Probing the low mass pseudoscalar in the flipped two-Higgs-doublet model

Dilip Kumar Ghosh,[1,*] Biswarup Mukhopadhyaya,[2,†] Sirshendu Samanta,[2,‡] and Ritesh K. Singh[2,§]

[1]*School of Physical Sciences, Indian Association for the Cultivation of Science, 2A & 2B, Raja S.C. Mullick Road, Kolkata 700032, India*

[2]*Department of Physical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, 741246, India*

(Received 28 May 2025; accepted 19 September 2025; published 17 October 2025)

The phenomenology of the flipped two-Higgs-doublet model (2HDM) is relatively less explored so far, as compared to the other, commonly discussed, types. It is found that this scenario, like several others, admits of a light neutral pseudoscalar A in the mass range 20–60 GeV, consistently with all current experimental data and theoretical constraints. However, the fact that such a pseudoscalar decays overwhelmingly into a $b\bar{b}$ pair makes its identification at the Large Hadron Collider (LHC) a challenging task. After identifying the region of the flipped 2HDM parameter space yielding a light pseudoscalar, we identify a useful search channel in the process $pp \to AZ(Z^*) \to b\bar{b}\ell^+\ell^-$. A cut-based analysis, followed by one based on boosted decision trees, shows that the light-A scenario in flipped 2HDM should be detectable with rather high statistical significance at the high-luminosity LHC run, even after including systematic uncertainties. Furthermore, part of the parameter space, especially around $m_A = 30$–40 GeV, is amenable to detection at the discovery level within Run-2 itself.

DOI: 10.1103/nfzx-96x1

I. INTRODUCTION

In the standard model (SM) of weak and electromagnetic interactions, spontaneous breaking of the $SU(2)_L \times U(1)_Y$ local electroweak gauge symmetry into $U(1)_{\text{EM}}$ plays a central role, where *a single complex* $SU(2)$ *scalar doublet* plays a crucial role [1]. However, despite the discovery of a neutral spinless particle with mass around 125 GeV broadly matching the Glashow-Salam-Weinberg Description [2], the question as to whether it is "the Higgs" or "a Higgs" is yet to be settled. Physicists have thus explored scenarios with more than one Higgs doublet (and even larger $SU(2)$ multiplets) [3–8], perhaps encouraged by the observed replication of fermion families. The first step in such theorization is two Higgs doublet models (2HDM) [3,4].

The four major forms of 2HDM, aimed at suppressing flavor changing neutral currents (FCNC) [9], are classified as type I, type II, lepton-specific, and flipped 2HDM [3,4]. They differ from each other in the formulation of the Yukawa interactions, with various kinds of $\mathbb{Z}_2$ couplings to ensure that both of the scalar doublets do not couple to both $T_3 = +1/2$ and $-1/2$ fermion-antifermion pairs. Accelerator phenomenology of these different types has been studied at various levels of detail, excepting the fourth category, namely, flipped 2HDM, on which, to the best of our knowledge, only a limited number of studies have been performed [10–12]. Here we report a study on the flipped 2HDM, with a specific theme.

The theme is woven around a light pseudoscalar physical state, which is present in all four of the aforementioned models. Such a particle cannot be very light in types I and II [13,14], because of its unsuppressed quark couplings. The lepton-specific scenario admits of a relatively light pseudoscalar that dominantly couples to leptons, and has been studied in various contexts [15–20]. This work investigates the flipped 2HDM scenario, which, too, accommodates a light pseudoscalar (around 20 GeV or more), satisfying all the direct search limits from the Large Hadron Collider (LHC). In view of the fact that suggestions about a relatively light spinless particle appear in the literature [21,22], such possibilities are worthy of investigation in all models of extended scalar sectors.

The big challenge in unravelling the signature of a light pseudoscalar (A) in flipped 2HDM is that it decays dominantly in the $b\bar{b}$ channel, which in general tends to get submerged in huge QCD backgrounds. The solution lies in some associated production processes, which provide useful tags. Keeping this in mind, our investigation

*Contact author: tpdkg@iacs.res.in
†Contact author: biswarup@iiserkol.ac.in
‡Contact author: ss21rs027@iiserkol.ac.in
§Contact author: ritesh.singh@iiserkol.ac.in

reveals that it is profitable to use the AZ production channel, mediated by the lighter SM-like scalar h (while the heavier scalar H, too, can mediate this, though its contribution is rather small in regions of the parameter space corresponding to a light A). The Z, via its leptonic decays, offers the much-needed tag.

The twist in the tale here is provided by the fact that even in this channel, a clear distinction of the A-signal requires an invariant mass peak of the $b\bar{b}$-pair at m_A. This requires, among other things, a reasonably high b-jet identification efficiency, which in turn requires each b to have hardness above a certain threshold. However, that tends to suppress the event rates with an on-shell Z, once the acceptance cuts are imposed. In other words, one is pushed to the edge of the allowed phase space of the AZ (Z^*) system. We find that the solution lies in sacrificing the demand for an on-shell Z, demanding instead an invariant mass peak of $b\bar{b}$ pairs with adequate hardness, and looking for the signal

$$pp \to (b\bar{b})_{m_A} + \ell^+\ell^-, \qquad (\ell = e, \mu).$$

We identify the regions corresponding to a light and detectable A (20–50 GeV) in the flipped 2HDM parameter space, and find that a cut-based analysis itself can yield signals with statistical significance (including 10%–20% systematics) range up to 15σ at high-luminosity LHC (HL-LHC) even with an off-shell Z. The significance improves further on carrying out an analysis based on Boosted Decision Trees (BDT) [23]. The results of both kinds of analyses are reported here.

Two of the novel features of our analysis are as follows:

(i) Existing studies in the low-mass region rely on fully hadronic $b\bar{b}$ decays, which suffer from overwhelming QCD backgrounds. In contrast, our analysis leverages the $Z^* \to \ell^+\ell^-$ decay, offering a less background-prone final state and improved signal significance.

(ii) The four final-state particles ($b\bar{b}\ell^+\ell^-$) reconstruct to a mass near 125 GeV, which serves as a natural discriminator against the dominant QCD background. This significantly enhances the signal-to-background ratio, making it an effective search strategy.

We stress upon the fact that the purpose of this investigation is to probe the case of light pseudoscalar and not a generic, exhaustive probe of the flipped 2HDM. Because of this thematic focus, we specifically identify regions where such a light pseudoscalar exists consistently with all phenomenological and theoretical constraints. The Markov Chain Monte Carlo (MCMC) algorithm used to scan the parameter space has thus been subjected to a bias in the low m_A range, in addition to the aforesaid constraints. The relevant parameter regions that emerge are therefore not covering the entire range of parameters in this model. And we have selected some representative benchmark points out of these regions, which bring out the viability of our proposed strategy.

