent PYTHIA parton shower simulation using the FxFx scheme [60]. The POWHEG v2 [61–63] event generator is used to simulate $\mathrm{t\bar{t}}$ production [64] and the t-channel [65] and Wt-channel [66] production of single top quarks at NLO accuracy. Multijet events are simulated at LO accuracy using PYTHIA.

The theoretical production cross sections for SM processes are calculated with NLO and next-to-NLO precision [55, 65–69]. The production cross sections for gluino and top squark-antisquark pairs are calculated with NLO plus next-to-leading-logarithmic precision [70].

The PYTHIA program with the CUETP8M1 tune [71] is used to describe parton showering and hadronisation for all simulated samples except top quark-antiquark production, which used the CUETP8M2T4 tune [72]. The NNPDF3.0 LO and NLO parton distribution functions [73] are used with the LO and NLO event generators, respectively. Minimum bias events are overlaid with the hard scattering event to simulate pileup interactions, with the multiplicity distribution matched to that observed in pp collision data. The resulting events undergo a full detector simulation using the GEANT4 [74] package. The analysis described in section 8 focuses on identifying the decay products of the LLP, characterised by tracks with large impact parameters, calorimeter deposits, and secondary vertices. Interactions of R hadrons with the detector material that can result in signatures with short tracks or anomalous energy loss are therefore not considered in this study.

5. The LLP jet algorithm

In this section, the DNN architecture and its technical implementation are presented. The use of a single parameterised DNN for hypothesis testing and the application of domain adaptation (DA) to samples of pp collision data recorded by the CMS experiment are presented for the first time.

5.1. Jet labelling

Generator-level information from MC programs is often used to label a jet according to its initiating parton for supervised learning. A standard procedure known as 'ghost' labelling [43] determines the jet flavour by clustering the reconstructed final-state particles and the generator-level b and c hadrons into jets. Only the directional information of the four-momentum of the generator-level (ghost) hadron is retained to prevent any modification to the four-momentum of the corresponding reconstructed jet. Jets containing at least one ghost hadron are assigned the corresponding flavour label, with b hadrons preferentially selected over c hadrons. Similarly, labels are defined for jets originating from gluons (g) or light-flavour (uds) quarks.

The LLP jet tagger adopts the ghost labelling approach for jets originating from SM background processes. However, a complication arises when applying ghost labelling to the jets originating from $\tilde{\mathrm{g}}$ decays. The quark and antiquark produced in the gluino decay can interact with each other, potentially leading to one or more jets that do not point in the same direction as the quarks. Two examples of $\tilde{\mathrm{g}}\rightarrow\mathrm{q\bar{q}}\tilde{\chi}_1^0$ decays are shown in figure 1. For each example, the final-state particles resulting from the hadronisation of one of the quarks are sufficiently diffuse that they are clustered into multiple distinct jets. By definition, ghost tagging cannot account for multiple jets originating from a single ghost particle, and it may even fail to associate the ghost particle with any of the jets if the jets are sufficiently distanced in η–ϕ space.

An alternative labelling scheme is defined for jets originating from gluino decays, which can be extended to other LLP decays. All stable SM particles are grouped according to their simulated vertex position and linked to one of the quark daughters from the LLP decay. All stable SM particles, except neutrinos, are clustered into generator-level jets using the anti-k_{T} algorithm [35, 36] with a distance parameter of 0.4. Given that constituent particles in a jet may originate from different vertices, the momentum $\vec{p}$ of a given jet is shared between vertices according to the vectorial momentum sum $\sum_i\vec{p}_i$ of the constituent particles i in a jet that share the same generator-level vertex v. Per jet, the jet-vertex shared momentum fraction f_v with respect to vertex v is then defined as

$$f_v=\frac{\left(\sum_i\vec{p}_i\middle|\,i\in v\right)\cdot\vec{p}}{\vec{p}^2},\quad f_v\in[0;1],\quad\sum_{v}^{\text{vertices}}f_v=1. \tag{1}$$

Each jet is then associated to the vertex v from which the majority of its momentum originates, i.e. $\hat{v}=\mathrm{argmax}(f_v)$. This criterion prevents the coincidental association of jets containing very few or very soft

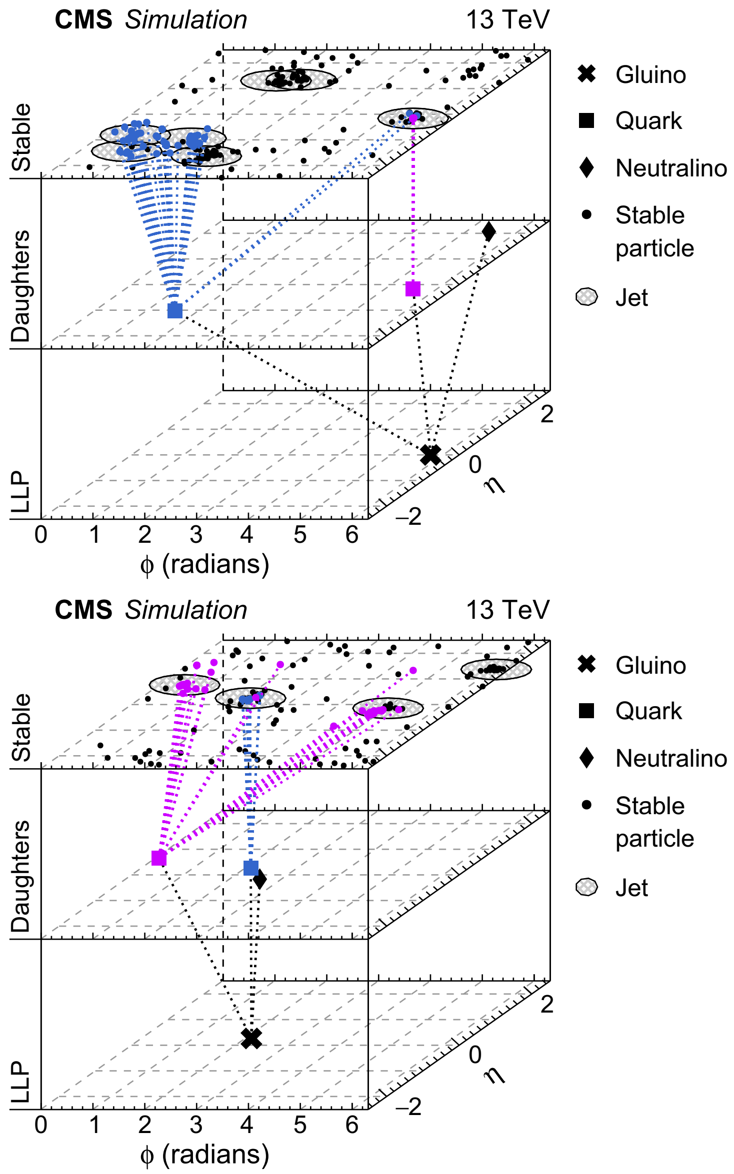

Figure 1. Two example $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ decay chains, constructed from information provided by the MADGRAPH5_aMC©NLO [55] and PYTHIA [56] programs. The positions of various particles in the η–ϕ plane are shown: the LLP ($\tilde{g}$) and its daughter particles ($q\bar{q}\tilde{\chi}_1^0$) are shown in the lower and middle planes, respectively; the upper plane depicts the location of the stable particles after hadronisation, with shaded ellipses overlaid to indicate the reconstructed jets. Each quark and its decay is assigned a unique colour. The dotted lines indicate the links between parent and daughter particles.

particles from a gluino decay to a vertex for which the majority of the constituent particles stem from initial-state radiation or the underlying event. A reconstructed jet is uniquely associated with a generator-level jet and they adopt the same label if their axes are aligned within a cone $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$. The LLP label is prioritised over the other jet labels to prevent ambiguities. Jets from the split SUSY samples that are not labelled as LLP by this scheme can still comprise a non-negligible fraction of displaced particles and are thus discarded to prevent class contamination. Non-LLP jets for the DNN training are instead taken from simulated samples of SM backgrounds. Applying an artificial 20% contamination of the LLP jet class from pileup jets, to test the robustness of the labelling procedure, leads to no discernible effect on the tagger performance given the statistical precision of the study.

