

Probing the inert doublet model using jet substructure with a multivariate analysis

Akanksha Bhardwaj,^{1,2,*} Partha Konar^{1,†} Tanumoy Mandal,^{3,‡} and Soumya Sadhukhan^{4,§}

¹Physical Research Laboratory (PRL), Ahmedabad - 380009, Gujarat, India

²Indian Institute of Technology, Gandhinagar-382424, Gujarat, India

³Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden

⁴Department of Physics and Astrophysics, University of Delhi, Delhi 110 007, India



(Received 26 June 2019; published 30 September 2019)

We explore the challenging but phenomenologically interesting hierarchical mass spectrum of the inert doublet model where relatively light dark matter along with much heavier scalar states can fully satisfy the constraints on the relic abundance and also fulfill other theoretical as well as collider and astrophysical bounds. To probe this region of parameter space at the LHC, we propose a signal process that combines up to two large radius boosted jets along with substantial missing transverse momentum. Aided by our intuitive signal selection, we capture a hybrid process where the di-fatjet signal is significantly enhanced by the mono-fatjet contribution with minimal effects on the SM di-fatjet background. Substantiated by the sizable mass difference between the scalars, these boosted jets, originally produced from the hadronic decay of massive vector bosons, still carry the inherent footprint of their root. These features implanted inside the jet substructure can provide additional handles to deal with a large background involving QCD jets. We adopt a multivariate analysis using a boosted decision tree to provide a robust mechanism to explore the hierarchical scenario, which would bring almost the entire available parameter space well within reach of the 14 TeV LHC runs with high luminosity.

DOI: [10.1103/PhysRevD.100.055040](https://doi.org/10.1103/PhysRevD.100.055040)

I. INTRODUCTION

The Standard Model (SM) of particle physics encapsulates our knowledge of fundamental interactions of the particle world with all its glory. Until now, apart from a few minor exceptions, the SM is in perfect agreement with all the high energy collider experiments like the Large Hadron Collider (LHC) experiments at CERN. The reputation of the SM, being a complete theory, gets tarnished when it cannot explain the presence of tiny yet nonzero masses of the neutrinos that are already established in neutrino oscillation experiments. Observations of cosmic microwave background radiation in various experiments unambiguously establish that 26% of the energy budget of our Universe is made up of an inert, stable component, termed “dark matter” (DM). The SM does not contain any particle that can satisfy the observed density of the DM, along with

explaining its other properties. The inert doublet model (IDM) is proposed [1,2] as a minimal extension of the SM that can provide an inert weakly interacting DM candidate, stabilized by the discrete symmetry of the model. The SM is extended with an extra $SU(2)_L$ scalar doublet which is odd under a discrete \mathbb{Z}_2 symmetry, and thus stabilizes the lightest neutral scalar of the model to be an ideal DM candidate. The neutrino mass can also be arranged in this setup by introducing lepton portals [3].

Inspection of the dark sector of the IDM, as done in Refs. [4,5], reveals that only small islands of parameter space can satisfy the full relic density of the DM dictated by the WMAP and the Planck results. Only light DM with a mass close to half of the Higgs mass can produce full relic density through the resonant Higgs portal annihilation. Even so, this requires a large mass difference between the DM with the additional beyond the SM (BSM) scalars in the model. Otherwise, the co-annihilation effects in a degenerate mass spectrum reduce the relic density to underproduce DM. Another part of the parameter space where one can explain the full DM relic density is for the DM mass, $m_{DM} \gtrsim 550$ GeV and even that is possible for an extremely degenerate BSM scalar mass spectrum. Among the various DM scenarios discussed above, heavy DM with a hierarchical spectrum can accommodate only about a few percent of the observed relic density and thus makes this

*akanksha@prl.res.in

†konar@prl.res.in

‡tanumoy.mandal@physics.uu.se

§physicsoumya@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

scenario uninteresting to probe at the LHC. One of the earlier collider studies of the IDM is performed in Ref. [6]. The dilepton and trilepton signatures at the LHC originating from the IDM have been investigated in Refs. [7,8]. Although the degenerate heavy DM scenario can provide full relic density, the inertness of the model leads to kinematic suppression in heavy DM production at colliders and therefore, makes the signal very weak. Moreover, detection of the soft decay products from such a compressed mass spectrum remains challenging due to poor signal efficiency. This scenario is probed using charged track signatures by the CMS Collaboration [9].

Among the two light DM scenarios in this model, the degenerate BSM scalar mass spectrum can satisfy only about 10% of the observed DM relic density. Nonetheless, this case is probed at the LHC through the mono jet search in Ref. [10]. We motivate our framework with the light DM along with the hierarchical mass spectrum where the full DM relic density is achieved. Although very challenging due to tiny production cross sections of the unstable heavy BSM scalars at the LHC, their significant mass differences with the DM candidate give rise to interesting signal topology characterized by two boosted jets along with large missing transverse energy (MET) from the DM production. This gives us a scope to employ sophisticated multivariate analysis equipped with jet substructure variables to isolate the signal from an overwhelmingly large SM background. The search for BSM scalars for this case is studied in the dijet plus MET channel, in a recent study [11]. All these searches do not exhibit bright discovery potential even with the highest possible LHC luminosity. We look to explore a suitable discovery potential of this scenario in this paper.

To reiterate the scenario under consideration, the heavier BSM scalars (a pseudoscalar A and charged Higgs H^\pm) reside in the mass range ($\sim 250\text{--}700$ GeV) much higher than the small mass window ($55\text{--}80$ GeV) where the DM candidate can lie. Hence, the hierarchical mass difference is large enough for these heavy scalars to decay dominantly to vector bosons ($V = \{W, Z\}$), which in turn are sufficiently boosted and corresponding hadronic deposits at the calorimeter behave like large radius fatjets, characterized by the jet radius, $R \sim p_T/2m_V \gtrsim 0.8$. Accompanied with large MET acquired by the undetected pair of DM particles, the presence of these fatjets in the signal brings additional variables like fatjet mass (M_J) and subjettness (τ_{21}), which carry the characteristics of boosted W/Z decay. These observables are perceived as the ones that can distinguish well between the signal and the background, dominantly coming from the SM $V + \text{jets}$, as only a tiny fraction of QCD jets mimic as boosted jets. Still, when the overwhelmingly large cross section of the background is pitted against the suppressed signal cross section due to inertness of the model, even a tiny fraction can overshadow the signal to deny a significant discovery potential.

There is a possibility of mono-fatjet signal topology [12,13] with roughly 1 order of magnitude higher cross section than the di-fatjet one. In this case, although we have a more significant cross section, the corresponding background becomes uncontrollably large. Therefore, the mono-fatjet topology alone is not sufficient to achieve discovery significance. The di-fatjet topology, on the other hand, also by itself is not enough to produce discovery significance due to tiny production cross section. In this paper, we propose a hybrid topology where the signal selections are designed aiming the di-fatjet topology but also allow a substantial fraction of the mono-fatjet signal. In doing this, we not only gain in the signal but at the same time, the enormous mono-fatjet background can also be tamed down.

A probe of the IDM using a cut-based analysis (CBA) in the di-fatjet plus MET channel, has failed to reach discovery significance of 5σ in any of our chosen benchmark points. From the nature of event distribution profiles, it is observed that the two variables viz. M_J and τ_{21} are very powerful to separate a tiny signal from the enormous SM background. The discovery significance in CBA is still elusive even if these jet substructure variables are used to the hilt. A multivariate analysis (MVA) in general performs better than a CBA, if appropriate kinematic variables are used in the analysis. So, a sophisticated MVA involving jet substructure variables quoted above is imperative to achieve better discovery potential in the IDM. With the events selected only after baseline cuts (defined later), the signal is still too tiny compared to the background to train the MVA setup. Therefore, the baseline selection criteria should be accompanied by stronger selection cuts at the baseline level to cut down the large background without harming the signal too much before passing events to MVA. These cuts should be chosen optimally; otherwise, if they are very similar to the one used in CBA, eventually reduce the predictive power of MVA. Finally, our selection cuts are designed in a way such that it allows signal to consist of two high- p_T fatjets. Along with the pure di-fatjet signal, substantial contribution comes from the events with mono-fatjet that mimics the signal. This admixture significantly increases the signal cross section but simultaneously bring in some extra background processes in the picture. We perform an MVA analysis coupled with jet substructure variables to achieve improved signal vs background discrimination, which could not be realized in the CBA. It helps us to reach a significantly higher LHC discovery potential in the di-fatjet plus MET channel of the IDM.

This paper is organized as follows. In Sec. II we briefly discuss the IDM, outlining its scalar sector. Next, in Sec. III, we invoke all the possible theoretical, collider, and astrophysical constraints applicable to the IDM, to ascertain the viability of hierarchical BSM scalar sector along with the presence of a light DM. Subsequently, in Sec. IV, we discuss four possible DM scenarios depending

on the DM mass and its mass differences to the other BSM scalars to motivate our choice of benchmark points. To define our analysis setup, we list the possible IDM processes contributing to our signal process, di-fatjet plus MET channel in Sec. V. We also discuss all possible SM backgrounds for this channel. At this point, we present our sample benchmark points covering our region of interest, which is also consistent with all the discussed constraints. In Sec. VI, we first use the baseline cuts and then introduce two fatjet specific observables and study how these perform to increase the signal vs background ratio and obtain the LHC reach for all the benchmark points. In Sec. VII, we improve our probe using MVA to recast the signal vs background numbers with nonrectangular cuts and therefore, having better sensitivity for the LHC search. Finally, we summarize our results and conclude in Sec. VIII.

II. INERT DOUBLET MODEL

We first discuss the traditional IDM where one adds an additional $SU(2)_L$ complex scalar doublet Φ_2 apart from the SM Higgs doublet Φ_1 , which are, respectively, odd and even under a discrete \mathbb{Z}_2 symmetry, i.e., $\Phi_1 \rightarrow \Phi_1$, $\Phi_2 \rightarrow -\Phi_2$. The most general scalar potential that respects the electroweak symmetry $SU(2)_L \otimes U(1)_Y \otimes \mathbb{Z}_2$ of the IDM can be written as [5]

$$\begin{aligned} V(\Phi_1, \Phi_2) = & \mu_1^2 \Phi_1^\dagger \Phi_1 + \mu_2^2 \Phi_2^\dagger \Phi_2 + \lambda_1 (\Phi_1^\dagger \Phi_1)^2 \\ & + \lambda_2 (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 \\ & + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} [(\Phi_1^\dagger \Phi_2)^2 + \text{H.c.}], \end{aligned} \quad (1)$$

where Φ_1 and Φ_2 both are hypercharged, $Y = +1$, and can be written as

$$\Phi_1 = \begin{pmatrix} G^+ \\ \frac{v+h+iG^0}{\sqrt{2}} \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} H^+ \\ \frac{H+iA}{\sqrt{2}} \end{pmatrix}. \quad (2)$$

Here h is the SM Higgs with G^+ , G^0 being the charged and neutral Goldstone bosons, respectively. The charged scalar H^+ is present in Φ_2 , along with the neutral scalars, H , A , respectively, being CP even and CP odd. For the vacuum expectation values (VEVs) of the two doublets, we adopt the notation $\langle \Phi_1 \rangle = v/\sqrt{2}$, $\langle \Phi_2 \rangle = 0$, keeping in mind the exact nature of the \mathbb{Z}_2 symmetry. The zero VEV of Φ_2 is responsible for the inertness of this model. Since all the SM fermions are even under \mathbb{Z}_2 , the new scalar doublet does not couple to the SM fermions and thus having no fermionic interactions. The scalar-gauge boson interactions originate through the kinetic term of the two doublets

$$\mathcal{L}_{\text{kin}} = (D_\mu \Phi_1)^\dagger (D^\mu \Phi_1) + (D_\mu \Phi_2)^\dagger (D^\mu \Phi_2). \quad (3)$$

All parameters in the scalar potential are assumed to be real in order to keep the IDM CP invariant.