The organization of the paper is as follows. A brief description of the flipped 2HDM and parameter space constraints, discussed in Sec. II, serves as the justification for our chosen benchmark points. Then, in Sec. III, we study the collider characteristics of the signal and backgrounds. We then reported results based on a cut-based approach and subsequently on boosted decision trees (BDT), in Secs. IV and V, respectively. We summarize and conclude in Sec. VI.

II. FLIPPED 2HDM: THE MODEL, ITS PARAMETERS, AND CONSTRAINTS

Being a popular extension of the SM, the 2HDM enriches the scalar sector with an additional complex $SU(2)_L$ doublet. Two scalar doublets are denoted by Φ_1 and Φ_2, and it gets the vacuum expectation value (vev) v_1 and v_2, respectively, after electroweak symmetry breaking (EWSB) [24] with $v = \sqrt{v_1^2 + v_2^2}$, where v is the vev of the SM. Out of eight fields of the two complex scalar doublets, three fields are absorbed to give the mass of three gauge bosons ($Z, W^\pm$) and yield five physical scalar states with two CP-even neutral scalar (h, H), a CP-odd neutral scalar (A) and a pair of charged scalar ($H^\pm$). To avoid FCNC [9] at the tree level, an extra $\mathbb{Z}_2$ symmetry is routinely imposed. This $\mathbb{Z}_2$ is defined in different ways, ensuring that no set of quarks or leptons, with $T_3 = +$ or $-\frac{1}{2}$, couples to both of the scalar doublets.

Neglecting CP-violation, the scalar potential can be written as,

$$\begin{aligned} V_{\text{scalar}} = {}& m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 + \lambda_1 (\Phi_1^\dagger \Phi_1)^2 \\ & + \lambda_2 (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) \\ & + \lambda_4 (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) \\ & + \left\{ -m_{12}^2 \Phi_1^\dagger \Phi_2 + \frac{\lambda_5}{2} (\Phi_1^\dagger \Phi_2)^2 + \text{H.c.} \right\}. \end{aligned} \quad (1)$$

Although the quadratic term $m_{12}^2 \Phi_1^\dagger \Phi_2$ softly breaks the $\mathbb{Z}_2$ symmetry, it does not introduce the FCNC further. The mixing angle $\beta = \tan^{-1}(\frac{v_2}{v_1})$ arises during mass matrix diagonalization of the charged scalar and pseudoscalar, whereas α, the mixing angle for neutral scalar, can be expressed in terms of the model parameters.

Depending on the scalar-doublet fermions couplings in the Yukawa potential, the 2HDM can be categorized into four types (Type-I, Type-II, Type-X, and flipped). In this work, we are dealing with the flipped type 2HDM where, in the Yukawa interactions, down-type quarks couple with Φ_1 and up-type quarks and leptons couple with Φ_2. Thus, in the mass eigen basis, the Yukawa interactions can be written as [3],

$$\mathcal{L}_{\rm Yukawa} = -\sum_f \frac{m_f}{v}(\xi_f^h \bar{f} f h + \xi_f^H \bar{f} f H - i\xi_f^A \bar{f}\gamma_5 f A) - \frac{\sqrt{2}}{v}[V_{ud}^{\rm CKM}(m_u \xi_u^A \bar{u}_R d_L + m_d \xi_d^A \bar{u}_L d_R)H^+ + m_\ell \xi_\ell^A \bar{\nu}_L \ell_R H^+ + {\rm H.c.}], \quad (2)$$

where

$$\xi_u^h = \frac{\cos\alpha}{\sin\beta}, \quad \xi_u^H = \frac{\sin\alpha}{\sin\beta}, \quad \xi_u^A = \cot\beta,$$
$$\xi_d^h = -\frac{\sin\alpha}{\cos\beta}, \quad \xi_d^H = \frac{\cos\alpha}{\cos\beta}, \quad \xi_d^A = \tan\beta,$$
$$\xi_\ell^h = \frac{\cos\alpha}{\sin\beta}, \quad \xi_\ell^H = \frac{\sin\alpha}{\sin\beta}, \quad \xi_\ell^A = -\cot\beta. \quad (3)$$

The model has a total of ten real parameters: $\{m_{11}, m_{22}, m_{12}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, v_1, v_2\}$, of which six of them are independent, when we put $m_h \approx 125$ GeV, and $\sqrt{v_1^2 + v_2^2} = 246$ GeV along with two tadpole equations of electroweak symmetry breaking conditions. For each value of the thus emerging independent parameters, one obtains $m_h, m_H, m_{H^\pm}, m_A$ as the masses of the light and the heavy neutral scalar, charged scalar, and pseudoscalar, respectively. The masses can now be expressed,

$$m_H^2 = M^2 s_{\alpha-\beta}^2 + \left(\lambda_1 c_\alpha^2 c_\beta^2 + \lambda_2 s_\alpha^2 s_\beta^2 + \frac{\lambda_{345}}{2} s_{2\alpha} s_{2\beta}\right) v^2, \quad (4)$$

$$m_h^2 = M^2 c_{\alpha-\beta}^2 + \left(\lambda_1 s_\alpha^2 c_\beta^2 + \lambda_2 c_\alpha^2 s_\beta^2 - \frac{\lambda_{345}}{2} s_{2\alpha} s_{2\beta}\right) v^2, \quad (5)$$

$$m_A^2 = M^2 - \lambda_5 v^2, \quad (6)$$

$$m_{H^\pm}^2 = M^2 - \frac{\lambda_4 + \lambda_5}{2} v^2, \quad (7)$$

where $M^2 = m_{12}^2/(s_\beta c_\beta)$ and, for an angle θ, $s_\theta (c_\theta)$ represents $\sin\theta (\cos\theta)$, and $\lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5$. When it comes to the phenomenology in the alignment limit, $\sin(\beta - \alpha) \approx 1$ [25].

The parameter space of flipped 2HDM, like all other extended EWSB sector scenarios, is constrained by several theoretical issues and experimental results. We outline them below (for details, see Ref. [26]).

(i) Vacuum stability. The demand for a stable minimum of the scalar potential at the EWSB scale requires [27] $\lambda_{1,2} > 0, \lambda_3 > -\sqrt{\lambda_1 \lambda_2}$, and $\lambda_3 + \lambda_4 - |\lambda_5| > -\sqrt{\lambda_1 \lambda_2}$.