5.2. Deep neural network architecture to predict the jet class

The architecture used to predict the jet class, inspired by the DeepJet algorithm, is shown in figure 2. The DNN considers approximately 600 input features, which can be grouped into four categories: up to 25 charged (neutral) PF candidates, each described by 17 (6) features and ordered by impact parameter significance (transverse momentum); up to four secondary vertices (ordered by transverse displacement significance), each described by 12 features; and 14 global features associated with the jet. The network also

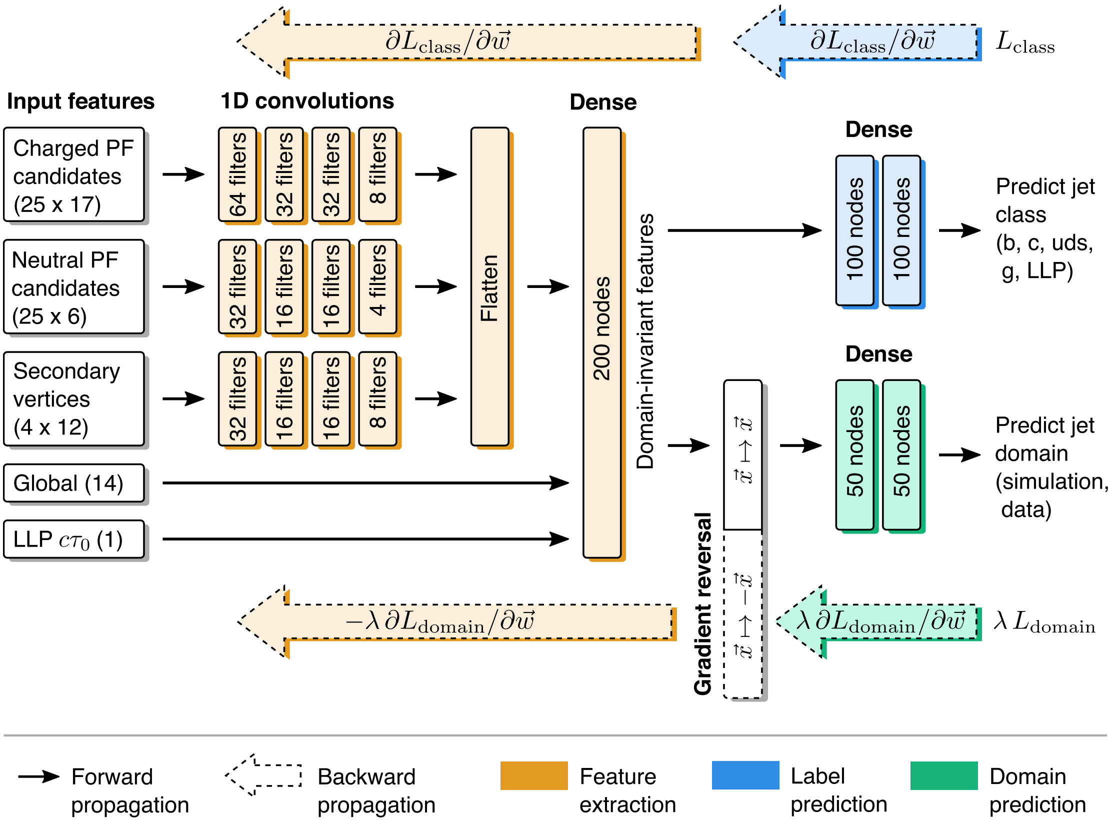

Figure 2. An overview of the DNN architecture, which comprises convolutional and dense layers; the numbers of filters and nodes are indicated. Dropout layers and activation functions are not shown. The input features are grouped by object type and ($m \times n$) indicates the maximum number of objects (m) and the number of features per object (n). The gradients of the class (L_{class}) and domain (L_{domain}) losses with respect to the weights $\vec{w}$, used during backward propagation, are shown.

considers jets containing zero reconstructed secondary vertices. Zero padding is used to accommodate the variable numbers of PF candidates and secondary vertices.

Each charged PF candidate is described by the following features: the p_T relative and perpendicular to the jet axis, the $\Delta\eta$ with respect to the jet axis, the track quality, and the transverse and three-dimensional impact parameters (and their significances) of the track. Each neutral PF candidate is described by its energy, the fractions of its energy deposited within the ECAL and HCAL subdetectors, the compatibility with the photon hypothesis, the compatibility with the pileup hypothesis as determined by the PUPPI algorithm [75, 76]. Charged and neutral PF candidates are also described by the collinearity with respect to the jet axis and the nearest secondary vertex. The features that describe each reconstructed secondary vertex include the three-dimensional displacement (and significance) with respect to the primary pp collision vertex, the number of associated tracks, and the following quantities determined from the four-momenta of the associated tracks: p_T, the $\Delta\eta$ with respect to the jet axis, and the invariant mass. The global jet features comprise the jet momentum and pseudorapidity, the number of constituent PF candidates, the number of reconstructed secondary vertices, and several high-level engineered features used by the CSV b tagging algorithm [5].

Four sequential layers of one-dimensional convolutions with a kernel size of one are used, with each layer comprising 64, 32, 16, 8, or 4 filters depending on the group of input features. Per particle candidate or vertex, each convolutional layer transforms the features from the preceding layer according to its filter size. By choosing a small filter size for the final layer, the overall operation can be viewed as a compression. After each layer, a leaky rectified linear (LeakyReLU) [77] activation function is used. Dropout [78] layers are interleaved throughout the network with a 10% dropout rate to mitigate overfitting. After the final convolutional layer, the compressed feature vectors are flattened and concatenated along with the global jet features. The input parameter $c\tau_0$ shown in figure 2 is described in section 5.3.

The resulting feature vector is fed into a multilayer perceptron, a series of dense layers comprising 200, 100, or 50 nodes. The softmax activation function is used in the last layer. Categorical cross entropy is used for the loss function to predict the jet class probability. The DNN provides an estimate of the probabilities for the following jet classes: LLP jet, b or c jet, uds (light-flavour quarks) jet, and g jet. The latter two classes are frequently combined when evaluating the network performance to give a light-flavour (udsg) jet class.

5.3. Network parameterisation according to $c\tau_0$ the proper decay length

The experimental signature for a jet produced in a gluino decay depends strongly on $c\tau_0$. Information from all CMS detector systems is available if the gluino decay occurs promptly, in the vicinity of the pp collision region, while information can be limited if the decay occurs in the outermost detector systems. Hence, $c\tau_0$ is introduced as an input parameter to the dense network, as indicated in figure 2. This parameterised approach [32] allows for hypothesis testing with a single network for values of $c\tau_0$ that span six orders of magnitude: $10\ \mu\text{m} < c\tau_0 < 10$ m. During the network evaluation for jets from both signal and SM processes, the $c\tau_0$ value is given by the signal hypothesis under test.

5.4. Domain adaptation by backward propagation

The simulated event samples are of limited accuracy and do not exhaustively reproduce all features observed in the pp collision data. Hence, a neural network may produce different identification efficiencies when evaluated on simulated samples and pp collision data if the training of the neural network relies solely on simulation. Domain adaptation is a technique that attempts to construct a feature representation that is invariant with respect to the domain from which the features are obtained. In this study, the domain is either pp collision data or simulation, and the domain-invariant representation is obtained from the output of the feature extraction subnetwork, as indicated in figure 2. Hence, for the subsequent task of classifying jets, a similar performance is expected for both simulated events and pp collision data [79].

In this paper, we use DA by backpropagation of errors [34]. To achieve this, an additional branch is added after the first dense layer of 200 nodes. Binary cross entropy is used for the loss function to predict the jet domain probability. During training, the combined loss $L_{\text{class}} + \lambda L_{\text{domain}}$ is minimised, where λ is a hyperparameter that controls the magnitude of the jet domain loss. A gradient reversal layer is inserted in the domain branch directly after the first dense layer. This special layer is only active during backward propagation and reverses the gradients of the domain loss L_{domain} with respect to the network weights $\vec{w}$ in the preceding layers. This forces the feature extraction subnetwork to focus less on domain-dependent features because its weights effectively minimise $L_{\text{class}} - \lambda L_{\text{domain}}$ instead. Only features that are common to both pp collision data and simulation are retained. For a full mathematical description of the gradient reversal layer, see reference [34].

5.5. Deep neural network training

Supervised training of the DNN is performed to predict the jet class and domain. The Adam optimiser [80] is used to minimise the loss function with respect to the network parameters. The DNN training relies on simulated events of gluino production from split SUSY models, and multijet and $\mathrm{t\bar{t}}$ production. Jets in these samples that have an uncorrected p_{T} greater than 20 GeV and satisfy the loose requirements outlined in section 2 are considered for the DNN training. Approximately 20 million jets are used. Event samples of split SUSY models are utilised over a wide range of $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ hypotheses to ensure adequate generalisation of the DNN over the model parameter space of interest. The models considered by the DNN training have seven $c\tau_0$ values that differ by factors of ten and span the range $10\ \mu\text{m} < c\tau_0 < 10$ m. For the jet domain prediction, 1.2 million jets are drawn from the pp collision data, as well as from simulated $\mathrm{W}(\rightarrow \ell\nu)$+jets and $\mathrm{t\bar{t}}$ events that are weighted according to their respective SM cross sections; all events are required to satisfy the μ+jets CR requirements.

The DNN training comprises a few tens of cycles over the full set of event samples (epochs). Each epoch is batched into subsamples of 10 000 jets containing approximately 2000 jets from each class: LLP, b, c, uds, and g. For each batch, the LLP jets are sampled randomly from split SUSY events that are generated according to the full $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}, c\tau_0)$ parameter space of interest. The jets from SM processes are drawn randomly (from a larger sample) to obtain the same binned p_{T} and η distributions available for the LLP jet class. This resampling technique is done to ensure an adequate generalisation that is largely independent of kinematical features related to the physics process. The $c\tau_0$ values assigned to the SM jets are generated randomly according to the $c\tau_0$ distribution obtained per batch for the LLP jet class.

The training of the domain branch uses batches of 10 000 jets, drawn from samples of μ+jets data and simulated $\mathrm{W}(\rightarrow \ell\nu)$+jets and $\mathrm{t\bar{t}}$ events. The DNN is trained using the class and domain batches simultaneously. Events in the domain batch are assigned the same $c\tau_0$ values as used by the class batch. For the domain batch, only the six highest p_{T} jets are used, and jets may be reused multiple times per epoch.