Here, the electroweak symmetry breaking takes place through the SM doublet Φ_1 getting a VEV, and after this, the masses of the physical scalars at tree level can be written as

$$\begin{aligned} m_h^2 &= 2\lambda_1 v^2, \\ m_{H^\pm}^2 &= \mu_2^2 + \frac{1}{2}\lambda_3 v^2, \\ m_H^2 &= \mu_2^2 + \frac{1}{2}(\lambda_3 + \lambda_4 + \lambda_5)v^2 = m_{H^\pm}^2 + \frac{1}{2}(\lambda_4 + \lambda_5)v^2, \\ m_A^2 &= \mu_2^2 + \frac{1}{2}(\lambda_3 + \lambda_4 - \lambda_5)v^2 = m_{H^\pm}^2 + \frac{1}{2}(\lambda_4 - \lambda_5)v^2. \end{aligned} \quad (4)$$

Here, m_h is the SM-like Higgs boson mass, and $m_{H(A)}$ are the masses of the CP -even (odd) scalars from the inert doublet, while m_{H^\pm} is the charged scalar mass. Either of the neutral scalars can be the DM candidate in this IDM framework since DM observations cannot probe the CP behavior. For the present analysis, we consider the CP -even scalar H as the DM candidate, which corresponds to negative values of λ_5 parameter. We define $\lambda_3 + \lambda_4 + \lambda_5 = \lambda_L$, which can be either positive or negative. The relations between the λ 's and the scalar masses get modified when the QED corrections are considered for both the scalar masses and scalar potential parameters. As the inert scalars do not couple to the SM quarks, higher-order QCD corrections are negligible for these parameters.

Compared to the SM, only the scalar sector is modified in the IDM. Similar to the SM, λ_1 and μ_1^2 are determined by $m_h \approx 125$ and $v \approx 246$ GeV. There are five free parameters in the scalar sector of the IDM viz. λ_L , λ_2 , m_A , m_{H^\pm} , and m_H that are expressed in terms of the five scalar potential parameters, μ_2^2 and $\lambda_{2,3,4,5}$, as shown in Eq. (4). The new doublet, being inert to the SM fermionic sector, does not introduce any additional new parameters in this setup. In IDM the contribution of the self-coupling parameter λ_2 is mostly limited to fixing unitarity and stability of the potential. It does not affect the scalar masses and their phenomenology. The Higgs portal coupling λ_L to the chosen DM candidate H determines the rate of the DM annihilation through the Higgs and, therefore, is an essential parameter in the DM sector along with the DM mass $m_{\text{DM}} = m_H$. The collider phenomenology of the IDM depends on the scalar masses m_{H^\pm} , m_A and m_H , as the mass differences between them play a significant role in proposing the search channels for different scenarios.

III. CONSTRAINTS ON THE INERT DOUBLET MODEL

The IDM parameter space is constrained from various theoretical as well as experimental considerations. In the

model, we have an extra doublet which brings extra parameters in the scalar potential. Therefore, it is imperative to check whether the extended potential is bounded from below, i.e., stable at tree level along with the potential parameters being within the unitary and perturbative regime. With the presence of the extra doublet, oblique parameters should be reexamined with respect to the presence of a light DM and custodial symmetry breaking, respectively. The presence of light scalars can also upset the LEP constraints and the Higgs invisible decay limits. Since DM is present in the model, we should satisfy the observed relic density keeping the DM-nucleon interactions below the DM direct detection reach.

A. Theoretical constraints

The scalar sector is modified in the IDM. We ensure the enlarged potential is stable, i.e., not unbounded from below and the global minimum is a neutral one. We also checked if the potential parameters are perturbative at the tree level along with satisfying unitarity bounds.

Bounded from below.—Theoretical constraints on quartic potential parameters (λ 's) can arise from restricting the scalar potential in Eq. (3) not to produce large negative numbers for large field values, i.e., $V > 0 \forall \Phi_i \rightarrow \pm\infty$. The mixed quartic terms can be combined to form complete square terms and demanding their coefficients to be positive, leads to the following conditions¹:

$$\begin{aligned} \lambda_1 > 0, \quad \lambda_2 > 0, \quad \lambda_3 + 2\sqrt{\lambda_1\lambda_2} > 0, \\ \lambda_3 + \lambda_4 + \lambda_5 + 2\sqrt{\lambda_1\lambda_2} > 0. \end{aligned} \quad (5)$$

Because of the introduction of new scalars, there are possibilities of having multiple minima. For the inert vacuum to be the global minimum, we restrict it from being charged by ensuring the condition

$$\lambda_4 + \lambda_5 < 0. \quad (6)$$

We also ensure that the global minimum is the inert vacuum as opposed to an inertlike one, with the imposition of the condition [15,16]

$$\frac{\mu_1^2}{\sqrt{\lambda_1}} - \frac{\mu_2^2}{\sqrt{\lambda_2}} > 0. \quad (7)$$

Perturbativity and unitarity.—We form the S matrix with the amplitudes computed from the $2 \rightarrow 2$ scalar scattering processes taking into account all the other quartic terms in the scalar potential. The eigenvalues of the S matrix turn out to be some combinations of these couplings.

¹Alternately, for a scalar potential with many quartic couplings, one can consider formulating the copositive matrices to guarantee the boundedness of the potential [14].

The perturbative unitarity constraints on those eigenvalues are $|\Lambda_i| \leq 8\pi$, where the scattering matrix provides us the combinations as

$$\begin{aligned} \Lambda_{1,2} &= \lambda_3 \pm \lambda_4; \quad \Lambda_{3,4} = \lambda_3 \pm \lambda_5; \quad \Lambda_{5,6} = \lambda_3 + 2\lambda_4 \pm 3\lambda_5; \\ \Lambda_{7,8} &= -\lambda_1 - \lambda_2 \pm \sqrt{(\lambda_1 - \lambda_2)^2 + \lambda_4^2}; \\ \Lambda_{9,10} &= -3\lambda_1 - 3\lambda_2 \pm \sqrt{9(\lambda_1 - \lambda_2)^2 + (2\lambda_3 + \lambda_4)^2}; \\ \Lambda_{11,12} &= -\lambda_1 - \lambda_2 \pm \sqrt{(\lambda_1 - \lambda_2)^2 + \lambda_5^2}. \end{aligned} \quad (8)$$

B. Collider constraints

Precision measurements at the LEP and the LHC contributed in pinning down the trace of new physics effects in different forms. The effects of hierarchical heavy BSM scalar mass spectrum and the presence of a lighter DM are under consideration. After the discovery of the Higgs boson, the LHC also measured its properties. Two such measurements, the Higgs decay to $\gamma\gamma$ and its invisible decay are important to consider in the context of IDM.

Oblique correction constraints.—The oblique parameters S , T , and U , proposed by Peskin and Takeuchi [17], are different combinations of the oblique corrections, i.e., radiative corrections to the two-point functions of the SM gauge bosons. The S parameter encodes the running of the neutral gauge boson two-point functions ($\Pi_{ZZ}, \Pi_{Z\gamma}, \Pi_{\gamma\gamma}$) in the lower energy range, from zero momentum to the Z pole. Therefore, the S parameter is sensitive to the presence of light particles with masses below m_Z , which is the case here due to the presence of the light DM. The T parameter, on the other hand, measures the difference between the WW and the ZZ two-point functions, Π_{WW} and Π_{ZZ} , at zero momentum. Mass splitting of the scalars inside a $SU(2)_L$ doublet breaks the custodial symmetry which modifies T . In the IDM, the mass splittings between the neutral and the charged scalars are controlled by the T parameter. The experimentally measured values of oblique parameters that we use in our analysis are [18]

$$S = 0.05 \pm 0.11; \quad T = 0.09 \pm 0.13; \quad U = 0.01 \pm 0.11. \quad (9)$$

$h \rightarrow \gamma\gamma$ signal strength constraint.—The signal strength for the $h \rightarrow \gamma\gamma$ channel is given by the following ratio [19,20],

$$R_{\gamma\gamma} = \frac{\sigma(pp \rightarrow h)}{\sigma(pp \rightarrow h_{\text{SM}})} \times \frac{\text{BR}(h \rightarrow \gamma\gamma)}{\text{BR}(h_{\text{SM}} \rightarrow \gamma\gamma)}. \quad (10)$$

In the IDM, the Higgs production rate is similar to that of the SM, as it is gluon fusion dominated in both the models. So, in the IDM, the ratio can be approximated as

$$R_{\gamma\gamma} = \frac{\text{BR}(h \rightarrow \gamma\gamma)_{\text{IDM}}}{\text{BR}(h \rightarrow \gamma\gamma)_{\text{SM}}}. \quad (11)$$

Combined CMS and ATLAS fit in the diphoton channel provides a 2σ limit on this observable as [21],

$$R_{\gamma\gamma} = 1.14^{+0.38}_{-0.36}. \quad (12)$$

Presence of a charged Higgs in the $h \rightarrow \lambda\lambda$ decay loop can induce a significant shift in this ratio for large values of hH^+H^- coupling. In the IDM, this coupling depends on λ_3 , which is also related to the charged Higgs mass and this parameter is constrained from the allowed range of the ratio $R_{\gamma\gamma}$ that can deviate from unity.²

Constraint from the Higgs invisible decay.—Another constraint from the Higgs data, applicable for the scenario when Higgs can decay to a pair of DM particles with a mass $m_{\text{DM}} < m_h/2$. The invisible decay width is given by

$$\Gamma(h \rightarrow \text{Invisible}) = \frac{\lambda_L^2 v^2}{64\pi m_h} \left(1 - \frac{4m_{\text{DM}}^2}{m_h^2}\right)^{\frac{1}{2}}. \quad (13)$$

The latest ATLAS constraint on the invisible Higgs decay is [22]

$$\text{BR}(h \rightarrow \text{Invisible}) = \frac{\Gamma(h \rightarrow \text{Invisible})}{\Gamma(h \rightarrow \text{Invisible}) + \Gamma(h \rightarrow \text{SM})} < 22\%.$$

In the case for light DM when the Higgs decay to a pair of DM particles is kinematically allowed, this limit can significantly constrain the parameter space of the IDM.

LEP bounds.—A reinterpretation of the neutralino search results at the LEP-II has ruled out the parameter regions [23,24] that satisfy the following three conditions:

$$m_H < 80, m_A < 100, \quad \text{and} \quad m_A - m_H > 8 \text{ GeV}. \quad (14)$$

Reinterpretation of chargino search results at the LEP-II has put a bound [25] on the charged Higgs mass as

$$m_{H^\pm} > 70 \text{ GeV}. \quad (15)$$

A hierarchical IDM scalar spectrum is not restricted from these constraints. Moreover, due to large mass gap in the spectrum, $Z \rightarrow HA, W^\pm \rightarrow HH^\pm, W^\pm \rightarrow AH^\pm$ off-shell decays have a negligible effect on the total width of the W and Z bosons, that are very precisely measured at the LEP experiments.

²In the IDM, only the Higgs decay rate to $\gamma\gamma$ can deviate from the SM value at the leading order. As that deviation is within the experimental limit, the Higgs boson here easily satisfies all Higgs signal data.

C. Astrophysical constraints

This model contains a DM candidate, the CP -even scalar in Φ_2 . Therefore, astrophysical constraints on this model consist of the DM relic density and the direct probe of DM in the Xenon and LUX experiments.

Relic density.—There are unputdownable observational pieces of evidence of the presence of DM in a vast range of length scale, starting from intergalactic rotation curve to the latest Planck experiment data. That suggests the current density of the DM comprises approximately 26% energy budget of the present Universe. The observed abundance of DM is usually represented in terms of density parameter Ω as [26]

$$\Omega_{\text{DM}} h^2 = 0.1187 \pm 0.0017, \quad (16)$$

where the observed Hubble constant is $\mathcal{H}_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$. The rate of DM annihilation to the SM particles is inversely proportional to the relic of the DM, and therefore constraints are imposed to avoid the overproduction of the relic in the IDM. We compute the DM relic density numerically with MicrOmega [27], implementing the IDM details there.