(ii) Perturbative unitarity at the electroweak scale. It requires that the coupled-channel gauge and Higgs boson $2 \to 2$ scattering matrix should have its highest eigenvalue bounded above by the optical theorem [28,29] for every partial wave. Also, $|\lambda_i| < 4\pi$ for every $i \in \{1, 2, 3, 4, 5ex\}$.

(iii) The signal strengths of the scalar in all channels should be constrained with LHC data at the 2σ level. This is verified with the help of the HiggsSignals package (which is a part of the HiggsTools) [30,31]. For example, for low m_A, decays like $h \to AZ$ are constrained by HiggsSignals.

(iv) Limits from search of new scalars. The limits coming from searches for the heavier neutral scalars in all channels are obtained from the code HiggsBounds (also part of HiggsTools) [32].

(v) Electroweak precision observables. These provide constraints via the limits described on the oblique electroweak parameters S, T, and U [33–35].

(vi) Flavor observables. Constraints come mainly from $b \to s\gamma$ [36] and $B_s \to \mu^+\mu^-$ [37]. These restrict mainly the $m_H^\pm - \tan\beta$ space.

We scan the parameter space of flipped 2HDM using a MCMC algorithm, with a bias toward $m_A \in$ [15 GeV, 90 GeV], respecting all the constraints above. The MCMC scan finds the following ranges only for the allowed parameters: $\lambda_1 \in [0.1, 1.6]$, $\lambda_2 \in [0.05, 0.16]$, $\lambda_3 \in [11.1, 12.5]$, $\lambda_4 \in [-6.3, -4.3]$, $\lambda_5 \in [5.6, 7.6]$, $m_{12} \in [114, 285]$,

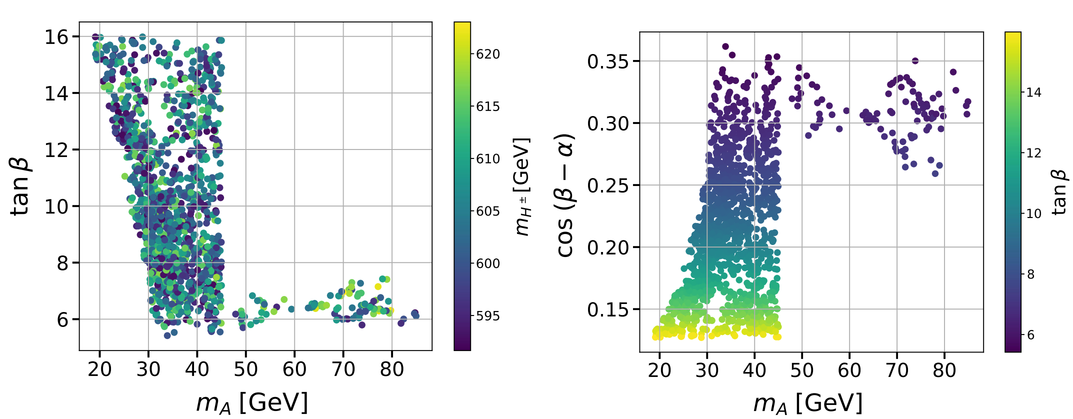

FIG. 1. Allowed scattered points in $m_A - \tan\beta$ plane on the left panel with third axis $m_{H^\pm}$ (color-coded) and in $m_A - \cos(\beta - \alpha)$ plane on the right panel with third axis $\tan\beta$ (color-coded), once we impose all theoretical and experimental constraints discussed in the text.

TABLE I. Chosen parameter points for different benchmarks where the pseudoscalar mass is between 20 and 60 GeV. The bold-faced parameters $\tan\beta$ and m_A are the most important model parameters that decide the signal cross section.

	Benchmark points				
Parameters	BP1	BP2	BP3	BP4	BP5
λ_1	0.418	0.341	0.193	0.349	0.242
λ_2	0.0645	0.0674	0.0676	0.0672	0.0664
λ_3	11.50	12.40	11.70	11.80	12.10
λ_4	−5.79	−5.29	−4.79	−4.76	−5.12
λ_5	5.76	7.11	6.98	7.05	6.99
m_{12} (GeV)	152.0	256.0	266.0	272.0	259.0
$\tan\beta$	**15.0**	**6.4**	**5.8**	**5.6**	**6.2**
m_A (GeV)	**20.0**	**30.0**	**40.0**	**50.0**	**60.0**
$\cos(\beta-\alpha)$	0.14	0.31	0.34	0.34	0.32
$\sin(\beta-\alpha)$	0.99	0.95	0.94	0.94	0.95
$\mathcal{B}(h\to AZ)$	13%	14%	0%	0%	0%
m_H (GeV)	591.0	659.0	653.0	658.0	655.0
$m_H^{\pm}$ (GeV)	592.0	614.0	598.0	600.0	609.0

$\tan\beta \in [5.3, 26.6]$. The left panel of Fig. 1 shows a scatter plot of allowed $\tan\beta$ vs m_A with $m_{H^\pm}$ along the color-coded third axis. On the right panel, the same points are plotted in the $\cos(\beta-\alpha) - m_A$ plane with $\tan\beta$ along the color-coded third axis, where the remaining four parameters have been marginalized over. All the theoretical and experimental constraints discussed above have been used to filter the points that are included in the plot. As is obvious from Fig. 1, $\tan\beta \lesssim 6$–7 are consistent with all constraints for $m_A \gtrsim 45$ GeV. The benchmark points listed in Table I have been guided by this consideration, and they illustrate these regions in the parameter space, which makes an optimistic prediction for our suggested signal rate, as elaborated in the next section.