The DNN is initially trained to predict only the jet class to determine the optimal scheduling of the learning rate α, which decays from an initial value of 0.01 according to $\alpha = 0.01/(1+\kappa n)$ where n is the epoch number and κ is the decay constant. The classifier performance is optimal for $\kappa = 0.1$ and only weakly

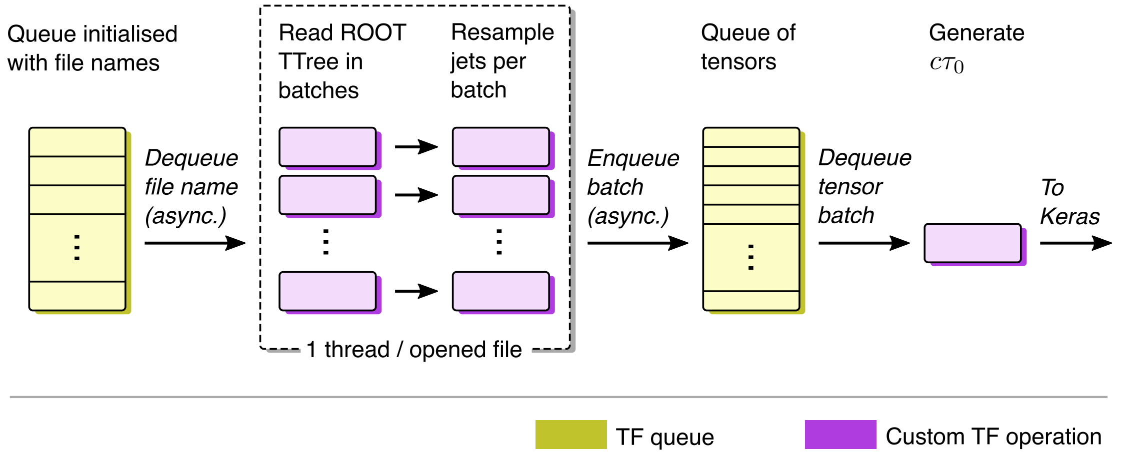

Figure 3. A schematic of the input pipeline for training the DNN, which uses the TensorFlow (TF) queue system with custom operation kernels for reading ROOT trees from disk, (p_T, η) resampling for SM jets, and generating random $c\tau_0$ values for jets from SM backgrounds and data.

dependent on κ. The DNN is then trained to predict both the jet class and domain, and the λ hyperparameter is increased according to $\lambda = \lambda_0[2/(1+\mathrm{e}^{-0.2n}) - 1]$ with $\lambda_0 = 30$, such that λ increases from 0 to $0.9\lambda_0$ after 15 epochs. The parameterisations used to evolve the values of the α and λ hyperparameters during the DNN training were chosen from several trials to ensure reproducible and optimal performance.

5.6. Workflow

The Keras v2.1.5 [81] software package is used to implement the DNN architecture. The TensorFlow v1.6 [82] queue system is used to read and preprocess files for the DNN training. A schematic overview of the pipeline used to preprocess batches of jets for the class prediction is given in figure 3, while a similar queue is also used for the domain prediction. At the beginning of each epoch, a queue holding a randomised list of the input file names is initialised. File names are dequeued asynchronously in multiple threads. For each thread, ROOT v6.18.00 [83] trees contained in the files are read from disk to memory in batches using a TensorFlow operation kernel, developed in the context of this paper. The resulting batches are resampled to achieve the same distributions in p_T and η for all jet classes and are enqueued asynchronously into a second queue, which caches a list of tensors. The DNN training commences by dequeuing a randomised batch of tensors and generating $c\tau_0$ values for all SM jets within the batch. The advantages of this system lie in its flexibility to adapt to new input features or samples on-the-fly. The reading of ROOT trees and the (p_T, η) resampling for the SM jets proceeds asynchronously in multiple threads, managed by TensorFlow, on the CPU while the network is being trained. A demonstration of the workflow can be found at reference [84].

6. Validation with pp collision data

In the absence of DA, the LLP jet probability $P(LLP|ctau)$ obtained from the simulation of the relevant SM backgrounds and the CR data can differ significantly, with deviations of up to 50% in the binned counts of $P(LLP|ctau)$ distributions. A similar event-level variable is $P_{\max}(\mathrm{LLP}|c\tau_0)$, which is defined as the maximum value of $P(LLP|ctau)$ obtained from all selected jets in a given event. A comparison of the $P_{\max}(\mathrm{LLP}|c\tau_0)$ distributions obtained from pp collision data and simulated events in both the μ+jets and μμ+jets CRs, using a DNN trained with and without DA, is shown in figure 4. The use of DA in the network leads to a significant improvement in the level of agreement in the binned counts of $P_{\max}(\mathrm{LLP}|c\tau_0)$ for the two domains of data and simulation, with only small residual differences remaining. This improvement is expected for the μ+jets CR, as the same events are used to train, evaluate, and optimise the domain branch of the DNN. The μμ+jets CR, comprising a statistically independent event sample, validates the method.

An estimate of the uncertainty in $P_{\max}(\mathrm{LLP}|c\tau_0)$ due to simulation mismodelling is determined from jets in a statistically independent sample of μμ+jets events that satisfy $p_T(\mu\mu) < 100$ GeV. The magnitude of the uncertainty is assessed by weighting up or down the simulated events by the factor $w^{\pm} = \prod_i^{\text{jets}}\Big(1 \pm \big(\xi_i(\mathrm{LLP}) - 1\big)\Big)$, where $\xi_i(\mathrm{LLP})$ is the ratio of counts from data and simulation in bin i of the $P(LLP|ctau)$ distribution. The ratios of event counts binned according to $P_{\max}(\mathrm{LLP}|c\tau_0)$ from pp collision data and simulation, as well as the corresponding uncertainty estimates, are shown in the lower panels of figure 4. The ratios are closer to unity, with reduced uncertainties, following the application of DA. The level

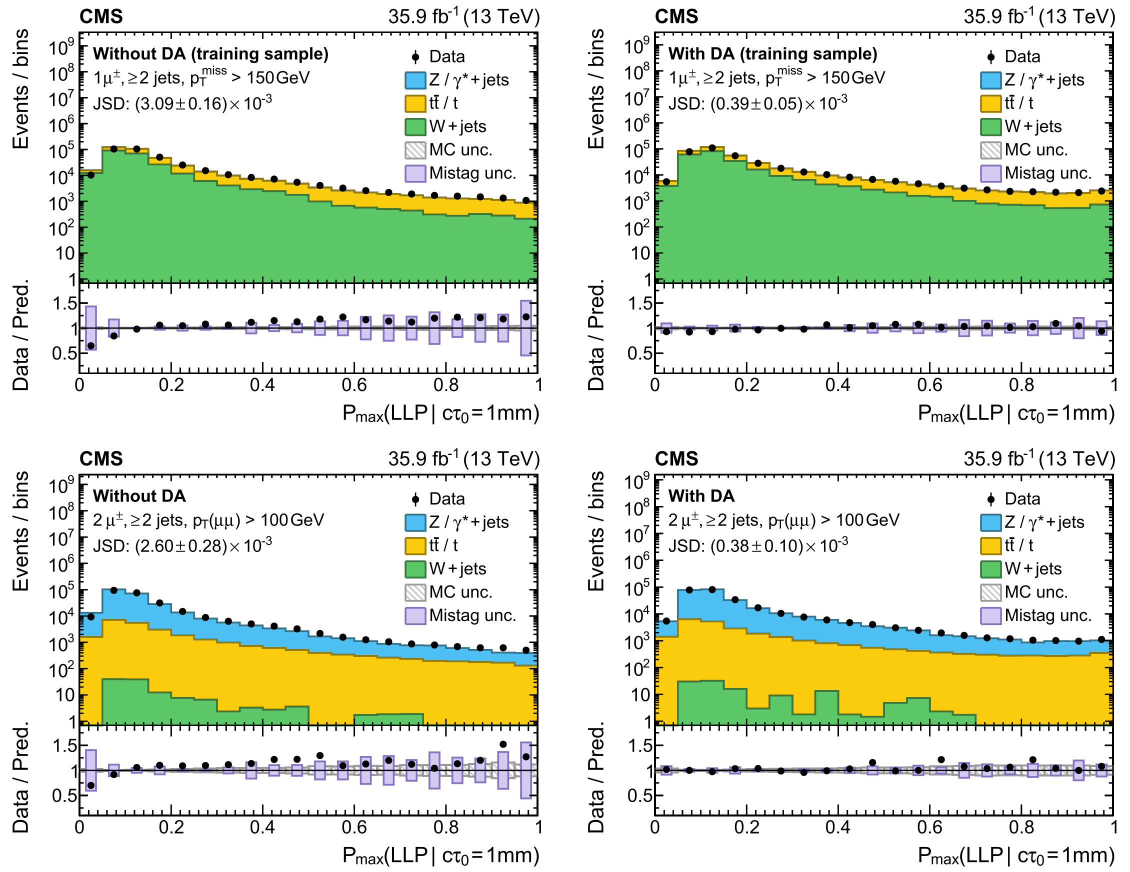

Figure 4. Distributions of the maximum probability for the LLP jet class obtained from all selected jets in a given event, $P_{\max}(\mathrm{LLP}|c\tau_0)$. The distributions from pp collision data (circular marker) and simulated events (histograms) are compared in the μ+jets (upper row) and μμ+jets (lower row) CRs, using a DNN trained without (left column) and with (right column) DA. All probabilities are evaluated with $c\tau_0 = 1$ mm. The Jensen–Shannon divergence (JSD) is introduced in the text. The lower subpanels show the ratios of the binned yields obtained from data and Monte Carlo (MC) simulation. The statistical (hatched bands) and systematic (solid bands) uncertainties due to the finite-size simulation samples and the simulation mismodelling of the mistag rate, respectively, are also shown.

of agreement between data and simulation is further quantified by the Jensen–Shannon divergence (JSD) [85], a measure of similarity between two probability distributions that is bound to [0, 1] and takes a value of zero for identical distributions. The JSD is reduced by an order of magnitude following the application of DA. The quoted uncertainties in JSD reflect the finite sizes of the data and simulated samples.