Direct detection constraints.—Along with the constraints from the relic abundance measurement in the Planck experiment, there exist strict bounds on the DM-nucleon cross section from the DM direct detection experiments like Xenon100 (Xenon1T) [28] and more recently from LUX [29]. For scalar DM considered in this work, the spin independent DM-nucleon scattering cross section mediated by the SM Higgs is given as [1]

$$\sigma_{\text{SI}} = \frac{\lambda_L^2 f^2}{4\pi} \frac{\mu^2 m_n^2}{m_h^4 m_{\text{DM}}^2}, \quad (17)$$

where $\mu = m_n m_{\text{DM}} / (m_n + m_{\text{DM}})$ is the DM-nucleon reduced mass and $\lambda_L = (\lambda_3 + \lambda_4 + \lambda_5)$ is the quartic coupling involved in the DM-Higgs interaction. A recent estimate of the Higgs-nucleon coupling is $f = 0.32$ [30], although the full range of allowed values is $f = 0.26\text{--}0.63$ [31]. As shown in Fig. 2 later, the Xenon1T upper bound on the DM-nucleon scattering can put a stringent limit on the allowed λ_L values that constrain the Higgs-DM coupling.

IV. POSSIBLE SEARCHES AND BENCHMARKING

We explore the DM paradigm inside the IDM, discussing the variation of DM relic density with various model parameters. The relic density dependence on DM mass for both the hierarchical and degenerate nature of the BSM mass spectrum is presented in Fig. 1 for different λ_L values. The nature of the mass spectrum is quantified by two mass differences $\Delta M_1 \equiv m_{H^\pm} - m_{\text{DM}}$ and $\Delta M_2 \equiv m_A - m_{\text{DM}}$, which are assumed to be equal (ΔM) in the plots. Effects of the hierarchical mass spectrum with $\Delta M = 100 \text{ GeV}$ and

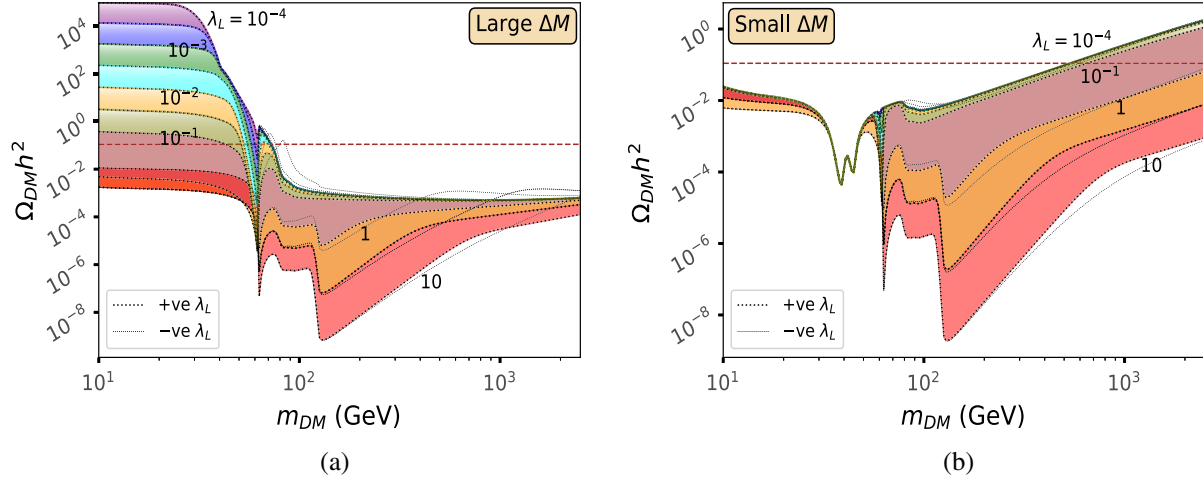


FIG. 1. Variation of relic density $\Omega_{DM}h^2$ shown as a function of dark matter mass m_{DM} in the inert doublet model. Band of colors with thick dotted lines considering different λ_L values. Corresponding negative values of λ_L are shown with thin dotted lines. DM relic abundance strongly depends upon the mass difference ΔM between dark matter candidates from additional BSM scalar masses; left and right plots correspond to the large and small values of it, respectively. One can identify four DM paradigms inside the IDM parameter space, as discussed in Table I.

the degenerate mass spectrum with $\Delta M = 1$ GeV on DM relic density are depicted in Fig. 1(a) and Fig. 1(b), respectively. We also point out how sign reversal of λ_L can alter the relic density dependence on the DM mass.

We roughly divide the DM paradigm of the IDM into four different cases depending on the DM mass and the nature of the mass spectrum, specified by $\Delta M_1, \Delta M_2$. These four cases are showcased in Table I. In each scenario, we discuss the thermal DM relic abundance along with the phenomenological study done to probe the BSM scalars.

Case I.—We first consider a case of light DM with mass, $m_{DM} \lesssim 80$ GeV together with all other heavy scalars within a narrow mass range. This case is severely constrained from the LEP data which rules out $m_{DM} < 45$ GeV. Precise LEP measurements of the Z width constrains $Z \rightarrow AH$ decay together with the conditions in Eq. (14). Even for the DM mass above 45 GeV, the degenerate nature of the spectrum ensures that all the inert scalars take part in the

co-annihilation processes and reduce the relic density to somewhat below 10% of the total relic. Instead of both of the mass differences ΔM_i tiny, if one of them is taken to be large, only the scalar with smaller ΔM dominantly affect the extent of DM co-annihilation. Sharp dip appears when the DM mass is at $m_h/2$ due to the resonant production peak in the DM annihilation through the Higgs portal. Furthermore, some additional shallow dips in the relic density are also observed when the WW and the ZZ annihilation modes open up, enhancing the annihilation cross section. In this low mass region, the DM annihilation is contributed dominantly through the Higgs portal, and thus the sign of λ_L does not affect the relic density. This scenario is explored at the LHC in the mono-jet signal, as discussed in Ref. [10].

Case II.—Here, we consider the light DM with the hierarchical scalar mass spectrum, i.e., large mass differences (ΔM_i) with both of the other heavy scalars. Because of this large mass difference between H and

TABLE I. Illustration of four DM paradigms inside the IDM parameter space, comparing DM mass as well as scalar mass hierarchy. Available DM density as a fraction of the required relic density is also pointed out for these cases. We study the phenomenologically interesting but challenging region marked by case II.

ΔM	M_{DM}	
	Small	Large
Small	Case I $M_{DM} < 80$ GeV $\Delta M \sim \mathcal{O}(1-10)$ GeV Relic density $\sim 10\%$	Case III $M_{DM} \sim 550$ GeV $\Delta M \sim \mathcal{O}(1)$ GeV Relic density $\sim 100\%$
Large	Case II $M_{DM} < 80$ GeV $\Delta M \sim \mathcal{O}(100)$ GeV Relic density $\sim 100\%$	Case IV $M_{DM} \sim 550$ GeV $\Delta M \sim \mathcal{O}(10-100)$ GeV Relic density $\sim 1\%$

A/H^\pm , the LEP Z -width measurements do not constrain such a low DM mass. DM annihilates only through the Higgs portal and therefore for tiny λ_L , the relic is over-produced. However, the entire relic density can be described at larger λ_L values, which are progressively bounded from the DM direct detection data from LUX and Xenon1T. Contrasting this with the degenerate case as pointed out in case I, here the co-annihilation effect is absent in the annihilation cross section and increases the relic density to produce a full relic in the range $m_{\text{DM}} \sim 53\text{--}70$ GeV depending on different λ_L values. Phenomenologically this parameter range is quite interesting although detection of such very light DM along with much heavier other scalars is challenging at the collider. One has to encounter a very small production cross section along with an extremely large SM background where the signal characteristics are very backgroundlike. The LHC potential of this case is studied in the dijet plus MET channel in Ref. [11]. Here, we take up this scenario for further analysis.

Case III.—If we move towards the heavier DM regime, a degenerate mass spectrum can provide full relic density at around $m_{\text{DM}} \sim 550$ GeV.³ Exact mass depends strongly on the value of λ_L parameter. From a 10% relic for $m_{\text{DM}} \sim 100$ GeV, it steadily increases as the $HH \rightarrow WW, ZZ$ annihilations open up and the cross section decreases with mass. The $HHVV$ coupling turns out to be $\lambda_{HHVV_i} \sim (4m_{\text{DM}}\Delta M_i/v^2 + \lambda_L)$ in the limit DM and other heavy scalars are mass degenerate, i.e., $\Delta M_1 \approx \Delta M_2 \rightarrow 0$. We explored this part of the parameter space earlier with both $\Delta M_i = 1$ GeV. Even if the DM annihilation rate increases with the DM mass, that increase is strongly suppressed due to tiny mass differences between the different BSM scalars in a nearly degenerate mass spectrum. The DM relic density, along with being inversely dependent on the annihilation cross section, also is directly proportional to the DM mass. Therefore, the interplay of these two competing effects finally ends up in a gradual increase in the DM relic density. The quartic coupling essentially depends only on λ_L in $\Delta M_i \rightarrow 0$, even then a λ_L sign reversal does not affect the DM pair annihilation. This scenario is phenomenologically interesting but quite challenging to probe at the LHC. This extremely compressed scenario can be probed at the LHC with identifying the charged track signal of a long-lived charged scalar [9].

Case IV.—In the heavier DM regime with hierarchical mass spectrum where both the mass differences are large, e.g., $\Delta M_1 \approx \Delta M_2 \sim 100$ GeV, the annihilation cross section increases with the DM mass. This happens due to rapid increase of the DM-gauge boson quartic couplings with its

mass, i.e., $\lambda_{HHVV_i} = 2(2m_{\text{DM}} + \Delta M_i)\Delta M_i/v^2 + \lambda_L$, for any general ΔM_i , which is a result of the large mass difference between the BSM scalars. Enhancement in the DM annihilation leads to drop of the relic density with increasing m_{DM} producing up to a few percents of the full observed value. Here, λ_L dependence is mostly overshadowed by large mass differences and does not affect much. In the case of very distinct choices of ΔM_i values, the DM annihilation would be dominated by the scalar having a larger mass difference through this enhanced coupling while the other one would contribute negligibly. Therefore the DM scenario in case III can be envisaged as a limiting case of case IV.

Among the four DM scenarios in the IDM as described above and also summarized in Table I, two cases have emerged as phenomenologically exciting. Light DM ($m_{\text{DM}} \sim 50\text{--}80$ GeV) with hierarchical mass spectrum with a substantial mass gap ($\Delta M_i \gtrsim 100$ GeV) with other heavy scalars can provide the full observed DM relic density. On the other side, we get a rather heavy DM ($m_{\text{DM}} \sim 550$ GeV) with an extremely degenerate mass spectrum, which can also provide the required relic density. Both the scenarios are challenging to probe, as the heavier BSM scalars are difficult to produce in the inert model and essentially confront with large irreducible SM backgrounds.