It may be noticed that λ_3 is rather close to its perturbative limit at the benchmark points used in our analysis. This is essentially due to the following requirements: (a) $m_{H^\pm} \gtrsim 600$ GeV from flavor physics constraints, (b) $|m_{H^\pm} - m_H|$ must be small due to limits from precision electroweak observables, (c) $m_h \approx 125$ GeV, and (d) $m_A \lesssim 50$ GeV in the context of our investigation, where we are looking for *signals of a light pseudoscalar* in flipped 2HDM. How the above constraints work can be understood from Eqs. (4)–(7). The apparent alternatives, which can allow relatively small values of λ_3, are larger magnitudes of m_{12} and/or the quartic couplings $|\lambda_{4,5}|$. However, these options are limited by the four aforementioned requirements. For example, $m_{H^\pm}$ or m_H tend to get lowered below acceptable limits, or m_A ceases to be small compared to $m_{H^\pm}$. Note that relatively high values ($\gtrsim 6$) of $\tan\beta$ have been chosen for $m_A = 20$ and 30 GeV. For higher values of m_A, substantial event rates are estimated only for $\tan\beta \lesssim 6$. The sharp line in Fig. 1 restricting higher $\tan\beta$ for $m_A \gtrsim 45$ is mainly due

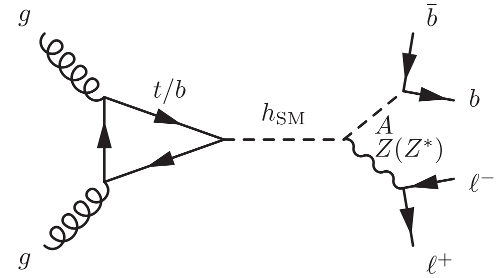

FIG. 2. Feynman diagram for the signal process where the SM Higgs is produced through ggf and decaying further into a Z boson and a pseudoscalar.

to limits from $gg \to A \to b\bar{b}$, which has been looked for recently in CMS [38].

We have already noted why one is pushed to a "corner" of the parameter space where tree-level λ_3 is close to its perturbative limit. This apparently brings in the possibility of the theory becoming nonperturbative around a TeV or so. However, it is nonetheless interesting to explore the implications of a light pseudoscalar in flipped 2HDM, albeit from a purely phenomenological viewpoint, without worrying about the UV completion of the theory. It may always be possible to have additional degrees of freedom around the TeV scale, not yet accessible to the LHC, which temper the growth of the quartic couplings as one reaches the TeV-scale threshold. Therefore, it is worthwhile to study the LHC implications of this scenario, keeping in mind that any experimental evidence of a light pseudoscalar and simultaneous hints of a flipped 2HDM may strongly suggest new degrees of freedom (possibly fermionic) around the TeV scale. Discussions on such possibilities are beyond the scope of the present study.

III. COLLIDER STUDY

A. Parton level characteristics of the signal

As discussed earlier, in the flipped 2HDM, there is no strong limit from the experimental side, forbidding a low mass pseudoscalar with $b\bar{b}$ as the dominant decay channel of the pseudoscalar with $\mathcal{B}(A\to b\bar{b}) \approx 99\%$. However, such a decay channel faces a stiff challenge from backgrounds. Keeping this in mind, we consider the channel where the 125 GeV SM Higgs is produced via gluon fusion (ggf) before decaying into the $AZ(Z^*)$ mode, followed by the leptonic decay (e, μ) of the real or virtual Z boson. Figure 2 represents the contributing Feynman diagram for the signal process. The presence of dileptons in the final states serves as a useful tag that keeps backgrounds under control. Since $(\beta-\alpha) \approx \pi/2$ in the alignment limit, the signal rates depend basically on m_A and $\tan\beta$. The latter controls both the hAZ couplings [$\sim\cos(\beta-\alpha)$] and the branching ratio for $A\to b\bar{b}$ ($\sim\tan\beta$). It should be noted that the production diagram in Fig. 2 includes contributions from both t and b quarks. The H-mediated contribution is

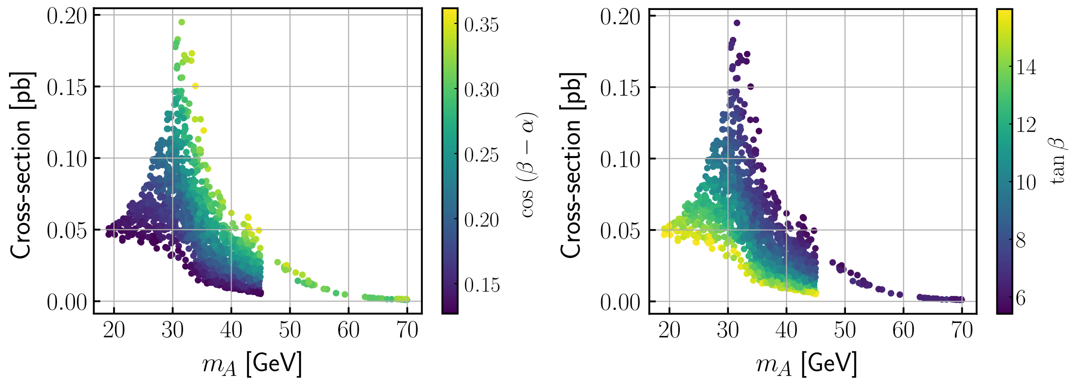

FIG. 3. Left panel of the figure shows the variation of the cross section (pb) for the process, $pp \to AZ(Z^*) \to b\bar{b}\ell^+\ell^-$ through the gluon-gluon fusion with third axis $\cos(\beta-\alpha)$ (color coded). The right panel of the figure represents the cross-section variation with the pseudoscalar mass keeping $\tan\beta$ on the third axis, which indicates the cross-section decreases with increasing $\tan\beta$ for all pseudoscalar masses.

found not to measure up to the h-mediated one in spite of this. On the whole, the contribution of the heavier neutral scalar H turns out to be at most 2% of the total contribution to our stipulated final state rate, although the HZA coupling strength is bigger than that for hZA in the alignment limit. The primary reason for this is the kinematic suppression of the H-mediated channel. Moreover, the $t\bar{t}h$ coupling dominates at the production level, despite the inclusion of the b-quark triangle. Lastly, the H-induced contributions require higher x values, so that suppression of the parton distribution function takes place.

The left panel of Fig. 3 shows the cross section for $pp \to AZ(Z^*) \to b\bar{b}\ell^+\ell^-$ as a function of m_A, with $\cos(\beta-\alpha)$ as the third axis (color coded) and the left panel with $\tan\beta$ as the third axis (color coded). The plot corresponds to outputs from DELPHES [39], through which events generated at the parton level are passed after showering in PYTHIA [40], imposing only the basic, trigger-level cuts. It shows that the rates fall sharply for $\tan\beta$ approaching 10 and beyond. Therefore, such a signal of a light pseudoscalar has hopes of distinction for low $\tan\beta$. This has influenced our choice of benchmark points. For bigger m_A ($\approx$60) GeV, where there is a further kinematic suppression of the rate, the signal is detectable for the smallest possible values of $\tan\beta$ consistent with flavor constraints.