The application of DA leads to significantly reduced biases and uncertainties in the modelling of $P(LLP|ctau)$ and related variables in the signal-depleted CRs. This behaviour would translate into an improved treatment for the estimation of SM backgrounds in the SR and the associated experimental systematic uncertainties. However, only modest gains in sensitivity to new high-mass particle states may be expected, as the dominant uncertainties arise from the finite-size samples of pp collision data and simulated events.

7. Performance

The performance of the tagger is demonstrated using simulated event samples for split SUSY benchmark models with an uncompressed mass spectrum, as defined in section 4, and $c\tau_0$ values of 1 mm and 1 m. The two values of $c\tau_0$ give greater weight to the roles of the tracker and calorimeter systems, respectively. Negligible SM background contributions are expected for the 1 m scenario. An inclusive sample of $\mathrm{t\bar{t}}$ events is used to provide both light-flavour (udsg) jets, through initial-state radiation and hadronic decays of the W boson, and b jets, with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$.

The efficiency of the tagger to identify correctly the LLP jet class depends on the chosen working point, defined by a threshold requirement on the jet class probability. The mistag rates for the remaining jet classes also depend on the same working point. The receiver operating characteristic (ROC) curves that provide the LLP jet tagging efficiency and the mistag rate for the udsg jet class as a function of the working point are shown in figure 5. The uncertainties indicated by the shaded bands are determined from the standard

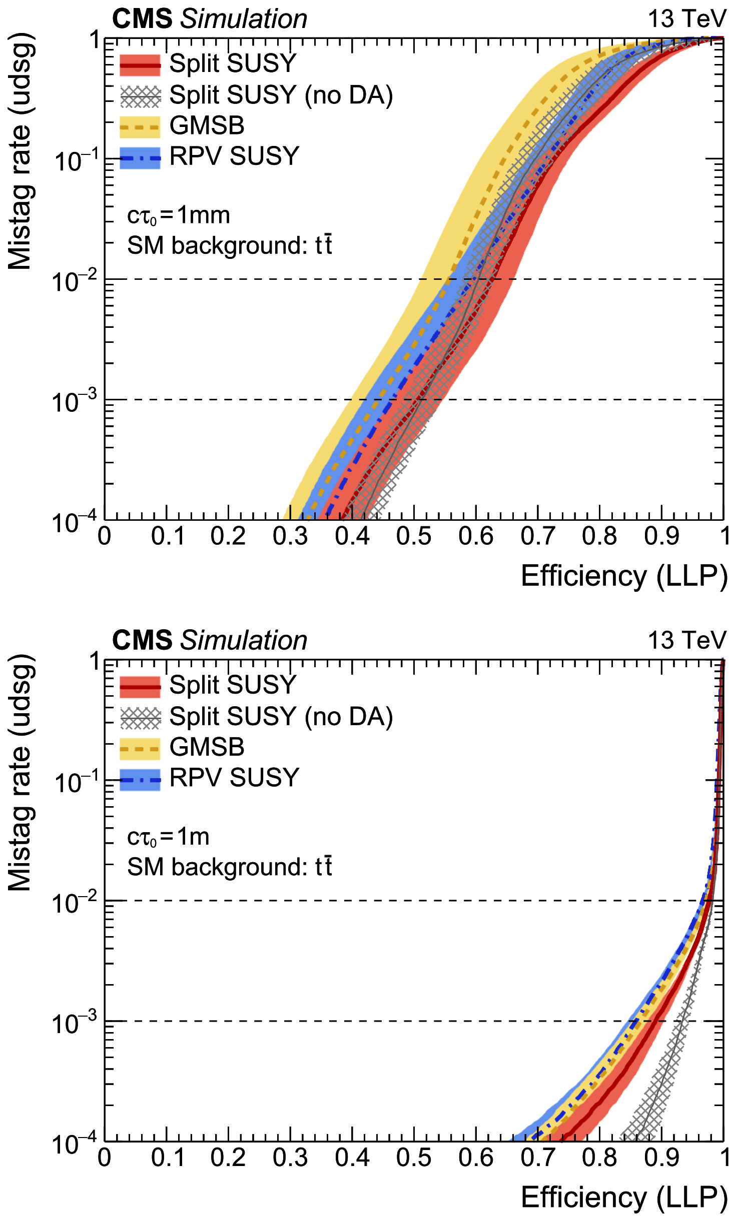

Figure 5. The ROC curves illustrating the tagger performance for the split (solid line), GMSB (dashed line), and RPV (dot-dashed line) SUSY benchmark models, defined in section 4, assuming $c\tau_0$ values of 1 mm (upper) and 1 m (lower). The thin line with hatched shading indicates the performance obtained with a DNN training using split SUSY samples but without the DA. The jet sample is defined in the text.

deviation obtained from a ten-fold cross validation. Given a mistag rate of 0.01%, equivalent to a background rejection factor of 10 000, efficiencies of 40 and 70% are obtained for split SUSY models with $c\tau_0$ values of 1 mm and 1 m, respectively. These efficiencies decrease by a factor ≈0.6 if the DNN relies solely on the global jet features. Additional studies reveal that the tagger efficiency for the LLP jet class falls to 25% in order to achieve the same mistag rate of 0.01% for the (SM) b jet class when considering split SUSY models with $c\tau_0 = 1$ mm.

Figure 5 also shows ROC curves when evaluating the DNN using simulated events from the GMSB and RPV SUSY benchmark models, as defined in section 4. The jets originate from uds quarks (gluons) from the gluino decay in the case of split (GMSB) SUSY models, and b hadrons from the top squark decay for RPV SUSY. A similar level of performance is observed for these SUSY models, in which the LLP jets have a different underlying flavour. Furthermore, the ROC curves indicated by the thick and thin solid curves illustrate the tagger performance when the DNN is trained with or without DA, respectively, for the split SUSY benchmark models. Studies demonstrate a comparable performance for split SUSY models with and without DA for the region $c\tau_0 \leq$ 10 mm. For larger values of $c\tau_0$, the performance is overestimated in the absence of DA, as indicated by the $c\tau_0 = 1$ m scenario shown in figure 5. This is because the DNN is able to exploit patterns in the features obtained from simulation that are not representative of those obtained from data.

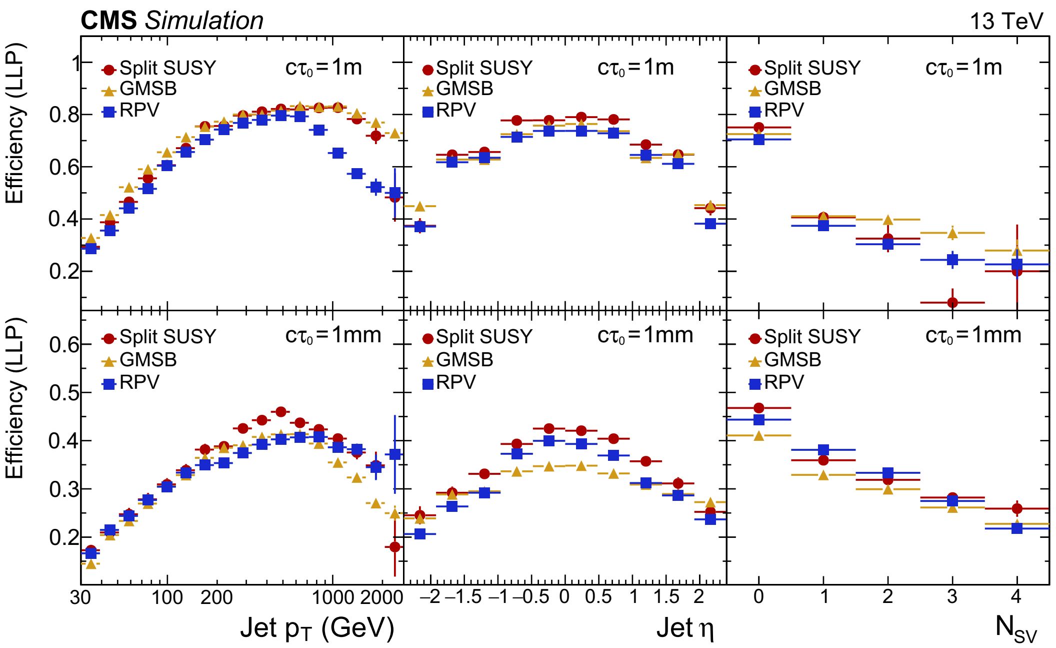

Figure 6. The LLP jet tagging efficiency as a function of the jet p_T, η, and N_{SV} using a working point that yields a mistag rate of 0.01% for the udsg jet class, as obtained from an inclusive sample of simulated $t\bar{t}$ events. The efficiency curves are shown separately for the split (circular marker), GMSB (triangle marker), and RPV (square marker) SUSY benchmark models, defined in section 4, assuming $c\tau_0$ values of 1 m (upper row) and 1 mm (lower row).

The LLP jet tagging efficiency is shown in figure 6 as a function of the jet p_T, η, and the number of reconstructed secondary vertices within the jet (N_{SV}) for a working point that yields a mistag rate of 0.01% for the udsg jet class obtained from simulated $t\bar{t}$ events. Efficiencies are highest for high-p_T, centrally produced jets with $N_{SV} = 0$. The latter observation demonstrates the complementary performance of the tagger with respect to a more standard approach of relying on reconstructed secondary vertices to identify displaced jets.