Now, we particularly focus on the low DM mass ($50\text{--}70$ GeV) with hierarchical mass spectrum, i.e., $\Delta M_1, \Delta M_2 \gtrsim 200$ GeV for our phenomenological study. To demonstrate the exact numerical evaluation, in Fig. 2, we explore the $m_{\text{DM}} - \lambda_L$ parameter plane of the IDM for a light DM with $\Delta M_i = 100$ GeV applying the constraints from the DM relic density measurements, the DM direct detection experiments and the constraint from the Higgs

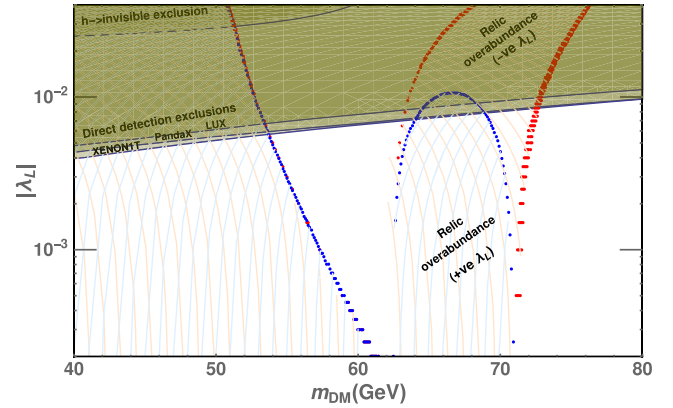


FIG. 2. Allowed DM relic abundance in $m_{\text{DM}} - |\lambda_L|$ parameter space extracted for a set of M_H^\pm, M_A , and λ_2 values. Blue (red) dots are the points where observed abundance is exactly satisfied as in Eq. (16) for +ve (−ve) values of λ_L . The shaded area under this curve represents over abundance and is thus excluded. Two other shaded regions at the upper portion of the plot are excluded from invisible decay of Higgs and DM direct detection constraints from XENON1T, Panda, and LUX, respectively.

³Recently, this limit is brought down to the DM mass ~ 400 GeV, as shown in Ref. [32], by introducing right handed neutrinos, whose late decay to the DM compensates the under-produced DM relic density seen previously.

invisible decay. This choice of 100 GeV is representative since major annihilation modes for DM are through the Higgs portal and the parameter space of this plot is equally valid for larger ΔM choices. Blue (red) dots are the points where the observed DM relic abundance is exactly satisfied as in Eq. (16) for +ve (-ve) values of λ_L . The shaded area under this curve represents DM over abundance and thus is excluded. Two other constraints can come from the invisible decay of the Higgs and the DM direct detection constraints from XENON1T, which are shown in the same plane in two other shaded regions in the upper portion of the plot, respectively. The DM direct detection constraints from LUX (Xenon1T) put stringent upper bound on λ_L , for all values of light DM. All other constraints described above, provide weaker bounds in this parameter space [33].

With our understanding of allowed DM mass and λ_L parameter in the light DM scenario, we now attempt to comprehend other remaining parameters. To do so, we set these parameters to a particular choice from the allowed region of the relic density plot and then perform a scan over the remaining three parameters comprising of heavy scalar masses (M_H^\pm, M_A) and λ_2 . One such frame of the allowed parameter space after imposing the theoretical constraints (unitarity, perturbativity, etc.) along with the $R_{\gamma\gamma}$ constraint from Secs. III A, III B are shown by the blue scatter plots in Fig. 3. The red dots in the same plot represent the values of M_H^\pm and M_A , which satisfy the oblique parameter constraints. These constraints, primarily through the T parameter, force these heavy scalar masses M_H^\pm and M_A to be almost degenerate.

To study the specific low mass DM scenario within the IDM, we choose a set of seven benchmark points (BPs) from the allowed parameter space. These BPs covering heavy scalar mass between 250 to 550 GeV along with the corresponding input DM mass and λ parameter are

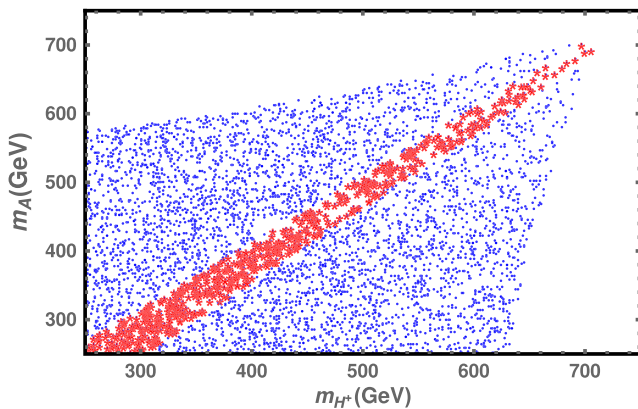


FIG. 3. The blue scatter plot shows the limits from positivity and perturbativity constraints in the (M_H^\pm, M_A) plane after fixing the DM mass and λ_L for all benchmark points. The red dots represent the allowed parameter space after imposing the constraints from the oblique parameters S and T . Similar allowed parameter space is found for other benchmark points.

TABLE II. Input parameters λ and the relevant scalar masses for some of the chosen benchmark points satisfying all the constraints coming from DM, Higgs, theoretical calculations, and low energy experimental data as discussed in the text. All the mass parameters are written in units of GeV. Standard choice of the other two parameters are fixed at $M_H = 53.71$ GeV and $\lambda_L = 5.4 \times 10^{-3}$.

Parameters	BP1	BP2	BP3	BP4	BP5	BP6	BP7
M_{H^\pm} (GeV)	255.3	304.8	350.3	395.8	446.9	503.3	551.8
M_A (GeV)	253.9	302.9	347.4	395.1	442.4	500.7	549.63
λ_2	1.27	1.07	0.135	0.106	3.10	0.693	0.285

summarized in Table II. It is worth noting that the choice of M_H and λ_L is for theoretical and experimental consistency, but the collider analysis proposed in this paper holds equally good for all the allowed points in Fig. 2. Radiative correction to the DM-Higgs portal coupling is calculated in Ref. [34], which allows slightly more parameter space.

V. COLLIDER ANALYSIS

We make use of various publicly available HEP packages for our subsequent collider study aiming for a consistent, reliable detector level analysis. We implement the IDM Lagrangian in FeynRules [35] to create the UFO [36] model files for the event generator MadGraph5 (v2.5.5) [37] which is used to generate all signal and background events. These events are generated at the leading order (LO) and the higher-order corrections are included by multiplying appropriate QCD K factors. We use CTEQ6L1 [38] parton distribution functions for event generation by setting default dynamical renormalization and factorization scales used in MadGraph5 [39]. Events are passed through Pythia8 [40] to perform showering and hadronization and matched up to two to four additional jets for different processes using MLM matching scheme [41,42] with virtuality-ordered Pythia showers to remove the double counting of the matrix element partons with parton showers. The matching parameter, Q_{CUT} is appropriately determined for different processes as discussed in Ref. [43]. Detector effects are simulated using Delphes [44] with the default CMS card. Fatjets are reconstructed using the FastJet [45] package by clustering Delphes tower objects. We employ Cambridge-Aachen (CA) [46] algorithm with radius parameter $R = 0.8$ for jet clustering. Each fatjet is required to have P_T at least 180 GeV. We use the adaptive boosted decision tree (BDT) algorithm in the TMVA framework [47] for MVA.

A. Signal topology

As discussed in the introduction, the hierarchical mass pattern in the IDM scalar sector, (i.e., $M_A \sim M_{H^\pm} \gg M_H$) provides us with interesting final states. Once a pair of

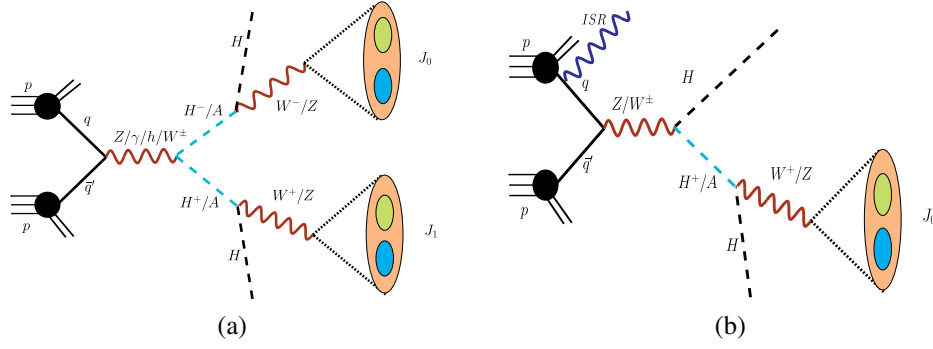


FIG. 4. Representative parton level diagrams of (a) di- V -jet plus missing energy ($2J_V + \cancel{E}_T$) and (b) mono- V -jet plus missing energy ($1J_V + \cancel{E}_T$).

heavy scalars (or one heavy scalar associated with a DM candidate) are produced at the LHC, they eventually decay dominantly producing two (or one) boosted vector bosons, each of which is decaying hadronically and thus producing V jet (J_V), where $V = \{W, Z\}$. These boosted V jets are always associated with large MET (\cancel{E}_T), an outcome of our inability to detect the DM pair at the detector. Representative Feynman diagrams of these signal topologies are demonstrated in Fig. 4. Among them, it must already be clear to the readers that the $1J_V + \cancel{E}_T$ channel alone, although being cross-section-wise bigger than $2J_V + \cancel{E}_T$, has less sensitivity at the LHC due to overwhelmingly large SM background. Therefore, we primarily focus on the $2J_V + \cancel{E}_T$ channel where the large background can be tamed down by employing jet substructure variables in an MVA framework. Our signal is not pure $1J_V + \cancel{E}_T$ or $2J_V + \cancel{E}_T$ topologies, rather it is an admixture of both processes. Note that the baseline selections (defined in Sec. VIB) are designed keeping $2J_V + \cancel{E}_T$ topology in mind. This keeps a large fraction of events from the $1J_V + \cancel{E}_T$ topology. In doing this we gain in the signal, but at the same time one can avoid extremely large background related to the $1J_V + \cancel{E}_T$ topology, (i.e., by demanding at least one J_V instead of two). Before moving on to the actual analysis, we give some useful details about these two signal topologies.

$2J_V + \cancel{E}_T$ channel.—This final state can arise in the IDM for the aforementioned benchmarks from the following three different channels:

$$\begin{aligned} pp &\rightarrow AH^\pm \rightarrow (ZH)(W^\pm H) \equiv 2J_V + \cancel{E}_T, \\ pp &\rightarrow H^+H^- \rightarrow (W^+H)(W^-H) \equiv 2J_V + \cancel{E}_T, \\ pp &\rightarrow AA \rightarrow (ZH)(ZH) \equiv 2J_V + \cancel{E}_T. \end{aligned} \quad (18)$$

Here, A and H^\pm decay to ZH and $W^\pm H$, respectively. As Z and W are originating from a heavy resonance, it is possible that they have sufficient boost to be reconstructed in a large radius jet. We do not distinguish a Z jet or a W jet and call them V jet as we always select fatjets with a broad mass range. A V jet possesses a two prong substructure,

i.e., energy distribution is centered around two subjet axes. We utilize the N -subjettiness ratio τ_{21} (defined later) to tag V jets.

$1J_V + \cancel{E}_T$ channel.—This final state can arise from the following two different channels:

$$\begin{aligned} pp &\rightarrow H^\pm H \rightarrow (W^\pm H)H \equiv 1J_V + \cancel{E}_T, \\ pp &\rightarrow AH \rightarrow (ZH)H \equiv 1J_V + \cancel{E}_T. \end{aligned} \quad (19)$$

Extra jets can arise in the final state due to initial state radiation (ISR) and can form another fatjet. So these channels can potentially mimic the $2J_V + \cancel{E}_T$ final state. We generate matched samples of this signal with up to two additional jets in the final state. In this topology, only one of the two fatjets has the V -jet-like structure and the other jet originates from the QCD radiation which mimics the fatjet characteristics. We find that the contributions to the $2J_V + \cancel{E}_T$ final state from the $1J_V + \cancel{E}_T$ topologies are quite significant and sometimes bigger than the $2J_V + \cancel{E}_T$ contribution itself after our final selection. This is mainly due to bigger production cross sections of $pp \rightarrow AH, H^\pm H$ processes and the tail events which satisfy the fatjet criteria of our analysis.⁴

The leading order production cross sections for the signal processes discussed above for different BPs are given in Table III. We have used NLO QCD k factors of 1.27 and 1.50 for the $q\bar{q}$ and the gg initiated productions for the signal [48].

B. Backgrounds

For our hybrid signal discussed in Sec. I as well as in Sec. VA, major backgrounds come from the following SM processes which we discuss briefly below. All these backgrounds are carefully included in our analysis.