The right panel of Fig. 3 demonstrates the likely detectability of the signal for different values of m_A, with $\tan\beta$ (color coded) along the third axis. $m_A \approx 20$ GeV yields somewhat low rates, because the final state particles often fail to satisfy the trigger. Similarly, kinematic suppression begins for $m_A \gtrsim 40$ GeV. It is around $m_A \approx 30$ GeV that the signal is of the highest strength.[1]

The relevance of some kinematic variables that prompt the choice of the acceptance cuts is indicated in Fig. 4, where the lepton pair invariant mass distribution is shown at the parton level, for three different benchmark points. Each plot is generated using two distinct values of the transverse momentum (p_T^b) trigger threshold for the b quark. We finally require that the total invariant mass of the two leptons and the two b-induced jets should exhibit a peak at m_h. This requires faithful reconstruction of the b jets, for which a minimum p_T^b of 20 GeV is more dependable than the initial parton-level trigger of 10 GeV. However, for final states resulting from an on-shell h, $p_T^b > 20$ GeV affects the hardness of each lepton. One thus notices a marked fall in the event rate for the $m_{\ell\ell}$ on the higher side, including the region corresponding to the Z peak. The events for lower $m_{\ell\ell}$ are therefore preferable if one wants to retain more events with a $\ell\ell b\bar{b}$ peak that is traced us back to the 125 GeV scalar as the source. Therefore, we refrain from demanding an on-shell Z as the source of the $\ell^+\ell^-$ pair. The situation is different for $m_A \geq 30$ GeV, for which there is scope for contributions from an on-shell Z. Here, the high A mass has not only relatively hard b's, but also an on-shell Z with p_T on the higher side, which yields substantial contributions for $m_A = 40$ GeV.

B. Backgrounds

While the signal rate has been computed at the leading order (LO), we have multiplied the LO estimates of the major backgrounds by the next-to-leading (NLO) k factor. This makes our prediction of the signal conservative. The main sources of background we have considered are:

$pp \to b\bar{b}\ell^+\ell^-$ This background arises from the production of a vector boson ($V = Z, \gamma^*$) in association with a $b\bar{b}$ pair. From the production of the di-Z boson, a pair of leptons and a pair of b jets could appear. Similar particles can also appear from an off-shell photon associated with the Z boson. A gluon splitting

[1]We shall see in the next section that, even in the "unfavorable" regions mentioned here, we obtain detectable signals with (a) the choice of appropriate cuts, and (b) the use of BDT.

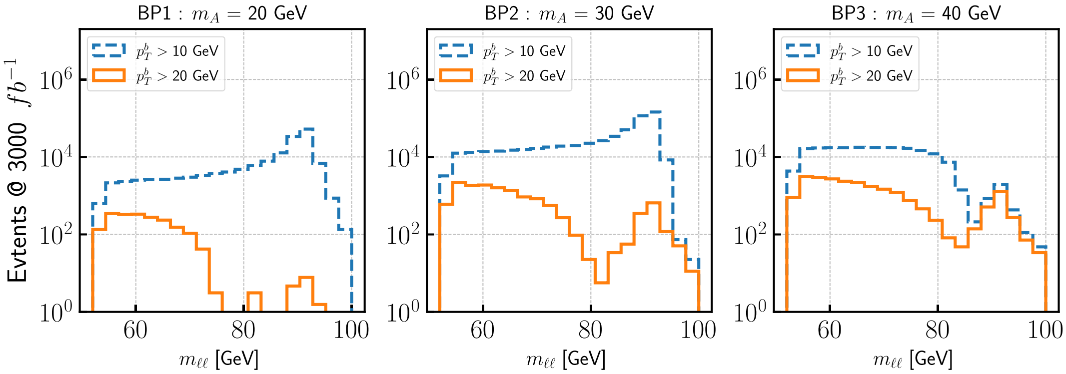

FIG. 4. Histogram of the invariant mass of leptons ($m_{\ell\ell}$) for three different signal benchmark points at the parton level. The y axis represents the number of events at 3000 fb^{-1} luminosity. The blue and orange histograms represent the number of events at the parton level with $p_T^b > 10$ GeV and $p_T^b > 20$ GeV, respectively.

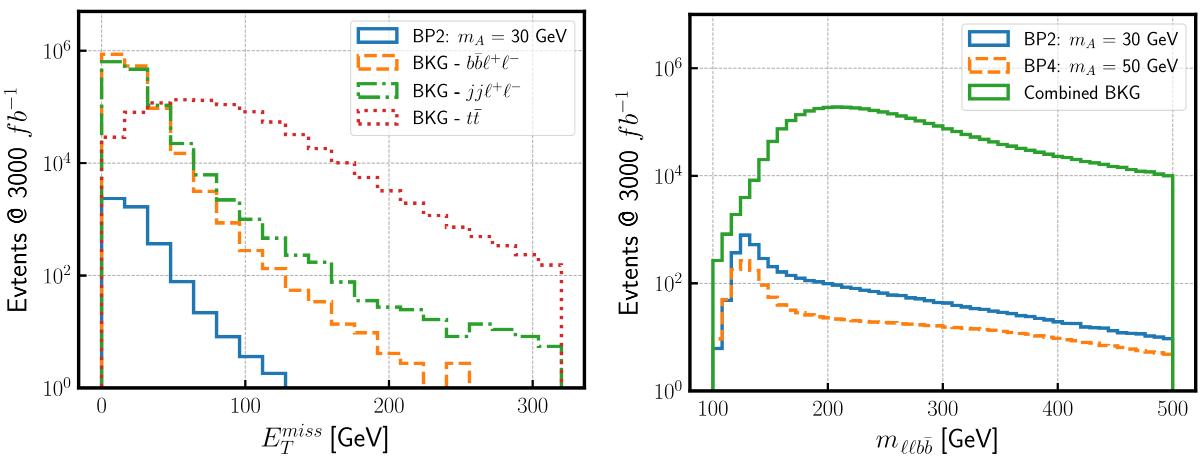

FIG. 5. Left panel shows the distribution of different backgrounds along with BP2, $m_A = 30$ GeV. The signal lies in the lower side of the $E_T^{\rm miss}$; this variable kills the $t\bar{t}$ background to a large extent, whereas the four-particle invariant mass is plotted in the right panel, where the signal peaks at the SM Higgs mass and the background deviates significantly and peaks near 210 GeV.

can also produce a pair of b jets with the association of a vector boson. In a realistic detector, jets and leptons can be misidentified. So, the jet tagging efficiencies have been folded. These efficiencies depend on the jets' transverse momentum (p_T). In contrast, the lepton misidentification rate is negligible. Two b jets are expected in the final state in the signal process, but misidentification can arise from multiple sources. The probability that a true b jet is correctly tagged follows a p_T-dependent relation [41], which results in approximately 60% tagging efficiency for the transverse momenta of b jets in the range of 30–50 GeV, where the signal populates.