The performance of the DNN parameterisation according to $c\tau_0$ is shown in figure 7. The upper panel shows the LLP jet tagging efficiency, using a working point that yields a mistag rate of 0.01% for the udsg jet class, as a function of the generated $c\tau_0$ value, for both an uncompressed and a compressed split SUSY model. Efficiencies in the range 40–80 (30–70)% are achieved for uncompressed (compressed) scenarios with 1 mm$\leq c\tau_0 \leq$ 10 m. The compressed scenarios are characterised by low-$\mathrm{p_T}$ jets, resulting in lower efficiencies as shown in figure 6.

The performance is further tested using two uncompressed split SUSY models with $c\tau_0 = 1$ mm and 1 m. The LLP jet tagging efficiency is obtained by evaluating the DNN for each value of $c\tau_0$ in the range 10 μm$\leq c\tau_0 \leq$ 10 m. Again, the efficiency is determined using a working point that is tuned for each evaluated $c\tau_0$ value to yield a mistag rate of 0.01% for the udsg jet class. The efficiency as a function of the evaluated $c\tau_0$ value is shown in figure 7 (lower). The maximum efficiency is obtained when the evaluated value of $c\tau_0$ approximately matches the parameter value of the split SUSY model. This behaviour may be used to characterise a potential signal contribution in terms of the model parameter $c\tau_0$. Finally, studies demonstrate that the parameterised approach does not significantly impact performance with respect to the training of multiple DNNs, one per $c\tau_0$ value.

8. Application to a search for LLPs

The performance of the LLP jet tagger is demonstrated by applying it to the search for long-lived gluinos with 10 μm$\leq c\tau_0 \leq$ 10 m. Expected limits on the theoretical production cross section for gluino pairs are determined. The search is performed using statistically independent samples of simulated events to ensure an unbiased evaluation of the tagger.

8.1. Categorisation of events and uncertainties

Candidate events that satisfy the SR requirements defined in section 3 are categorised according to: the number of jets in the event, N_{jet}; the number of jets for which $P(LLP|ctau)$ is above a predefined threshold, $P(LLP|ctau)$; and H_T. The resulting six categories are summarised in table 1. Models with an uncompressed

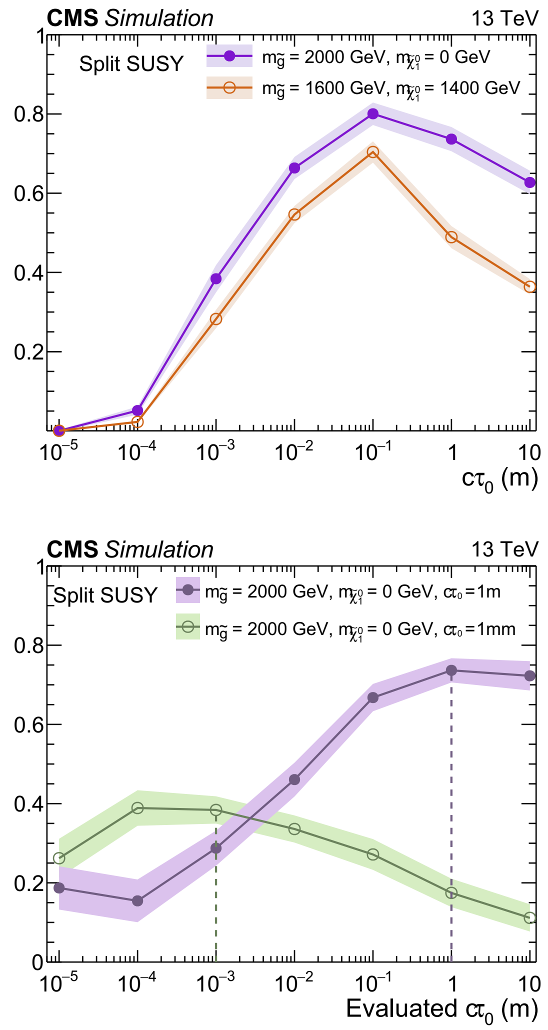

Figure 7. The LLP jet tagging efficiency, using a working point that yields a mistag rate of 0.01% for the udsg jet class obtained from an inclusive sample of simulated $\mathrm{t\bar{t}}$ events, when (upper) the DNN is evaluated as a function of the model parameter value $c\tau_0$ for an uncompressed and a compressed split SUSY model, and (lower) the DNN is evaluated over a range of $c\tau_0$ values for uncompressed split SUSY models generated with $c\tau_0 = 1$ mm and 1 m; the dashed vertical lines indicate equality for the evaluated and generated values of $c\tau_0$ for each model. The fixed model parameters are defined in the legends.

Table 1. The event counts and uncertainties for SM backgrounds and split SUSY models, as determined from simulation, in categories defined by H_T and $(N_\mathrm{jet}, P(LLP|ctau))$. The simulated samples are normalised to an integrated luminosity of 35.9 fb^{-1}. The uncompressed and compressed split SUSY models are defined in section 4. The value of $c\tau_0$ is assumed to be 1 mm. The uncertainties include both statistical and systematic contributions. Expected counts for events that satisfy $P(LLP|ctau) < 2$ are not shown.

H_T (GeV)	300–800	300–800	300–800	> 800	> 800	> 800
$(N_\mathrm{jet}, P(LLP\|ctau))$	(3–4, ≥ 2)	(5, ≥ 2)	(≥ 6, ≥ 3)	(3–4, ≥ 2)	(5, ≥ 2)	(≥ 6, ≥ 3)
$\mathrm{Z}(\to\nu\bar{\nu})$+jets	41 ± 39	6.5 ± 5.8	0.6 ± 0.4	3.3 ± 2.8	1.6 ± 1.2	0.1 ± 0.1
$\mathrm{W}(\to\ell\bar{\nu})$+jets	56 ± 44	11.6 ± 5.1	1.5 ± 0.5	3.6 ± 2.5	1.2 ± 3.0	<0.1
$\mathrm{t\bar{t}}$	40 ± 36	18 ± 16	1.9 ± 1.1	2.1 ± 1.3	3.2 ± 2.4	3.0 ± 2.1
Single top quark	5.7 ± 5.2	2.6 ± 2.2	0.3 ± 0.2	0.6 ± 0.4	0.5 ± 0.3	0.4 ± 0.3
Total SM	142 ± 69	39 ± 18	4.3 ± 1.3	9.7 ± 4.0	6.5 ± 4.1	3.5 ± 2.5
Uncompressed	<0.1	<0.1	<0.1	3.0 ± 2.9	3.8 ± 3.7	5.7 ± 5.5
Compressed	5.4 ± 5.0	4.2 ± 3.8	2.8 ± 2.5	1.1 ± 0.9	2.5 ± 2.2	4.5 ± 4.1

mass spectrum, $m_{\tilde{g}} - m_{\tilde{\chi}^0_1} \gtrsim 200$ GeV, are characterised by high values of H_T. Models with a compressed mass spectrum, $m_{\tilde{g}} - m_{\tilde{\chi}^0_1} \lesssim 200$ GeV, are characterised by lower values of H_T because of the limited kinematic phase space available to the jets from the gluino decay and an increased reliance on associated jet production from initial-state radiation. Events that satisfy $P(LLP|ctau) < 2$ are grouped into a single additional category, which is used to constrain the normalisation of simulated background events during the statistical evaluation.

The tagger is evaluated using simulated samples of all relevant SM backgrounds, described in section 4. The negligible background contribution from multijet events in the SR is not considered in this exploratory study. The predefined threshold on $P(LLP|ctau)$ is determined per $c\tau_0$ value such that the most sensitive event categories are nearly free of SM background contributions, while control over uncertainties due to the finite-size simulated samples is maintained. The $P(LLP|ctau)$ thresholds fall in the range 30%–50% and yield an LLP jet tagging efficiency of about 30%–90% for $c\tau_0 \geq 1$ mm. Table 1 summarises the expected counts and uncertainties for the contributions from SM backgrounds, in the various event categories, for $c\tau_0 = 1$ mm. The statistical uncertainty arising from the finite size of the simulated samples is the dominant contribution to the quoted uncertainties. Additional systematic contributions are described below. The expected event counts from the uncompressed and compressed benchmark models of split SUSY, defined in section 4, are also provided.

Several sources of systematic uncertainty in the SM background expectations are considered. An uncertainty of ±20% is assumed in the normalisation of each dominant background process, $\mathrm{W}(\rightarrow\ell\nu)$+jets, $Z(\rightarrow\nu\bar{\nu})$+jets, and single top quark and $\mathrm{t\bar{t}}$ production, which is motivated by theoretical uncertainties in the production cross sections and uncertainties in the experimental acceptance for the final-state leptons [86]. The uncertainty in the mistag rate for jets from SM processes is typically below 10%, determined by the procedure described in section 6. The jet energy is varied within its uncertainty and resolution [44], and the resulting shifts are propagated to $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The unclustered component of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is varied within its uncertainties. The uncertainty in the number of pileup interactions is determined by varying the inelastic pp cross section within its uncertainty of ±5% [87]. The renormalisation and factorisation scales of the aforementioned four dominant SM backgrounds are varied independently per process by factors of 0.5 and 2 [88] to estimate the migration of events between categories. An uncertainty of ±2.5% in the integrated luminosity is assumed [89].