⁴The motivation to choose the $2J_V + \cancel{E}_T$ channel is that one has larger features than in the case of $1J_V + \cancel{E}_T$ to handle the enormous background, where $1J_V + \cancel{E}_T$ also contributes to the signal $2J_V + \cancel{E}_T$ when the extra QCD jet mimics as a fatjet. The $1J_V + \cancel{E}_T$ is explored in the searches in Refs. [12,13].

TABLE III. Production cross sections for the signal processes that contribute to the $1J_V + \cancel{E}_T$ and $2J_V + \cancel{E}_T$ final states at the 14 TeV LHC. These numbers are for the $pp \rightarrow xy$ level before the decay of IDM scalars.

Benchmark points	$\sigma(pp \rightarrow xy)$ (fb)				
	AH^0	$H^\pm H^0$	AH^\pm	$H^+ H^-$	AA
BP1	34.54	62.53	12.62	7.96	0.50
BP2	18.71	34.12	6.22	4.23	0.40
BP3	11.43	20.84	3.50	2.59	0.34
BP4	7.11	13.32	2.05	1.70	0.28
BP5	4.63	8.44	1.22	1.13	0.24
BP6	2.84	5.32	0.71	0.76	0.19
BP7	1.95	3.70	0.45	0.56	0.16

$V + \text{jets}$.—There are the following two types of mono-vector boson processes that contribute dominantly in the background.

- (i) $Z + \text{jets}$: This is the most dominant background in our case. We generate the event samples by simulating the inclusive $pp \rightarrow Z + \text{jets} \rightarrow \nu\nu + \text{jets}$ process matched up to four extra partons. Here, invisible decay of Z gives rise to a large amount of \cancel{E}_T and QCD jets mimic as fatjets.
- (ii) $W + \text{jets}$: This process also contributes significantly in the background when W decays leptonically, and the lepton does not satisfy the selection criteria. This is often known as the lost lepton background. The neutrino comes from the W decay and contributes to missing energy and QCD jets mimic as fatjets. We generate the event samples by simulating inclusive $pp \rightarrow W + \text{jets} \rightarrow \ell_{(e,\mu)}\nu + \text{jets}$ process matched up to four extra partons.

In order to get statistically significant background events coming from the tail phase space region with large \cancel{E}_T , we apply a hard cut of $\cancel{E}_T > 100$ GeV at the generation level to generate these background events.

$VV + \text{jets}$.—Different diboson processes like WZ , WW , and ZZ also mimic the signal and contribute to the SM background. The $pp \rightarrow WZ$ process contributes most significantly among these three diboson channels when W decays hadronically, and Z decays invisibly. We call this background as $W_h Z_\nu$. Similarly, $W_h W_\ell$, where one W decays hadronically and the other leptonically, and $Z_h Z_\ell$ (a hadronic Z and a leptonic Z) can also contribute to the SM backgrounds when leptons remain unidentified. All the diboson processes are generated up to two extra jets with MLM matching. In this case, one of the fatjets can come from the hadronic decay of V , and the other can come from the hard partons.

Single top .—Single top production in the SM includes three types of processes viz. top associated with W (i.e., $pp \rightarrow tW$ process), s -channel single top process (i.e., $pp \rightarrow tb$), and t -channel single top process (i.e., $pp \rightarrow tj$). Among these, the associated production tW contributes

TABLE IV. Cross sections for the background processes considered in this analysis at the 14 TeV LHC. These numbers are shown with the QCD correction order provided in brackets.

Background process		σ (pb)
$V + \text{jets}$ [49,50]	$Z + \text{jets}$	6.33×10^4 [NNLO]
	$W + \text{jets}$	1.95×10^5 [NLO]
$VV + \text{jets}$ [51]	$WW + \text{jets}$	124.31 [NLO]
	$WZ + \text{jets}$	51.82 [NLO]
	$ZZ + \text{jets}$	17.72 [NLO]
Single top [52]	tW	83.1 [N ² LO]
	tb	248.0 [N ² LO]
	tj	12.35 [N ² LO]
Top pair [53]	$t\bar{t} + \text{jets}$	988.57 [N ³ LO]

significantly in the SM background for our signal topologies.

$t\bar{t} + \text{jets}$.—This can be a background for our signal topologies when it decays semileptonically, i.e., one of the top decays leptonically and the other decays hadronically. This background contains b jets. We control this background by applying a b veto. This background always has one V jet. Another fatjet can originate from an untagged b jet or QCD radiation.

Apart from the above background processes, we also calculate the contributions from triboson and QCD multijet processes. However, these contributions are found to be insignificant as compared to the background discussed above, and therefore, we neglect the contribution of these backgrounds in the analysis. The production cross sections with higher order QCD corrections for all the background processes considered in this analysis at the 14 TeV LHC are listed in Table IV.

VI. CUT-BASED ANALYSIS

We perform a CBA to estimate the sensitivity of observing the IDM signatures at the high luminosity LHC runs. It is evident that the signal cross sections are too small compared to the vast SM background. Therefore, one needs sophisticated kinematic observables for the isolation of signal events from the background events. Our signal processes always include at least a hadronically decaying vector boson that can provide a V -like fatjet. Therefore, we make use of the jet substructure variables for our purpose.

A. V -jet tagging: Jet substructure observables

Jet substructure observables have emerged as a powerful technique to search for new physics signatures at colliders. In our case, boosted W and Z bosons, originated from the decay of heavy IDM scalars (H^\pm , A), give rise to collimated jets that can form a large radius jet (fatjet). These fatjets have two-prong substructures. We utilize two jet

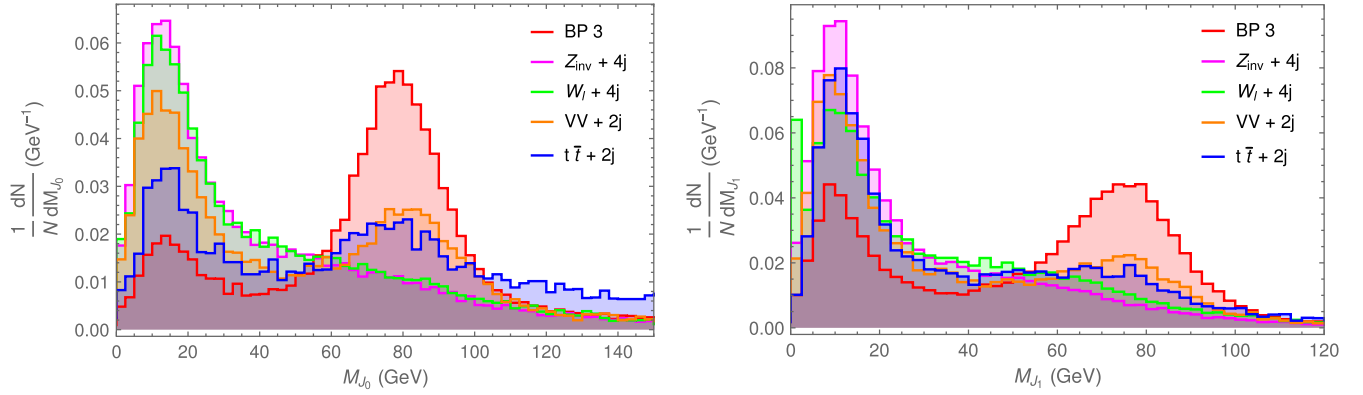


FIG. 5. Normalized distributions for invariant mass of the leading fatjet M_{J_0} (left) and subleading fatjet M_{J_1} (right) after the baseline selection cuts.

substructure observables viz. the jet-mass (M_J) and N -subjettiness ratio (τ_{21}). The M_J is a viable observable to classify the V jets from the fatjets originated from QCD jets. We calculate the jet mass as $M_J = (\sum_{i \in J} P_i)^2$, where P_i are the four-vector of energy hits in the calorimeter. The discrimination power of M_J reduces if extra contribution comes from the parton, which does not actually originate from the V decay. This results in broadening of the peak in the M_J distributions. Different jet grooming techniques are proposed to remove these softer and wide-angle radiations, such as trimming, pruning, and filtering [54–57]. We choose pruning for grooming the fatjets.

Pruned jet mass.—We performed the pruning with the standard method as prescribed in Refs. [55,56] to clean the softer and wide-angle emission by rerunning the algorithm and vetoing on such recombinations. At each step of recombination, one calculates the two variable z and ΔR_{ij} , where z is defined as $z = \min(P_{Ti}, P_{Tj})/P_{T_{i+j}}$ and ΔR_{ij} is the angular separation between two protojets. If $z < z_{\text{cut}}$ and $\Delta R_{ij} > R_{\text{fact}}$ then the i th and j th protojets are not recombined and the softer one is discarded. Here, z_{cut} and R_{fact} are parameters of the pruning algorithm. We have taken the default values of $R_{\text{fact}} = 0.5$ and $z_{\text{cut}} = 0.1$ as

suggested in Ref. [55]. In Fig. 5, we show the distributions for pruned jet mass for signal (BP3) and the important backgrounds. It is evident from these distributions that the peak around 80–90 GeV reflect the V -mass peak for the signal whereas for most of the background processes the peaks below 20 GeV reflect the fatjets mimic from a single prong hard QCD jet.

N -subjettiness ratio.— N subjettiness is a jet variable which determines the inclusive jet shape by assuming N subjets in it. It is defined as the angular separation of constituents of a jet with the nearest subjet axis weighted by the P_T of the constituents and can be calculated as [58,59]

$$\tau_N^{(\beta)} = \frac{1}{\mathcal{N}_0} \sum_i p_{i,T} \min \{ \Delta R_{i1}^\beta, \Delta R_{i2}^\beta, \dots, \Delta R_{iN}^\beta \}. \quad (20)$$

Here, i runs over the constituent particles inside the jet and $p_{i,T}$ is the respective transverse momentum. The normalization factor is defined as $\mathcal{N}_0 = \sum_i p_{i,T} R$ for a jet of radius R . In Fig. 6, the distribution for the N -subjettiness ratio for signal BP3 and leading background are shown. The value for τ_{21} is small for fatjets emerging from the signal than the background. The N -subjettiness ratio τ_{21} is

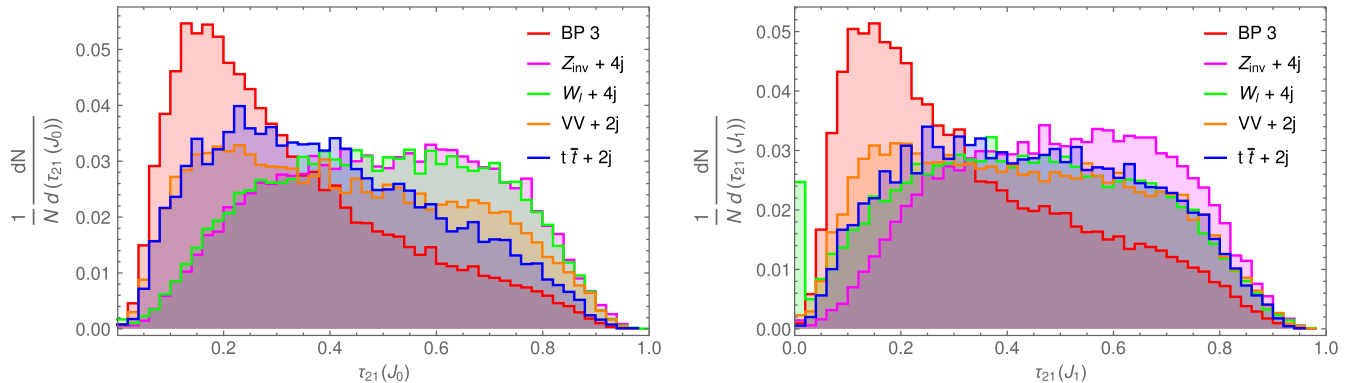


FIG. 6. Normalized distributions for N subjettiness of leading fatjet $\tau_{21}(J_0)$ (left) and subleading fatjet $\tau_{21}(J_1)$ (right) after the baseline selection cuts.