$pp \to jj\ell^+\ell^-$ This reducible background involves light-flavor jets ($j = u, d, s, c, g$) produced in association with a vector boson. Though the jets are not b jets, mistagging can cause this background to contaminate the signal region. Among light jets, charm jets are more likely to be mistakenly labeled as b jets. Their mistagging rate is around 15% for $p_T \approx 50$ GeV and follows a p_T-dependent relation [41]. For other light-flavor jets, the mistagging probability is significantly suppressed and approximately linear in p_T [41], yielding a mistagging rate of about 1% for jets with $p_T \approx 50$ GeV. Additionally, the probability of a light jet being misidentified as a lepton is highly suppressed, typically below 1%.

$pp \to t\bar{t}$ Top quark pair production is another significant background due to its high cross section, in which the top quarks decay into bottom quarks and W bosons, which decay into leptons of various flavors. However, semileptonic top quark decays result in missing transverse energy ($E_T^{\rm miss}$) which offers a distinctive feature that suppresses this background. We apply a suitable selection that can significantly reduce this background because the signal process does not involve $E_T^{\rm miss}$. The left panel of Fig. 5 shows the $E_T^{\rm miss}$ distribution for different processes in the background along with one benchmark point (BP2) of our signal.

Moreover, all these backgrounds can be further separated from the signal with an appropriate four-particle invariant mass ($m_{\ell\ell b\bar{b}}$) selection. The right panel of Fig. 5 shows the distribution of $m_{\ell\ell b\bar{b}}$, indicating a clear separation of the peak between signal and backgrounds.

We generate the Universal FeynRules Output (UFO) model file for the event generation using the package SARAH [42], and use the SPheno [43] package, which computes the Higgs masses, decay widths, branching ratios, and various couplings. At a center-of-mass energy of 14 TeV, we simulate all the signals and backgrounds at the parton level using the MadGraph5_AMC@NLO [44]. Further, showering and hadronization are carried out using PYTHIA8 [40] and DELPHES [39] for detector simulation. The anti-k_t technique [45] is used to cluster jets with a radius of 0.5. The NLO k-factors for the backgrounds $b\bar{b}\ell^+\ell^-$, $jj\ell^+\ell^-$, and $t\bar{t}$ are 1.7 [46], 1.3 [47], and 1.489 [48], respectively.

All the signals and backgrounds require triggers while passing through the detector simulator DELPHES. The following triggers in Eq. (8) have been deployed to pass the events in the signals and backgrounds.

$$\text{DELPHES trigger:} \begin{cases} p_T^{\ell} > 10\,\text{GeV}, & |\eta_\ell| < 2.5, \\ p_T^{j} > 20\,\text{GeV}, & |\eta_j| < 4.7, \\ \Delta R(j,j) > 0.4, \\ \Delta R(\ell,\ell) > 0.4, \\ \Delta R(j,\ell) > 0.4. \end{cases} \tag{8}$$

Furthermore, the selection cut requires two opposite-sign same-flavor leptons and two b jets. Table II shows the number of events for an integrated luminosity of 3000 fb^{-1} for all signals, together with the major backgrounds, after the DELPHES trigger and two different levels of the selection cut.

IV. RESULTS WITH CUT-BASED ANALYSIS

The signal channel has advantages in QCD background reduction because of two isolated leptons in the final state.

We perform our numerical analysis assuming an integrated luminosity of $\int \mathcal{L}dt = 3000$ fb^{-1} in the context of the HL-LHC at a center-of-mass energy of 14 TeV. In Table III, we represent the number of signal events corresponding to our BPs that survive after each stage of

TABLE II. Number of signal and different backgrounds events at the integrated luminosity 3000 fb^{-1} passed after DELPHES trigger and two different selection cuts.

	Number of events at $\mathcal{L} = 3000$ fb^{-1} at $\sqrt{s} = 14$ TeV LHC							
	Signals					Backgrounds		
Cuts	BP1	BP2	BP3	BP4	BP5	$b\bar{b}\ell^+\ell^-$	$jj\ell^+\ell^-$	$t\bar{t}$
DELPHES trigger [Eq. (8)]	0.15 M	0.563 M	0.20 M	71.22 K	13.80 K	49.47 M	1450.0 M	13.31 M
Selection cut: $N_{\ell^\pm} = 2$	75.12 K	0.26 M	85.30 K	28.33 K	5.64 K	25.0 M	720.6 M	3.15 M
Selection cut: $N_{b-jets} = 2$	978	4491	3000	1560	690	1.51 M	1.25 M	0.81 M

TABLE III. This is the cut-flow table for the different benchmark points. The $m_{\ell\ell b\bar{b}}$ and E_T^{miss} cuts are applied similarly for each benchmark point, but the invariant mass of two b jets ($m_{b\bar{b}}$) is chosen according to the pseudoscalar mass involved in the different benchmark points, which are shown diagonally in bold. The number of events is specified for an integrated luminosity of 3000 fb^{-1} for the corresponding cuts for signal and background processes.

	Number of events at $\mathcal{L} = 3000$ fb^{-1} at $\sqrt{s} = 14$ TeV LHC							
	Signals					Backgrounds		
Cuts	BP1	BP2	BP3	BP4	BP5	$b\bar{b}\ell^+\ell^-$	$jj\ell^+\ell^-$	$t\bar{t}$
DELPHES trigger + selection cuts	978	4491	3000	1560	690	1.51 M	1.25 M	0.81 M
110 GeV $\leq m_{\ell\ell b\bar{b}} \leq$ 145 GeV	314	1922	1510	716	181	9367	6573	9320
$E_T^{\text{miss}} \leq$ 40 GeV	305	1855	1457	697	177	9265	6408	1033
10 GeV $\leq m_{b\bar{b}} \leq$ 30 GeV	267	583	45	7	1	1358	1910	134
20 GeV $\leq m_{b\bar{b}} \leq$ 60 GeV	256	1834	1419	567	79	6804	5133	666
25 GeV $\leq m_{b\bar{b}} \leq$ 65 GeV	123	1749	1431	646	120	7105	4663	724
35 GeV $\leq m_{b\bar{b}} \leq$ 75 GeV	14	601	1272	680	166	6861	3627	747
45 GeV $\leq m_{b\bar{b}} \leq$ 85 GeV	7	89	527	591	174	5934	2600	652