The LLP jet tagging efficiency determined from simulated events of split SUSY models, ϵ_{MC}, may require the application of corrections to account for sources of potential mismodelling. The corrections are applied by reweighting simulated events as follows:

$$\begin{aligned} w &= \left(\frac{1-\epsilon_{\mathrm{data}}}{1-\epsilon_{\mathrm{MC}}}\right)^{(N_{\mathrm{LLP}}-N_{\mathrm{LLP}}^{\mathrm{tag}})} \left(\frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}}\right)^{N_{\mathrm{LLP}}^{\mathrm{tag}}} \\ &= \left(\frac{1-\mathrm{SF}\,\epsilon_{\mathrm{MC}}}{1-epsilon_{\mathrm{MC}}}\right)^{(N_{\mathrm{LLP}}-N_{\mathrm{LLP}}^{\mathrm{tag}})} \mathrm{SF}^{N_{\mathrm{LLP}}^{\mathrm{tag}}}, \end{aligned} \tag{2}$$

$$w = \left(\frac{1-\epsilon_{\mathrm{data}}}{1-\epsilon_{\mathrm{MC}}}\right)^{(N_{\mathrm{LLP}}-N_{\mathrm{LLP}}^{\mathrm{tag}})} \left(\frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}}\right)^{N_{\mathrm{LLP}}^{\mathrm{tag}}} = \left(\frac{1-\mathrm{SF}\,\epsilon_{\mathrm{MC}}}{1-\epsilon_{\mathrm{MC}}}\right)^{(N_{\mathrm{LLP}}-N_{\mathrm{LLP}}^{\mathrm{tag}})} \mathrm{SF}^{N_{\mathrm{LLP}}^{\mathrm{tag}}},$$

where N_{LLP} denotes the number of LLP jets and $N_{\mathrm{LLP}}^{\mathrm{tag}}$ the number of LLP jets that are also tagged per event. The scale factor $\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$ denotes a correction to ϵ_{MC}, and the product of ϵ_{MC} and the SF is bound to $[0, 1]$.

8.2. Signal hypothesis testing

A likelihood model is used to test for the presence of new-physics signals in the SR. The observed event count in each event category is modelled as a Poisson-distributed variable around the sum of the SM expectation and a potential signal contribution. The expected event counts from SM backgrounds are obtained from the simulated samples. The uncertainties resulting from the finite simulated samples are modelled using the Barlow–Beeston method [90]. The systematic uncertainties in the SM background estimates are accommodated in the likelihood model as nuisance parameters.

Hypothesis testing is performed using an Asimov data set [91] to provide expected constraints on simplified models of split SUSY. A modified frequentist approach is used to determine the expected upper limits at 95% confidence level (CL) on the theoretical gluino pair production cross section as a function of $m_{\tilde{\mathrm{g}}}$, $m_{\tilde{\chi}_1^0}$, and $c\tau_0$. The signal strength parameter, r_{UL}, expresses the upper limit on the production cross section relative to the theoretical value. Alternatively, expected lower limits at 95% CL on $m_{\tilde{\mathrm{g}}}$ can be determined as a function of $c\tau_0$. The approach is based on the profile likelihood ratio as the test statistic [92], the $\mathrm{CL_s}$ criterion [93, 94], and the asymptotic formulae [91] to approximate the distributions of the test statistic under the background-only and signal-plus-background hypotheses.

8.3. Interpretations

Figure 8 shows the negative log-likelihood from a maximum likelihood fit to the Asimov data as a function of both the SF and r/r_{UL}, where the signal strength parameter r represents the injected gluino pair production

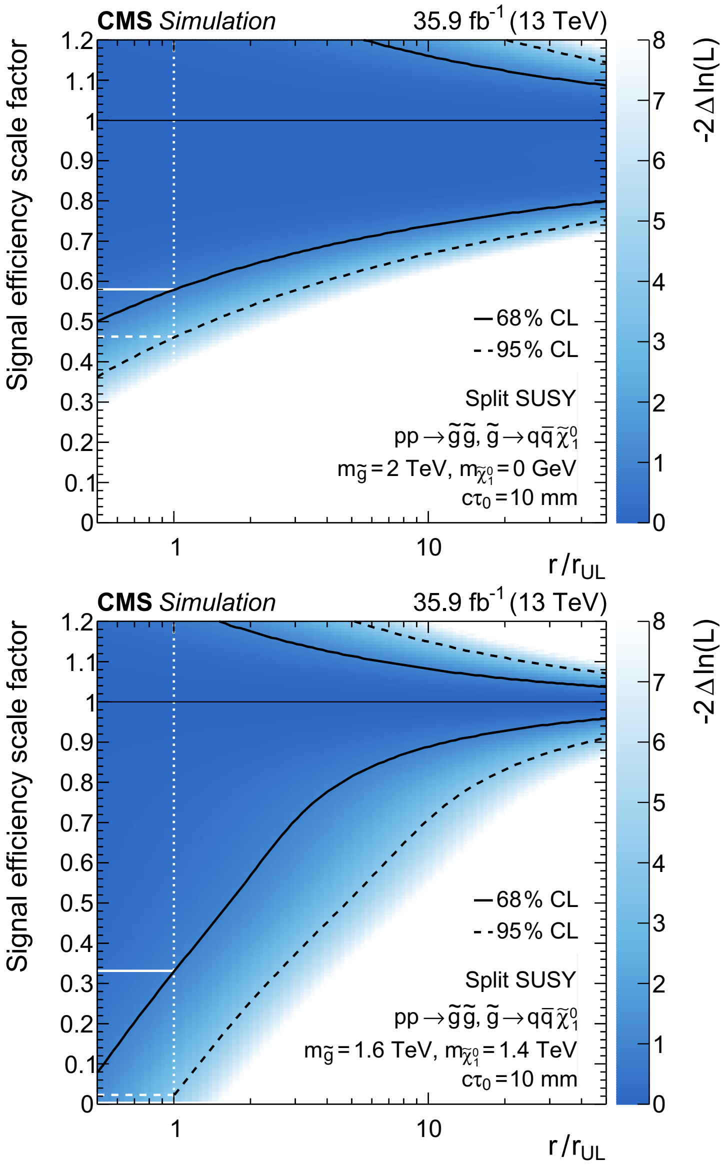

Figure 8. The negative log-likelihood of a maximum likelihood fit to the Asimov data set as a function of the signal efficiency scale factor and r/r_{UL} for a (upper) uncompressed and (lower) compressed scenario. The black solid (dashed) line indicate the 68 (95)% CL interval, while for $r = r_{\mathrm{UL}}$ (white dotted line) the white solid and dashed lines indicate the SF constraints at 68% and 95% CL, respectively. The product of the LLP jet tagger efficiency and the SF is bound to $[0, 1]$.

cross section relative to the theoretical value, for two split SUSY benchmark scenarios with an uncompressed and compressed mass spectrum, and $c\tau_0 = 10$ mm. The model assumptions are indicated in the figure legends. The SF is constrained to > 0.6 (> 0.3) at 68% CL for $r/r_{\mathrm{UL}} = 1$ for the uncompressed (compressed) model, and it is bound to $\lesssim 1.2$ by the condition SF $\epsilon_{\mathrm{MC}} \in [0, 1]$. The figure demonstrates that the likelihood is not degenerate with respect to the scale factor and signal strength. Instead, the (N_{jet}, $P(LLP|ctau)$) categorisation scheme allows the SF to be constrained *in situ* in the signal-plus-background fit when adding the SF as a nuisance parameter to the likelihood model, described in section 8.1.

Figure 9 summarises the expected lower limit on $m_{\tilde{g}}$ (95% CL) as a function of $c\tau_0$ for simplified models of split SUSY that assume the production of gluino pairs. The model assumptions are indicated by the legend in each panel. The upper (lower) panel presents the expected mass exclusions for models with an uncompressed (compressed) mass spectrum. A lower limit on the gluino mass in excess of 2.1 (1.5) TeV is obtained for models with an uncompressed (compressed) mass spectrum. The mass exclusions are determined by assuming an integrated luminosity of 35.9 fb^{-1}. This permits a comparison with the exclusions reported by an inclusive search for SUSY [27] in final states containing jets and $p_{\mathrm{T}}^{\mathrm{miss}}$, over the same range in $c\tau_0$ values. Significant gains in excluded values of $m_{\tilde{g}}$, of up to approximately 500 GeV, are expected for $c\tau_0 \gtrsim 1$ mm. The coverage is also competitive with respect to a dedicated reconstruction technique that is reported in reference [23]. For the region $c\tau_0 < 1$ mm, the tagger performance degrades

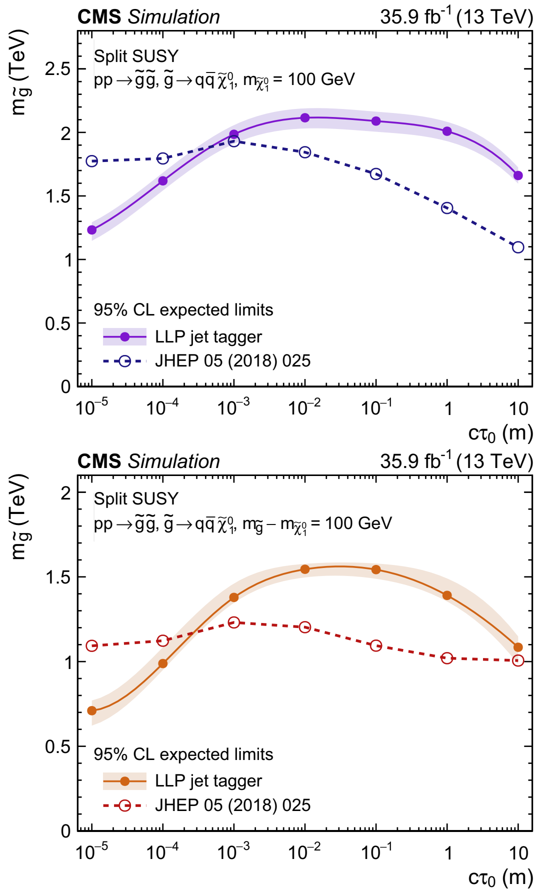

Figure 9. Expected 95% CL lower limits on $m_{\tilde{g}}$ as a function of $c\tau_0$ for split SUSY models with an uncompressed (upper) and a very compressed (lower) mass spectrum. The shaded bands indicate the total uncertainty from both statistical and systematic sources. The model assumptions are indicated by the legends. The results are compared to the expected limits obtained in reference [27], indicated by the dashed lines.

because of a limited ability to tag an LLP jet in the vicinity of the primary pp interaction vertex, while the reference search is able to exploit the distinguishing kinematical features of split SUSY events through a finer categorisation of candidate signal events. For the region $c\tau_0 > 1$ m, the LLPs frequently decay outside the experimental acceptance, which leads to an increased reliance on the presence of initial-state radiation.