TABLE V. After implementing the corresponding cut, the expected number of events and cut efficiency are shown for signal (BP3) for all possible channels that are contributing to the two $2J_V + \text{MET}$ final state, for an integrated luminosity of 3000 fb^{-1} at the 14 TeV LHC.

Cut	Signal BP3				
	AH^0	$H^\pm H^0$	AH^\pm	$H^+ H^-$	AA
Baseline $+\cancel{E}_T > 200 \text{ GeV}$	672.03 [100%]	1608.8 [100%]	711.62 [100%]	562.15 [100%]	64.5 [100%]
b -veto	474.24 [70.70%]	1291.74 [80.28%]	474.81 [66.66%]	426.85 [75.95%]	32.5 [50.49%]
$65 \text{ GeV} < M(J_0), M(J_1) < 105 \text{ GeV}$	79.50 [11.83%]	274.17 [17.04%]	171.83 [24.12%]	137.56 [24.48%]	4.87 [18.13%]
$\tau_{21}(J_0), \tau_{21}(J_1) < 0.35$	52.44 [7.88%]	171.79 [10.67%]	128.40 [18.02%]	101.98 [18.13%]	3.5 [5.18%]

close to zero if correctly identify the N -prong structure of the jet.

B. Event selection

We list our baseline selection criteria to select events for further analysis.

Baseline selection criteria.—

- (i) Events are selected with missing transverse energy $\cancel{E}_T > 100 \text{ GeV}$.
- (ii) We demand for at least two fatjets of radius parameter $R = 0.8$ constructed using the CA algorithm with fatjet transverse momentum $P_T(J) > 180 \text{ GeV}$.
- (iii) We apply the following lepton veto—events are rejected if they contain a lepton with transverse momentum $P_T(\ell) > 10 \text{ GeV}$ and pseudorapidity $|\eta(\ell)| < 2.4$.
- (iv) We further demand that the azimuthal separation $\Delta\phi$ between the fatjets and \cancel{E}_T , $|\Delta\phi(J, \cancel{E}_T)| > 0.2$. This minimizes the effect of jet mismeasurement contributing to \cancel{E}_T .

After primary selection, we apply the following final selection criteria on events satisfying the baseline selection criteria for final analysis.

Final selection criteria.—

- (i) After optimization with signal and background, the minimum \cancel{E}_T requirement is raised from 100 to 200 GeV.
- (ii) In order to reduce the huge background coming from the $tt + \text{jets}$, we apply a b veto with p_T -dependent b -tagging efficiency as implemented in Delphes. Here, jets are formed using the anti- k_t algorithm with radius parameter $R = 0.5$.
- (iii) We demand that the pruned jet mass of leading and subleading fatjets should be in $65 \text{ GeV} < M_{J_i} < 105 \text{ GeV}$ to tag J_V candidates.
- (iv) Further, to discriminate the fatjet J_V from the QCD jets, we look for the two-prong nature of the fatjet using N subjetteness and select the events with $\tau_{21}(J_i) < 0.35$ of the unpruned fatjet.

In Table V, we present the cut flow for the signal (BP3) associated with the cut efficiencies and the number of events for an integrated luminosity of 3000 fb^{-1} at the

TABLE VI. Cut flow for the SM backgrounds after corresponding cuts are implemented, for an integrated luminosity of 3000 fb^{-1} at the 14 TeV LHC.

Cut	Background				
	$Z_\nu + \text{jets}$	$W_\ell + \text{jets}$	$VV + \text{jets}$	Single-top	$tt + \text{jets}$
Baseline $+\cancel{E}_T > 200 \text{ GeV}$	3.22×10^6 [100%]	4.76×10^6 [100%]	1.47×10^5 [100%]	2.06×10^5 [100%]	3.81×10^5 [100%]
b -veto	2.69×10^6 [83.65%]	4.30×10^5 [9.22%]	1.13×10^5 [75.01%]	3.63×10^4 [16.90%]	4.60×10^4 [12.34%]
$65 \text{ GeV} < M(J_0), M(J_1) < 105 \text{ GeV}$	3.80×10^4 [1.19%]	1.67×10^4 [0.35%]	4.09×10^3 [2.81%]	1.96×10^3 [0.72%]	1.66×10^3 [0.43%]
$\tau_{21}(J_0), \tau_{21}(J_1) < 0.35$	1.30×10^4 [0.41%]	3.79×10^3 [0.07%]	1.62×10^3 [1.03%]	1.44×10^3 [0.56%]	3.84×10^2 [0.10%]

TABLE VII. Statistical significance of the signal for different benchmark points in di-fatjet + \cancel{E}_T analysis for an integrated luminosity of 3000 fb^{-1} at the 14 TeV LHC.

Benchmarks	BP1	BP2	BP3	BP4	BP5	BP6	BP7
Significance	1.9	2.9	3.2	2.9	1.9	1.6	1.1

14 TeV LHC. Similarly, Table VI represents the cut flow for the different backgrounds. From these numbers, it is explicit that the τ_{21} and M_J are powerful variables to have large background reduction with good signal acceptance. We can further infer from Table V that in spite of quite low efficiencies of AH and $H^\pm H$ channels to satisfy the $2J_V + \cancel{E}_T$ criteria, they give comparable contributions to the signal due to its large production cross section.

We compute the statistical signal significance using $\mathcal{S} = \mathcal{N}_S / \sqrt{\mathcal{N}_S + \mathcal{N}_B}$, where \mathcal{N}_S and \mathcal{N}_B represent the remaining number of signal and background events after implementing all the cuts. We show the statistical significance for different benchmark points in Table VII. The highest significance is found for BP3. We would like to emphasize that even after utilizing the novel techniques of jet substructure this particular region of parameter space is very challenging to probe with high sensitivity at the HL-LHC. In order to optimize our search further, we use MVA with jet substructure variables.

VII. MULTIVARIATE ANALYSIS

In the previous section, we present the reach of our model using a CBA. Although we have not achieved discovery significance of 5σ in any of our benchmark points, we see that the two variables viz. M_J and τ_{21} are very powerful to separate the tiny signal from the large SM background. In this section, we use a sophisticated MVA to achieve better sensitivity than a CBA. We would like to discuss two important points here. First, we have observed that MVA does not perform well if we use events selected just with the baseline cuts since the signal is too tiny compared to the overwhelmingly large background. Therefore, we need to apply, in addition to the baseline selection cuts, the following strong cut on the hardest fatjet mass, $M_{J_0} > 40 \text{ GeV}$ and b veto on jets to further trim

down the large background before passing events to MVA. These cuts are very effective to drastically reduce the background but not the signal and are optimally chosen such that it is not too close or too relaxed compared to the cuts used in CBA. If the extra strong cuts for MVA are too close to the cuts applied for the CBA, MVA will not give us an improved sensitivity. On the other hand, if they are too relaxed, the performance of MVA will degrade as the background will become too large. Although we select events with two high- p_T fatjets, we only demand the jet mass of the leading- p_T fatjet is greater than 40 GeV. This will pass a large fraction of mono-fatjet signal events along with the di-fatjet. Therefore, on the one hand, this will increase the signal. However, on the other hand, this will also increase the background.

In Table VIII, we show the number of signal ($1J_V$ and $2J_V$ categories) and background events at the 14 TeV LHC with 3000 fb^{-1} integrated luminosity. Observe that although we demand two fatjets in our selection, the number of $1J_V$ events that contribute to the signal are always bigger than the $2J_V$ contributions for all BPs. This is due to the fact that cross sections for $1J_V$ topologies are much bigger than the $2J_V$ topologies and also a significant fraction of $1J_V$ events pass the selection cuts. Therefore, it is necessary to design hybrid selection cuts, stricter than $1J_V$ but looser than $2J_V$, where both $1J_V$ and $2J_V$ topologies contribute. Our selection cuts are, therefore, optimally designed to achieve better sensitivity.

For our MVA, we use the adaptive BDT algorithm. We obtain two statistically independent event samples for the signal as well as for the background and split the dataset randomly 50% for testing and the rest for training purposes for both the signal and background. Note that there are multiple processes that are contributing to the signal and similarly for the background. In MVA, we construct the signal classes by combining both the $1J_V$ and $2J_V$ topologies that pass our MVA selection criteria. These different signal samples are separately generated at LO and then mixed according to their proper weights to obtain the kinematical distributions for the combined signal. Similarly, all different background samples are mixed to obtain similar distributions for the background class.

The final set of variables that are used in the MVA are decided from a larger set of kinematic variables by looking

TABLE VIII. Number of signal and background events at the 14 TeV LHC with 3000 fb^{-1} integrated luminosity. These numbers are obtained by applying $M_{J_0} > 40 \text{ GeV}$ and b -jet veto in addition to the baseline cuts defined in the text.

Topology	BP1	BP2	BP3	BP4	BP5	BP6	BP7
$1J_V$	1668	2025	2023	1472	1334	1190	920
$2J_V$	601	1112	1572	1254	979	948	608
Z	W	t	tt	WZ	ZZ	WW	Total
3.15×10^6	1.43×10^6	1.6×10^5	1.6×10^5	1.76×10^5	2.97×10^4	1.21×10^4	5.1×10^6

at their power of discrimination between signal and background classes. Four substructure variables for two fatjets, i.e., $M_{J_{0,1}}$ and $\tau_{21}(J_{0,1})$ has already proved to be very important discriminators in our CBA. Stronger transverse momenta cut for such jets are favorable to retain the correct classification of these variables. We already required reasonably high P_T criteria for both such jets. However, to construct the hybrid selection cuts $P_T(J_0)$ can still take a

role in determining the purity of the hardest fatjet J_0 . We also include relative separation between these fatjets $\Delta R(J_0, J_1)$ and the azimuthal angle separation between the leading fatjet from the missing transverse momentum direction $\Delta\phi(J_0, \cancel{E}_T)$. The scale of new physics is relatively high, and that is typically captured by some of the topology independent inclusive variables like H_T , \cancel{H}_T , \cancel{E}_T , etc. We utilize the global inclusive variable $\sqrt{\hat{S}_{\min}}$ proposed to