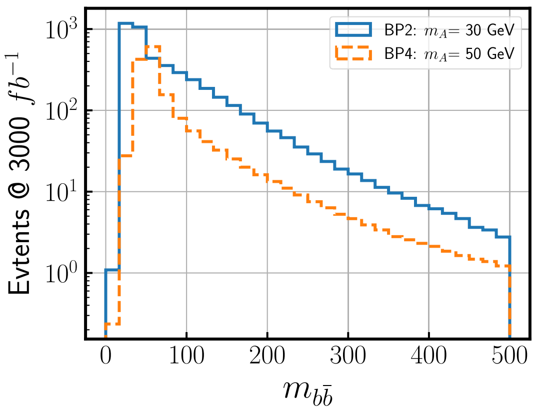

FIG. 6. Figure shows the distribution of the two b jets invariant mass ($m_{b\bar{b}}$) for two different signal benchmark points (BP2, BP5).

cuts. For comparison, we also show the different number of background events in the same table. For distinguishing the signal from the background, one of the most effective kinematic variables is the invariant mass of the two b jets and two leptons, denoted as $m_{\ell\ell b\bar{b}}$. The right side of Fig. 5 shows the distribution of the $m_{\ell\ell b\bar{b}}$ for BP2, $m_A = 30$ GeV, and BP4, $m_A = 50$ GeV, along with all combined backgrounds, where the signal peaks at the SM Higgs mass and the background population near 210 GeV. Although $m_{\ell\ell b\bar{b}}$ extends on the higher side with a long tail in signal, those events mainly come from mistagged b jets or a b jet produced after the parton shower whose mother is not the pseudoscalar. Those events that push the four-particle invariant mass higher are also accumulated in the long tail of the $m_{b\bar{b}}$ as shown in Fig. 6.

Between 110 GeV and 145 GeV, we put the cut on $m_{\ell\ell b\bar{b}}$, and it turns out that the signal events achieve a high survival rate for all benchmark points, but only 0.70% of events survive in the overall backgrounds. After applying this cut, the events in the signal corresponding to the on-shell Z boson peak in the $m_{\ell\ell}$ distribution are significantly suppressed. The remaining events mainly originate from lepton pairs production via an off-shell Z boson as shown in Fig. 7(a). This behavior is consistent with the parton-level characteristics of the signal, as discussed in Sec. III A, where high p_T^b are typically associated with off-shell Z boson. Figure 7(b) represents the $m_{\ell\ell}$ distribution for the backgrounds before and after the $m_{\ell\ell b\bar{b}}$ cut. After imposing this cut on the background, the Z peak also largely shrinks. Then we impose the missing energy cut, namely, $E_T^{\text{miss}} \leq 40$ GeV, to eliminate the $t\bar{t}$ background, mainly as shown before in Fig. 5. After imposing this, the $t\bar{t}$ background shrinks by more than 85%, where signal events are barely suppressed. Lastly, as two b jets in our signal are produced from the pseudoscalar, the invariant mass of these two b jets helps further in improving the signal over backgrounds. We apply different $m_{b\bar{b}}$ cuts for different benchmark points, as shown in the cut-flow Table III.

Various sources of uncertainty, such as luminosity uncertainty, parton distribution function uncertainty, etc., can affect collider experiments. We anticipate some degree of uncertainty in the HL-LHC as well. We, therefore, choose to present the signal significance with systematic uncertainties,

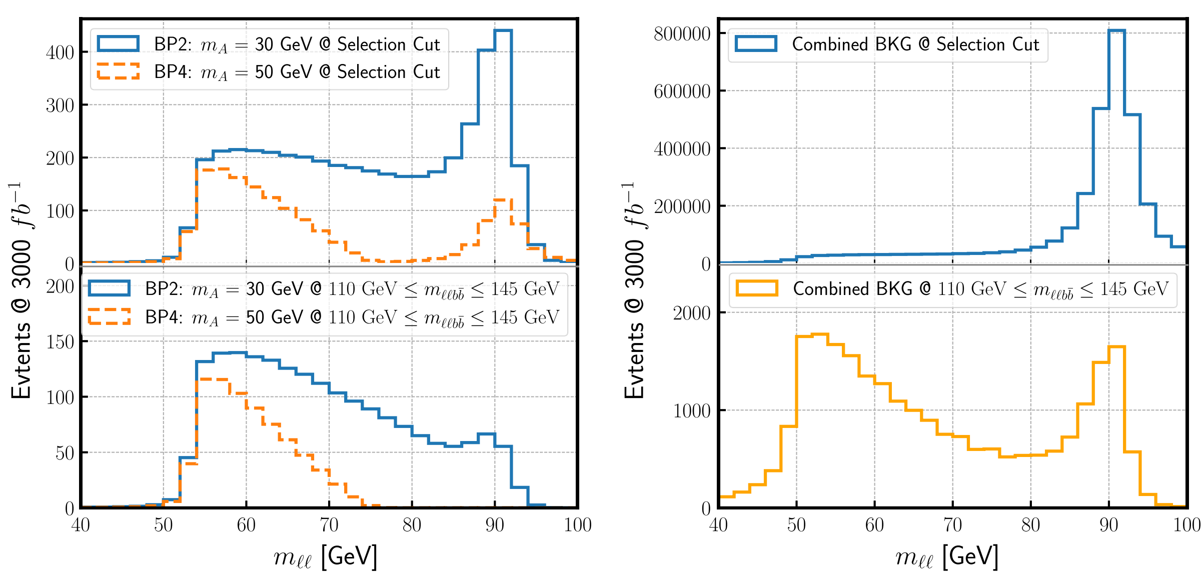

FIG. 7. Two panels of the figure show the distribution of the $m_{\ell\ell}$ for two different signals (BP2, BP5) and background before and after imposing the $m_{\ell\ell b\bar{b}}$ cut. After imposing the $m_{\ell\ell b\bar{b}}$ cut around the SM Higgs mass, mainly the off-shell contribution of the Z boson survives in the signal, which is consistent with the earlier discussion in Sec. III A using parton-level data. Left panel: Distribution of $m_{\ell\ell}$ for two signal benchmark points (BP2 and BP5) before and after imposing the $m_{\ell\ell b\bar{b}}$ cut. Right panel: Corresponding distributions for the backgrounds under the same conditions.

TABLE IV. This is the signal significances at the $\sqrt{s} = 14$ TeV LHC at $\mathcal{L} = 3000$ fb^{-1}, for two different systematics 10% and 20%. For the 10% case, a benchmark-specific $m_{b\bar{b}}$ cut is applied to determine the signal significance, while for the 20% case, all signal significance has been calculated using different $m_{b\bar{b}}$ cuts for all different benchmark points.