9. Summary

Many models of new physics beyond the standard model predict the production of long-lived particles (LLPs) in proton-proton (pp) collisions at the LHC. Jets arising from the decay of LLPs (LLP jets) can be appreciably displaced from the pp collisions. A novel tagger to identify LLP jets is presented. The tagger employs a deep neutral network (DNN) using an architecture inspired by the CMS DeepJet algorithm. Simplified models of split supersymmetry (SUSY), which yield neutralinos and LLP jets from the decay of long-lived gluinos, are used to train the DNN and demonstrate its performance.

The application of various techniques related to the tagger are reported. A custom labelling scheme for LLP jets based on generator-level information from Monte Carlo programs is defined. The proper decay length $c\tau_0$ of the gluino is used as an external parameter to the DNN, which allows hypothesis testing over several orders of magnitude in $c\tau_0$ with a single DNN. The DNN is trained using samples of both simulated events and pp collision data. The application of domain adaptation by backward propagation significantly improves the agreement of the DNN output for simulation and data, by an order of magnitude according to the Jensen–Shannon divergence, when compared to training the DNN with simulation only. The method is validated using signal-depleted control samples of pp collisions at a centre-of-mass energy of 13 TeV. The samples were recorded by the CMS experiment and correspond to an integrated luminosity of 35.9 fb^{-1}. Training the DNN with pp collision data does not significantly degrade the tagger performance. The tagger rejects 99.99% of light-flavour jets from standard model processes, as measured in an inclusive $t\bar{t}$ sample, while retaining approximately 30%–80% of LLP jets for split SUSY models with 1 mm$\leq c\tau_0\leq$ 10 m and a gluino-neutralino mass difference of at least 200 GeV.

Finally, the potential performance of the tagger is demonstrated in the framework of a search for split SUSY in final states containing jets and significant missing transverse momentum. Simulated event samples provide the expected contributions from standard model background processes. Candidate signal events were categorised according to the scalar sum of jet momenta, the number of jets, and the number of tagged LLP jets. Expected lower limits on the gluino mass at 95% confidence level are determined with a binned likelihood fit as a function of $c\tau_0$ in the range from 10 μm to 10 m. A procedure to constrain a correction to the LLP jet tagger efficiency in the likelihood fit is introduced. Competitive limits are demonstrated: models with a long-lived gluino of mass $\gtrsim$2 TeV, a neutralino mass of 100 GeV, and a proper decay length in the range 1 mm$\leq c\tau_0\leq$ 1 m are expected to be excluded by this search.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, PUT and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); NKFIA (Hungary); DAE and DST (India); IPM (Iran, Islamic Republic Of); SFI (Ireland); INFN (Italy); MSIP and NRF (Korea, Republic Of); MES (Latvia); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR, and NRC KI (Russian Federation); MESTD (Serbia); SEIDI, CPAN, PCTI, and FEDER (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU (Ukraine); STFC (United Kingdom of Great Britain and Northern Ireland); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, contract Nos. 675440, 752730, and 765710 (European Union); the Leventis Foundation; the A.P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the F.R.S.-FNRS and FWO (Belgium) under the 'Excellence of Science—EOS' – be.h project n. 30820817; the Beijing Municipal Science and Technology Commission, No. Z181100004218003; the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy—EXC 2121 'Quantum Universe' – 390833306; the Lendület ('Momentum') Program and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the New National Excellence Program ÚNKP, the NKFIA research grants 123842, 123959, 124845, 124850, 125105, 128713, 128786, and 129058 (Hungary); the Council of Science and Industrial Research, India; the HOMING PLUS program of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus program of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Ministry of Science and Education, grant no. 14.W03.31.002 6 (Russian Federation); the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia María de Maeztu, grant MDM-2015-0509 and the Programa Severo Ochoa del Principado de Asturias; the Thalis and Aristeia programs cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); the Kavli Foundation; the Nvidia Corporation; the SuperMicro Corporation; the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

Data availability

The authors will deposit their research data in accordance with the CMS data preservation, re-use and open access policy. Please also see the CMS Public Pages.

References

[1] Albertsson K *et al* 2018 Machine learning in high energy physics community white paper *18th Int. Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017), Proc., Seattle, USA, August 21-25, 2017* p 0220008 arXiv180702876

[2] Larkoski A J, Moult I and Nachman B 2017 Jet Substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning *Phys. Rept.* **841** 1–63

[3] ATLAS Collaboration 2008 The ATLAS experiment at the CERN Large Hadron Collider *JINST* **3** S08003

[4] CMS Collaboration 2008 The CMS experiment at the CERN LHC *JINST* **3** S08004

[5] CMS Collaboration 2018 Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV *JINST* **13** P05011 arXiv171207158

[6] ATLAS Collaboration 2019 Atlas b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV Submitted to *Eur. Phys. J. C* arXiv190705120

[7] Stoye M 2018 Deep learning in jet reconstruction at CMS *18th Int. Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)* p 042029

[8] CMS Collaboration 2018 Performance of the Deepjet b Tagging Algorithm Using 41.9/fb of Data From Proton-Proton Collisions at 13 TeV With Phase 1 CMS Detector *CMS Detector Performance Note* CMS-DP-2018-058 CERN http://cds.cern.ch/record/2646773

[9] Arkani-Hamed N and Dimopoulos S 2005 Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC *JHEP* **06** 073

[10] Giudice G F and Romanino A 2004 Split supersymmetry *Nucl. Phys.* B **699** 65 (Erratum: *Nucl. Phys. B* 706 (2005) 487)

[11] Hewett J L, Lillie B, Masip M and Rizzo T G 2004 Signatures of long-lived gluinos in split supersymmetry *JHEP* **09** 070

[12] Giudice G F and Rattazzi R 1999 Theories with gauge mediated supersymmetry breaking *Phys. Rept.* **322** 419

[13] Barbier R *et al* 2005 R-parity violating supersymmetry *Phys. Rept.* **420** 1

[14] Strassler M J and Zurek K M 2008 Discovering the Higgs through highly-displaced vertices *Phys. Lett.* B **661** 263

[15] Han T, Si Z, Zurek K M and Strassler M J 2008 Phenomenology of hidden valleys at hadron colliders *JHEP* **07** 008

[16] Alves D *et al* 2012 Simplified models for LHC new physics searches *J. Phys.* G **39** 105005

[17] Buchmüller O *et al* 2017 Simplified models for displaced dark matter signatures *JHEP* **09** 076

[18] Alimena J *et al* 2019 *Searching for Long-Lived Particles Beyond the Standard Model at the Large Hadron Collider* arXiv190304497

[19] Farrar G R and Fayet P 1978 Phenomenology of the production, decay and detection of new hadronic states associated with supersymmetry *Phys. Lett.* B **76** 575

[20] ATLAS Collaboration 2019 Search for long-lived neutral particles in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS calorimeter *Eur. Phys. J.* C **79** 481 arXiv190203094

[21] ATLAS Collaboration 2019 Search for heavy charged long-lived particles in the ATLAS detector in 36.1 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV *Phys. Rev.* D **99** 092007 arXiv190201636

[22] ATLAS Collaboration 2019 Search for long-lived particles produced in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer *Phys. Rev.* D **99** 052005 arXiv181107370

[23] ATLAS Collaboration 2018 Search for long-lived, massive particles in events with displaced vertices and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector *Phys. Rev.* D **97** 052012 arXiv171004901

[24] CMS Collaboration 2019 Search for long-lived particles using nonprompt jets and missing transverse momentum with proton-proton collisions at $\sqrt{s} = 13$ TeV *Phys. Lett.* B **797** 134876 arXiv190606441

[25] CMS Collaboration 2019 Search for long-lived particles decaying into displaced jets in proton-proton collisions at $\sqrt{s} = 13$ TeV *Phys. Rev.* D **99** 032011 arXiv181107991

[26] CMS Collaboration 2018 Search for long-lived particles with displaced vertices in multijet events in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev.* D **98** 092011 arXiv180803078

[27] CMS Collaboration 2018 Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in final states with jets and missing transverse momentum *JHEP* **05** 025 arXiv180202110