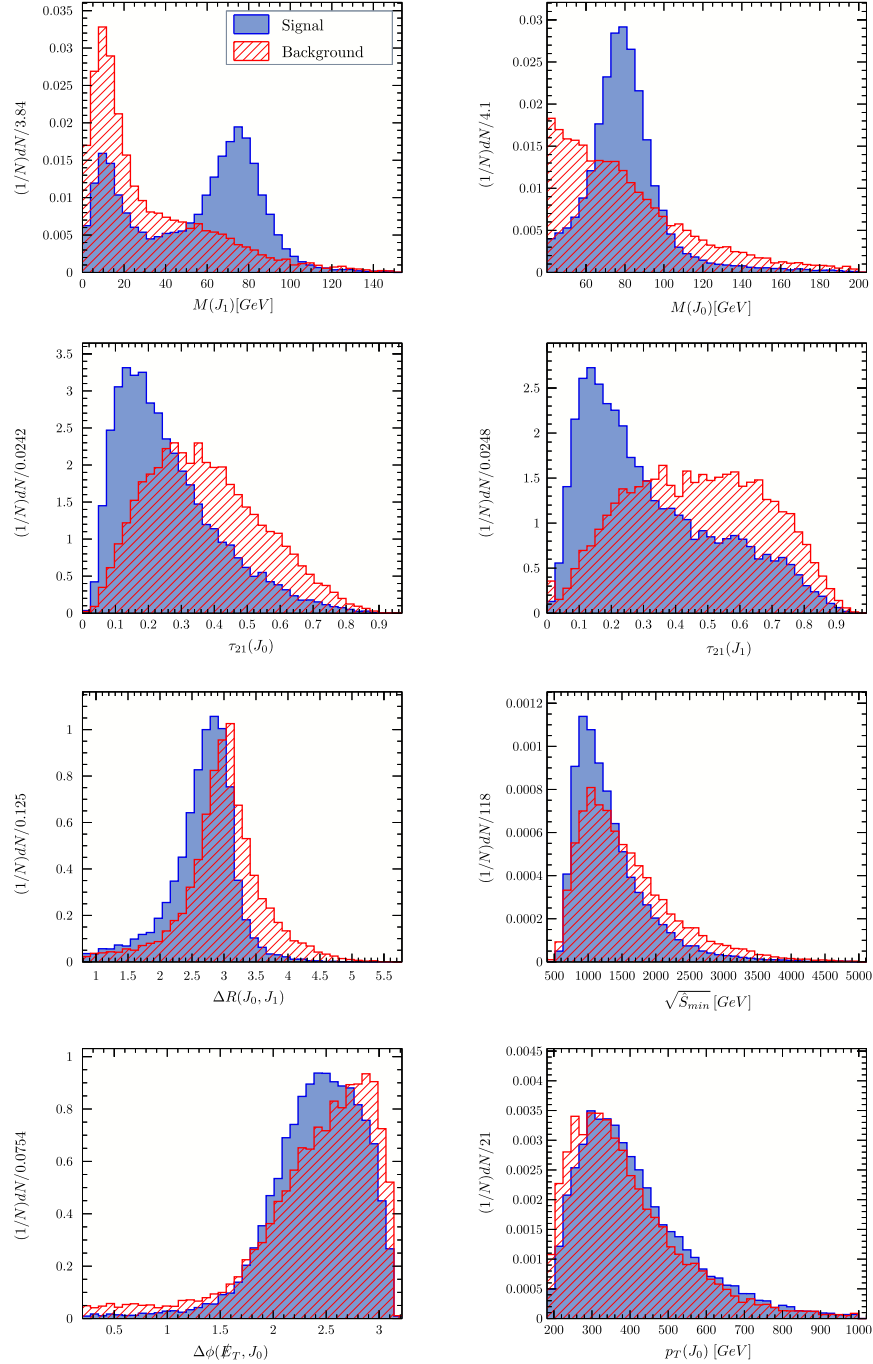


FIG. 7. Normalized distributions of the input variables at the LHC ($\sqrt{s} = 14$ TeV) used in the MVA for the signal (blue) and the background (red). Signal distributions are obtained for BP3 including $1J_V$ and $2J_V$ topologies and the background includes all the dominant backgrounds discussed in Sec. VB.

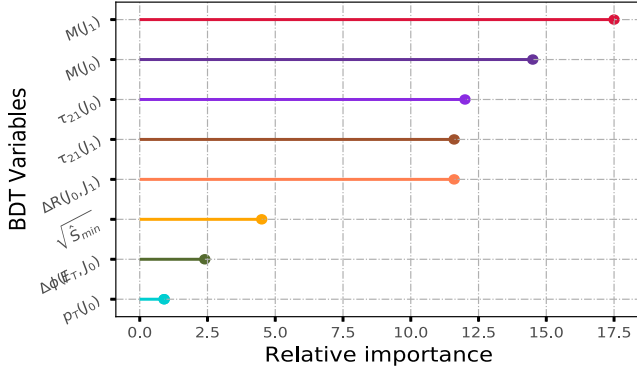


FIG. 8. Kinematic variables used for our MVA and their relative importance. We obtain these numbers from the TMVA package for the benchmark point BP3. Here, we show method unspecific relative importance. This can change slightly for different algorithms and their tuning parameters.

determine the mass scale of new physics for events with invisible particles such as ours [60–62]. This variable, constructed out of all reconstructed objects at the detector, demonstrates better efficiency compared to other inclusive counterparts. For example, we did not use \cancel{E}_T as a feature after baseline cut since it showed a high correlation with $\sqrt{S_{\min}}$ and turned out to be less important than it.

In Fig. 7, we show the normalized distributions of all eight input variables used in the MVA. Signal distributions are obtained for BP3 including $1J_V$ and $2J_V$ topologies and the background includes all the dominant backgrounds discussed in Sec. VB for the 14 TeV LHC. For the same benchmark scenario, method unspecific relative importance of all the kinematic variables are available during TMVA analysis and presented in Fig. 8. Moreover, we mostly keep

variables which are less correlated (or anticorrelated) for both the signal and the background. Relative importance is a measure that is used to rank the variables in MVA. In other words, a variable has better discriminatory power if it has greater relative importance. For this particular benchmark point, BP3, $M_{J_{0,1}}$ variables are very good discriminators according to their relative importance. The N -subjettiness variables, $\tau_{21}(J_{0,1})$, are also very good discriminators as expected. Note that, the relative importance can change for different benchmark points or different LHC energies etc., that can change the shapes of the variables. The linear correlation matrices for the signal and the background can be seen in Fig. 9. Observe that M_{J_1} and $\tau_{21}(J_1)$ variables are strongly anticorrelated. The correlation in the M_{J_1} and $\tau_{21}(J_1)$ variable is due to a mixture of $1J_V$ and $2J_V$ topology in the signal. However, we keep both of them in the MVA since both of them are very powerful discriminators for $2J_V$ topology.

Since the BDT algorithm is prone to overtraining, one should be careful while using it. This usually happens during the training of the algorithm due to inappropriate choices of the BDT specific parameters. One can avoid overtraining by checking the Kolmogorov-Smirnov probability during training. We train the algorithm for every benchmark point separately and ensure that the algorithm is not overtrained in our analysis. Note that the set of eight variables that are used in our analysis may not be the optimal ones. There is always the scope of improving the analysis by choosing a cleverer set of variables. But since the variables we use in MVA are very good discriminators, our obtained sensitivities are fairly robust.

In Fig. 10, we show the normalized BDT response for the signal and the background (training and test samples for both the classes) for BP3. One can clearly see that the BDT

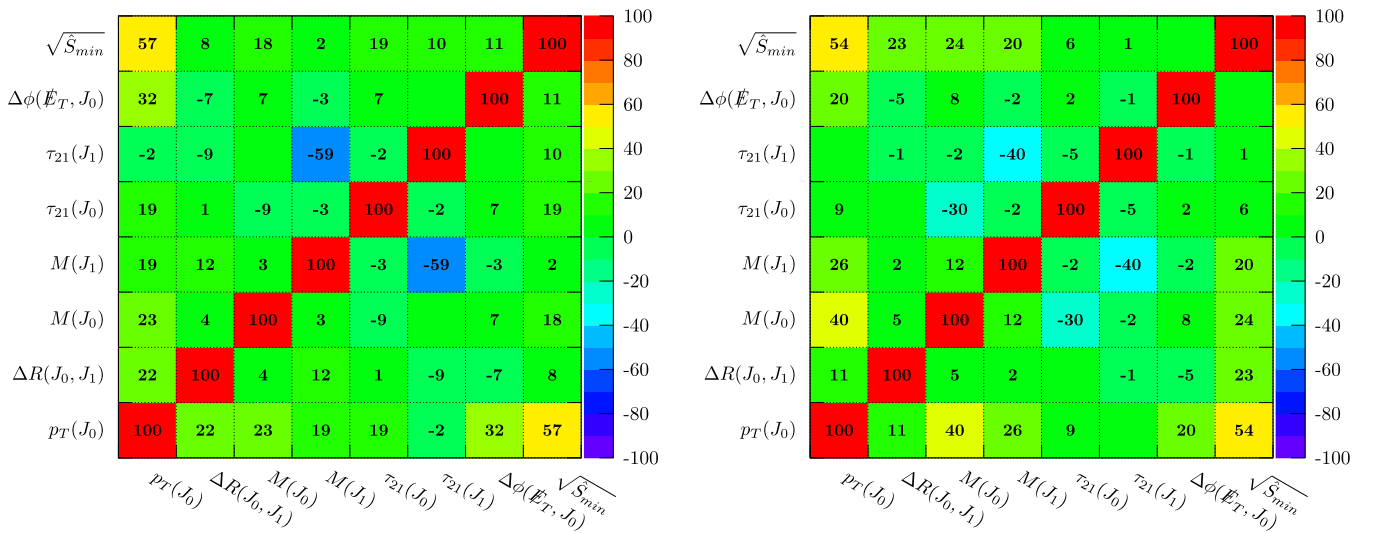


FIG. 9. The linear correlation coefficients (in %) for signal (left panel) and background (right panel) among different kinematical variables that are used for the MVA for BP3. Positive and negative signs of the coefficients signify that the two variables are positively correlated and negatively correlated (anticorrelated).

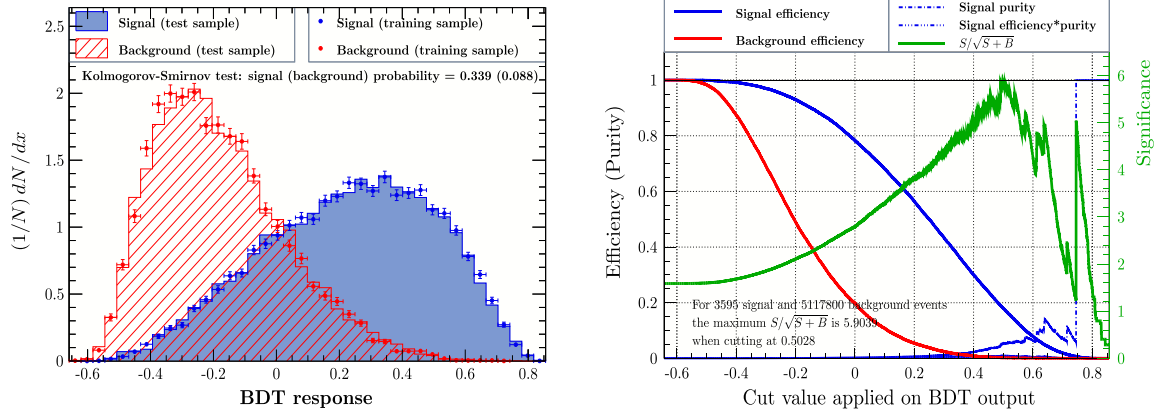


FIG. 10. (Left panel) Normalized BDT response distributions for the signal and the background for BP3. (Right panel) Cut efficiencies as functions of BDT cut values.

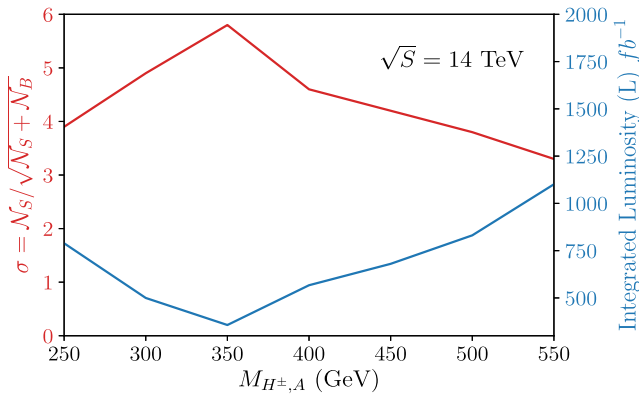


FIG. 11. Significance as a function of heavy scalar mass M_{H^\pm} at the 14 TeV LHC with 3000 fb^{-1} integrated luminosity. We also present required luminosity for the exclusion (2σ) of different benchmark points based on this heavy scalar mass.

responses for the signal and background classes are well separated. We apply a cut on the BDT responses, i.e., $\text{BDT}_{\text{res}} > \text{BDT}_{\text{cut}}$ and show the corresponding cut efficiencies for the signal (blue) and the background (red) and

TABLE IX. Total number of signal events are \mathcal{N}_S^{bc} (including $1J_V$ and $2J_V$ topologies as shown in Table VIII) and with number of background events \mathcal{N}_{SM} before BDT_{opt} cut. The number of signal and background events after the BDT_{opt} cut are denoted by \mathcal{N}_S and \mathcal{N}_B , respectively.

BP	\mathcal{N}_S^{bc}	BDT_{opt}	\mathcal{N}_S	\mathcal{N}_B	$\mathcal{N}_S / \sqrt{\mathcal{N}_S + \mathcal{N}_B}$
1	2269	0.45	412	10748	3.9
2	3137	0.42	596	14200	4.9
3	3595	0.50	635	10957	5.9
4	2726	0.52	504	11514	4.6
5	2313	0.51	404	8880	4.2
6	2138	0.58	385	9871	3.8
7	1528	0.55	278	6823	3.3
\mathcal{N}_{SM}	5117800

the significance (green) as functions of BDT_{cut} . The significance is computed using the formula $\sigma = \mathcal{N}_S / \sqrt{\mathcal{N}_S + \mathcal{N}_B}$, where \mathcal{N}_S and \mathcal{N}_B are the signal and background events that are survived after the $\text{BDT}_{\text{res}} > \text{BDT}_{\text{cut}}$ cut for a given integrated luminosity. The optimal BDT cut, BDT_{opt} is the cut for which the significance is maximized. In Table IX, we show \mathcal{N}_S , \mathcal{N}_B , and σ for different BPs for the 14 TeV LHC, considering an integrated luminosity of 3000 fb^{-1} . We also demonstrate this significance as a function of $M_{H^\pm, A}$ in Fig. 11 (red curve), whereas the blue curve represents the required luminosity for the 2σ exclusion of different BPs.

VIII. RESULTS AND DISCUSSION

The IDM is a simple theoretical framework with rich phenomenology providing possible DM candidates. We classify the model space in four categories depending on the masses of the scalars in the model as summarized in Table I. Some of them are quite interesting in view of the observed properties of the Z boson, Higgs and DM, together with fulfilling all the available theoretical constraints and from the low energy experiments. All such constraints on the IDM are critically analyzed to establish that a hierarchical BSM spectrum with a light DM ($m_{\text{DM}} \lesssim 80 \text{ GeV}$) provides an appealing scenario, as it fulfills the full observed relic density. Furthermore, additional constraints from the Higgs invisible decay and the DM direct detection limits leave us with little allowed parameter space left to be explored at the LHC, albeit a rather difficult region to explore.

Exploiting the fact that after production, the heavy BSM scalar essentially decays into boosted vector bosons together with light DM candidates, we propose a search strategy of a scenario consisting of two boosted fatjets with large MET. Hadronic decay from such boosted vector bosons carries distinctive substructures characteristically different from the single prong large radius QCD jets and can be distinguished with moderate efficiencies using jet substructure observables.

It turns out that our signal of boosted $2J_V + \cancel{E}_T$ also gets significant contributions from single heavy scalar productions with light DM, where the other second J_V is mimicked by a QCD jet, especially since the later production is roughly 1 order of magnitude higher than the two J_V processes. Essentially the di-fatjet signal, after our selection cuts, turns out to be a hybrid of di- J_V and mono- J_V signals. The corresponding background to the mono- J_V channel is also very large, which contributes to the overall background. The $V + \text{jets}$ SM processes are the dominant backgrounds to the above signal, and the sheer magnitudes of these backgrounds of order ~ 1000 pb make it very difficult to search for the BSM scalars of the IDM in any channel. We use intuitive application of jet substructure variables like the fatjet mass (M_J) and the N subjettness (τ_{21}), which encode the internal structure of the fatjets.

Even with these variables, it is extremely difficult to overcome the huge background and therefore, the best case cut-based analysis discovery potential remains restricted to less than 3σ . While cuts on these variables, as detailed in Tables V and VI, can bring down the background to less than the 1% level from the generated ones simultaneously bringing down the signal numbers to 10%–20%. In the end, we do not obtain any significant improvement in the discovery potential to make it cross the desired 5σ barrier for discovery. The best LHC sensitivity is obtained for the BP3 with $m_{H^\pm} \approx m_A \sim 350$ GeV and significance decreases both sides of the spectrum. With the increase of m_{H^\pm}, m_A , we get a higher boost for the decaying vector bosons, resulting in better discrimination power of the jet substructure variables. On the other hand, the presence of

heavier particles leads to the suppressed signal cross section. Therefore, the best signal to background sensitivity is obtained only in an intermediate mass range.

To improve the LHC discovery potential, an MVA is undertaken where we employ a total eight kinematic variables which try to devise a boosted decision tree and provide the optimum separation between signal and background. Instead of the rectangular cuts used in CBA, MVA can use the full potential of jet substructure variables to study the full hierarchical parameter space of the IDM, which is allowed after imposing all the theoretical and experimental constraints. The LHC sensitivity is improved to 5.6σ for BP3 using MVA. Hence, much of the parameter space in a well-motivated scenario within the IDM framework which provides a hierarchical BSM spectrum with light DM ($m_{\text{DM}} \lesssim 80$ GeV), along with an almost degenerate heavy charged Higgs and a pseudoscalar A within the mass range between 250–550 GeV, can be excluded with 1100 fb^{-1} integrated luminosity at the 14 TeV LHC.

ACKNOWLEDGMENTS

The work of A. B. and P. K. is supported by Physical Research Laboratory (PRL), Department of Space, Government of India and the computations were performed using the HPC resources (Vikram-100 HPC) at PRL. T. M. is financially supported by the Royal Society of Arts and Sciences of Uppsala as a researcher at Uppsala University. Part of the work done by S. S. is supported by the DS Kothari postdoctoral fellowship granted by the UGC.

-
- [1] R. Barbieri, L. J. Hall, and V. S. Rychkov, *Phys. Rev. D* **74**, 015007 (2006).
 - [2] M. Cirelli, N. Fornengo, and A. Strumia, *Nucl. Phys.* **B753**, 178 (2006).
 - [3] D. Borah, S. Sadhukhan, and S. Sahoo, *Phys. Lett. B* **771**, 624 (2017).
 - [4] A. Ilnicka, M. Krawczyk, and T. Robens, *Phys. Rev. D* **93**, 055026 (2016).
 - [5] A. Belyaev, G. Cacciapaglia, I. P. Ivanov, F. Rojas-Abatte, and M. Thomas, *Phys. Rev. D* **97**, 035011 (2018).
 - [6] Q.-H. Cao, E. Ma, and G. Rajasekaran, *Phys. Rev. D* **76**, 095011 (2007).
 - [7] E. Dolle, X. Miao, S. Su, and B. Thomas, *Phys. Rev. D* **81**, 035003 (2010).
 - [8] X. Miao, S. Su, and B. Thomas, *Phys. Rev. D* **82**, 035009 (2010).
 - [9] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **01** (2015) 096.
 - [10] A. Belyaev, T. R. F. P. Tomei, P. G. Mercadante, C. S. Moon, S. Moretti, S. F. Novaes, L. Panizzi, F. Rojas, and M. Thomas, *Phys. Rev. D* **99**, 015011 (2019).
 - [11] P. Poulou, S. Sahoo, and K. Sridhar, *Phys. Lett. B* **765**, 300 (2017).
 - [12] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **12** (2016) 083; **08** (2017) 035(E).
 - [13] M. Aaboud *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **10** (2018) 180.
 - [14] J. Chakraborty, P. Konar, and T. Mondal, *Phys. Rev. D* **89**, 095008 (2014).
 - [15] I. F. Ginzburg, K. A. Kanishev, M. Krawczyk, and D. Sokolowska, *Phys. Rev. D* **82**, 123533 (2010).
 - [16] B. Świeżewska, *Phys. Rev. D* **88**, 055027 (2013); **88**, 119903(E) (2013).
 - [17] M. E. Peskin and T. Takeuchi, *Phys. Rev. D* **46**, 381 (1992).
 - [18] M. Tanabashi *et al.* (Particle Data Group), *Phys. Rev. D* **98**, 030001 (2018).

- [19] A. Arhrib, R. Benbrik, and N. Gaur, *Phys. Rev. D* **85**, 095021 (2012).
- [20] B. Swiezewska and M. Krawczyk, *Phys. Rev. D* **88**, 035019 (2013).
- [21] G. Aad *et al.* (ATLAS and CMS Collaborations), *J. High Energy Phys.* **08** (2016) 045.
- [22] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **76**, 6 (2016).
- [23] E. Lundstrom, M. Gustafsson, and J. Edsjo, *Phys. Rev. D* **79**, 035013 (2009).
- [24] G. Belanger, B. Dumont, A. Goudelis, B. Herrmann, S. Kraml, and D. Sengupta, *Phys. Rev. D* **91**, 115011 (2015).
- [25] A. Pierce and J. Thaler, *J. High Energy Phys.* **08** (2007) 026.
- [26] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016).
- [27] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **149**, 103 (2002).
- [28] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. Lett.* **109**, 181301 (2012).
- [29] D. S. Akerib *et al.* (LUX Collaboration), *Phys. Rev. Lett.* **118**, 021303 (2017).
- [30] J. Giedt, A. W. Thomas, and R. D. Young, *Phys. Rev. Lett.* **103**, 201802 (2009).
- [31] Y. Mambrini, *Phys. Rev. D* **84**, 115017 (2011).
- [32] D. Borah and A. Gupta, *Phys. Rev. D* **96**, 115012 (2017).
- [33] M. Krawczyk, D. Sokolowska, P. Swaczyna, and B. Swiezewska, *J. High Energy Phys.* **09** (2013) 055.
- [34] S. Banerjee and N. Chakrabarty, *J. High Energy Phys.* **05** (2019) 150.
- [35] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, *Comput. Phys. Commun.* **185**, 2250 (2014).
- [36] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, *Comput. Phys. Commun.* **183**, 1201 (2012).
- [37] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [38] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *J. High Energy Phys.* **07** (2002) 012.
- [39] <https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/FAQ-General-13>.
- [40] T. Sjostrand, S. Mrenna, and P. Z. Skands, *J. High Energy Phys.* **05** (2006) 026.
- [41] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *J. High Energy Phys.* **01** (2007) 013.
- [42] S. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, A. Schalick, and S. Schumann, in *HERA and the LHC: A Workshop on the Implications of HERA for LHC Physics: Proceedings Part A*, CERN Yellow Reports: Conference Proceedings, edited by A. De Roeck and H. Jung (CERN, Geneva, 2006), <https://doi.org/10.5170/CERN-2005-014>.
- [43] <https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/IntroMatching>.
- [44] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [45] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [46] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *J. High Energy Phys.* **08** (1997) 001.
- [47] A. Hocker *et al.*, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [48] B. Hespel, D. Lopez-Val, and E. Vryonidou, *J. High Energy Phys.* **09** (2014) 124.
- [49] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Phys. Rev. Lett.* **103**, 082001 (2009).
- [50] G. Balossini, G. Montagna, C. M. Carloni Calame, M. Moretti, O. Nicrosini, F. Piccinini, M. Treccani, and A. Vicini, *J. High Energy Phys.* **01** (2010) 013.
- [51] J. M. Campbell, R. K. Ellis, and C. Williams, *J. High Energy Phys.* **07** (2011) 018.
- [52] N. Kidonakis, *Proc. Sci., DIS2015* (**2015**) 170 [[arXiv:1506.04072](https://arxiv.org/abs/1506.04072)].
- [53] C. Muselli, M. Bonvini, S. Forte, S. Marzani, and G. Ridolfi, *J. High Energy Phys.* **08** (2015) 076.
- [54] D. Krohn, J. Thaler, and L.-T. Wang, *J. High Energy Phys.* **02** (2010) 084.
- [55] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev. D* **80**, 051501 (2009).
- [56] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev. D* **81**, 094023 (2010).
- [57] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Phys. Rev. Lett.* **100**, 242001 (2008).
- [58] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **03** (2011) 015.
- [59] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **02** (2012) 093.
- [60] P. Konar, K. Kong, and K. T. Matchev, *J. High Energy Phys.* **03** (2009) 085.
- [61] P. Konar, K. Kong, K. T. Matchev, and M. Park, *J. High Energy Phys.* **06** (2011) 041.
- [62] A. J. Barr, T. J. Khoo, P. Konar, K. Kong, C. G. Lester, K. T. Matchev, and M. Park, *Phys. Rev. D* **84**, 095031 (2011).