[28] CMS Collaboration 2018 Search for decays of stopped exotic long-lived particles produced in proton-proton collisions at $\sqrt{s} = 13$ TeV *JHEP* **05** 127 arXiv180100359

[29] CMS Collaboration 2018 Search for new long-lived particles at $\sqrt{s} = 13$ TeV *Phys. Lett.* B **780** 432 arXiv171109120

[30] CMS Collaboration 2016 Search for long-lived charged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV *Phys. Rev.* D **94** 112004 arXiv160908382

[31] LHCb Collaboration 2017 Updated search for long-lived particles decaying to jet pairs *Eur. Phys. J.* C **77** 812 arXiv170507332

[32] Baldi P *et al* 2016 Parameterized neural networks for high-energy physics *Eur. Phys. J.* C **76** 235

[33] Graham P W, Kaplan D E, Rajendran S and Saraswat P 2012 Displaced supersymmetry *JHEP* **07** 149

[34] Ganin Y and Lempitsky V 2014 *Unsupervised Domain Adaptation by Backpropagation* arXiv14097495

[35] Cacciari M, Salam G P and Soyez G 2008 The anti-k_{T} jet clustering algorithm *JHEP* **04** 063

[36] Cacciari M, Salam G P and Soyez G 2012 FASTJET user manual *Eur. Phys. J.* C **72** 1896

[37] CMS Collaboration 2017 Particle-flow reconstruction and global event description with the CMS detector *JINST* **12** P10003 arXiv170604965

[38] CMS Collaboration 2015 Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV *JINST* **10** P08010 arXiv150202702

[39] CMS Collaboration 2015 Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV *JINST* **10** P06005 arXiv150202701

[40] CMS Collaboration 2018 Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV *JINST* **13** P06015 arXiv180404528

[41] CMS Collaboration 2011 Measurement of $\mathrm{B}\bar{\mathrm{B}}$ angular correlations based on secondary vertex reconstruction at $\sqrt{s} = 7$ TeV *JHEP* **03** 136 arXiv11023194

[42] CMS Collaboration 2014 Description and performance of track and primary-vertex reconstruction with the CMS tracker *JINST* **9** P10009 arXiv14056569

[43] Cacciari M and Salam G P 2008 *Pileup subtraction using jet areas* **659** 119 *Phys. Lett. B* arXiv07071378

[44] CMS Collaboration 2017 Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV *JINST* **12** P02014 arXiv160703663

[45] CMS Collaboration 2017 Jet Algorithms Performance in 13 TeV Data *CMS Physics Analysis Summary* CMS-PAS-JME-16-003 CERN https://cds.cern.ch/record/2256875

[46] CMS Collaboration 2019 Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector *JINST* **14** P07004 arXiv190306078

[47] CMS Collaboration 2017 The CMS trigger system *JINST* **12** P01020 arXiv160902366

[48] CMS Collaboration 2012 Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV *JINST* **7** P10002 arXiv:1206.4071

[49] CMS Collaboration 2011 Search for supersymmetry in pp collisions at 7 TeV in events with jets and missing transverse energy *Phys. Lett.* B **698** 196 arXiv11011628

[50] Dimopoulos S, Raby S and Wilczek F 1981 Supersymmetry and the scale of unification *Phys. Rev.* D **24** 1681

[51] Ibanez L E and Ross G G 1981 Low-energy predictions in supersymmetric grand unified theories *Phys. Lett.* B **105** 439

[52] Marciano W J and Senjanović G 1982 Predictions of supersymmetric grand unified theories *Phys. Rev.* D **25** 3092

[53] ATLAS Collaboration 2012 Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC *Phys. Lett.* B **716** 1 arXiv12077214

[54] CMS Collaboration 2012 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC *Phys. Lett.* B **716** 30 arXiv12077235

[55] Alwall J *et al* 2014 The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations *JHEP* **07** 079

[56] Sjöstrand T *et al* 2015 An introduction to PYTHIA 8.2 *Comput. Phys. Commun.* **191** 159

[57] Fairbairn M *et al* 2007 Stable massive particles at colliders *Phys. Rept.* **438** 1

[58] Kraan A C 2004 Interactions of heavy stable hadronizing particles *Eur. Phys. J.* C **37** 91

[59] Mackeprang R and Rizzi A 2007 Interactions of coloured heavy stable particles in matter *Eur. Phys. J.* C **50** 353

[60] Frederix R and Frixione S 2012 Merging meets matching in MC@NLO *JHEP* **12** 061

[61] Nason P 2004 A new method for combining NLO QCD with shower Monte Carlo algorithms *JHEP* **11** 040

[62] Frixione S, Nason P and Oleari C 2007 Matching NLO QCD computations with parton shower simulations: the POWHEG method *JHEP* **11** 070

[63] Alioli S, Nason P, Oleari C and Re E 2010 A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG box *JHEP* **06** 043

[64] Frixione S, Nason P and Ridolfi G 2007 A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction *JHEP* **09** 126

[65] Alioli S, Nason P, Oleari C and Re E 2009 NLO single-top production matched with shower in POWHEG: *s*- and *t*-channel contributions *JHEP* **09** 111 (Erratum: *JHEP* 02 (2010) 011)

[66] Re E 2011 Single-top Wt-channel production matched with parton showers using the POWHEG method *Eur. Phys. J.* C **71** 1547

[67] Gavin R, Li Y, Petriello F and Quackenbush S 2013 W physics at the LHC with FEWZ 2.1 *Comput. Phys. Commun.* **184** 208 arXiv12015896

[68] Gavin R, Li Y, Petriello F and Quackenbush S 2011 FEWZ 2.0: a code for hadronic *Z* production at next-to-next-to-leading order *Comput. Phys. Commun.* **182** 2388

[69] Czakon M and Mitov A 2014 Top++: a program for the calculation of the top-pair cross-section at hadron colliders *Comput. Phys. Commun.* **185** 2930

[70] Borschensky C *et al* 2014 Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV *Eur. Phys. J.* C **74** 174

[71] CMS Collaboration 2016 Event generator tunes obtained from underlying event and multiparton scattering measurements *Eur. Phys. J.* C **76** 155 arXiv151200815

[72] CMS Collaboration 2016 Investigations of the impact of the parton shower tuning in PYTHIA 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV *CMS Physics Analysis Summary* CMS-PAS-TOP-16-021 CERN

[73] Ball R D *et al* 2015 Parton distributions for the LHC Run II *JHEP* **04** 040 arXiv14108849

[74] Agostinelli A *et al* 2003 GEANT4—a simulation toolkit *Nucl. Instrum. Meth.* A **506** 250

[75] Bertolini D, Harris P, Low M and Tran N 2014 Pileup per particle identification *JHEP* **10** 059

[76] CMS Collaboration 2019 Pileup Mitigation at CMS in 13 TeV Data *CMS Physics Analysis Summary* CMS Physics Analysis Summary CMS-PAS-JME-18-001 CERN https://cds.cern.ch/record/2683784

[77] Glorot X, Bordes A and Bengio Y 2011 Deep sparse rectifier neural networks *Proc. of the Fourteenth Int. Conf. on Artificial Intelligence and Statistics* vol 15 p 315 PMLR

[78] Srivastava N *et al* 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res* **15** 1929

[79] Ryzhikov A and Ustyuzhanin A 2018 Domain adaptation with gradient reversal for MC/real data calibration *18th Int. Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)* p 042018

[80] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization *Proc. Third Int. Conf. for Learning Representations* arXiv14126980

[81] Chollet F *et al* 2015 KERAS https://keras.io

[82] Abadi M *et al* 2015 TENSORFLOW: large-scale machine learning on heterogeneous systems arXiv160304467

[83] Brun R and Rademakers F 1997 ROOT: an object oriented data analysis framework *New Computing Techniques in Physics Research V. Proc., 5th Int. Workshop, AIHENP '96, Lausanne, Switzerland, September 2-6, 1996* p 81

[84] Komm M *et al* 2019 ROOT-Based Preprocessing Pipeline for TENSORFLOW/KERAS https://github.com/matt-komm/ROOT-TF-pipeline

[85] Lin J 2006 Divergence measures based on the Shannon entropy *IEEE Trans. Inf. Theor.* **37** 145

[86] Kallweit S *et al* 2016 NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging *JHEP* **04** 021 arXiv151108692

[87] CMS Collaboration 2018 Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV *JHEP* **07** 161 arXiv180202613

[88] Kalogeropoulos A and Alwall J 2018 *The Syscalc Code: a Tool to Derive Theoretical Systematic Uncertainties* arXiv180108401

[89] CMS Collaboration 2017 CMS luminosity measurements for the 2016 data taking period *CMS Physics Analysis Summary* CMS-PAS-LUM-17-001 CERN http://cds.cern.ch/record/2257069

[90] Barlow R J and Beeston C 1993 Fitting using finite Monte Carlo samples *Comput. Phys. Commun.* **77** 219

[91] Cowan G, Cranmer K, Gross E and Vitells O 2011 Asymptotic formulae for likelihood-based tests of new physics *Eur. Phys. J.* C **71** 1554 (Erratum: *Eur. Phys. J. C* 73 (2013) 2501)

[92] ATLAS and CMS Collaborations, and LHC Higgs Combination Group 2011 Procedure for the LHC Higgs boson search combination in summer 2011 *Technical Report* CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11 CERN https://cds.cern.ch/record/1379837

[93] Junk T 1999 Confidence level computation for combining searches with small statistics *Nucl. Instr. Meth.* A **434** 435

[94] Read A L 2002 Presentation of search results: the CL_s technique *J. Phys.* G **28** 2693