	Significance ($\mathcal{S}$) at 10% systematics				
$m_{b\bar{b}}$ cut dependence	BP1	BP2	BP3	BP4	BP5
BP's specific cut	4.3	15.1	11.9	6.0	1.7
	Significance ($\mathcal{S}$) at 20% systematics				
10 GeV $\le m_{b\bar{b}} \le$ 30 GeV	4.1	8.8	0.7	0.1	0.01
20 GeV $\le m_{b\bar{b}} \le$ 60 GeV	2.0	14.5	11.2	4.5	0.6
25 GeV $\le m_{b\bar{b}} \le$ 65 GeV	1.0	13.9	11.4	5.2	0.9
35 GeV $\le m_{b\bar{b}} \le$ 75 GeV	0.1	5.1	10.7	5.6	1.4
45 GeV $\le m_{b\bar{b}} \le$ 85 GeV	0.06	0.8	4.9	5.5	1.6

$$\mathcal{S} = \sqrt{2}\left[(s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right]^{\frac{1}{2}}, \tag{9}$$

where s and b are signal and background event numbers. The systematic uncertainty in the background σ_b^2 is proportional to b and can be given as a percentage of the background. We tabulate the signal significance for all the benchmark points in Table IV, considering 10% systematics, with benchmark-specific $m_{b\bar{b}}$ cuts applied to all the points. Results with 20% systematics uncertainty are also shown there, where, in addition, the effects of the $m_{b\bar{b}}$ cut for each BP on other BPs are also shown. The expected signal significance for all the benchmark points is quite good.

The signal significance is strongest at about 15σ with 10% systematic at 3000 fb^{-1} luminosity for the BP2, where the pseudoscalar mass is 30 GeV. This signal can be seen at 330 fb^{-1} luminosity with 10% systematic at the discovery level 5σ. As the pseudoscalar mass increases, the signal relevance decreases. The signal significance for a 50 GeV pseudoscalar is near 6σ. For the pseudoscalar mass 60 GeV, it drops to 1.7σ. But, as will be covered in the next section, applying the boosted BDT method would improve even more.

V. RESULT USING BDT

In the collider, distinguishing signal events from the backgrounds is the fundamental challenge if the signal rate is sufficiently low. In our work, we see earlier that the simple cut-based analysis is not effective for the BP5, $m_A = 60$ GeV. Machine learning techniques, on the other hand, are powerful tools to enhance signal discrimination by capturing the complex correlation between multiple kinematic variables in the signal and backgrounds over different phase spaces where the cut-based approach mostly fails. In this work, we have used the BDT as a machine learning model. BDT works by iteratively training multiple decision trees on a set of variables and captures the nonlinear correlations.

TABLE V. Hyperparameters used in training the XGBoost classifier. These were selected based on iterative tuning to balance performance and overfitting.

Hyperparameter	Value	Hyperparameter	Value
`objective`	`binary: logistic`	`n_estimators`	1000
`eval_metric`	`logloss`	`learning_rate`	0.01
`max_depth`	3	`min_child_weight`	5
`gamma`	1	`subsample`	0.6
`colsample_bytree`	0.6	`reg_alpha`	1.0
`reg_lambda`	5.0	`tree_method`	`hist`

An important step in the BDT analysis is the selection of the discriminating variables, which could help separate the signal from the background. The most relevant kinematic variables include:

(i) Four-body invariant mass ($m_{\ell\ell b\bar{b}}$): The reconstructed four-body invariant mass in all benchmark signals peaks near the 125-GeV SM Higgs mass. So, in the four-body invariant mass, we allow only those events that lie between 100 GeV and 160 GeV before passing to the BDT. It helps BDT recognize the signal and background pattern better.

(ii) Two-leptons invariant mass ($m_{\ell\ell}$): As in our signal, two leptons produced mainly from an off-shell Z-boson would be a great discriminator for the BDT.

Additionally, we pass the many different features, such as two-body invariant masses m_{ij}, pairwise distances in the $\eta-\phi$ plane ΔR_{ij}, and azimuthal angle separation $\Delta\phi_{ij}$, for $i, j \in \{\ell_1, \ell_2, b_1, b_2\}$. We also add missing energy ($E_T^{\rm miss}$) and the scalar sum of the p_T of all visible particles (H_T). It is trained using a simulated dataset where events are classified as a signal or background. We employ XGBoost [49], an optimized gradient-boosting algorithm, to construct the decision trees. The set of hyperparameters employed for training the XGBoost classifier is summarized in Table V.

The simulated dataset is split into 60% for training, 25% for test purposes, and another 15% for validation during training. During the training of BDT, the discrimination of signals from a background heavily relies on selecting the input variables. So, after training, a feature importance ranking could be extracted from the trained model, which provides insight into which kinematic observables contribute the most to classification. The left panel of Fig. 8 illustrates the learning curve of the BDT, demonstrating that the model is well trained and does not exhibit overfitting. The middle panel of Fig. 8 represents the

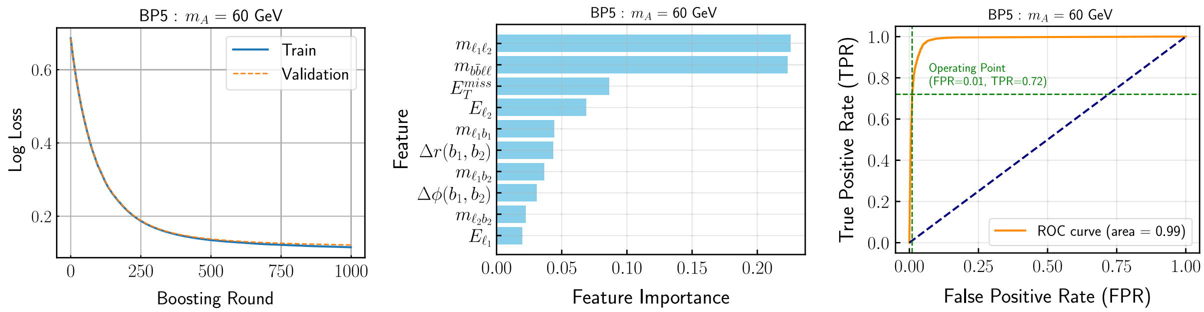

FIG. 8. For BP5, $m_A = 60$ GeV. The left panel shows the trained model learning curve on the training and validation datasets. The importance of the different input features during the training of BDT to discriminate the signal and background is shown in the middle panel. The right panel illustrates the ROC curve after training the BDT, with the AUC of 0.99.

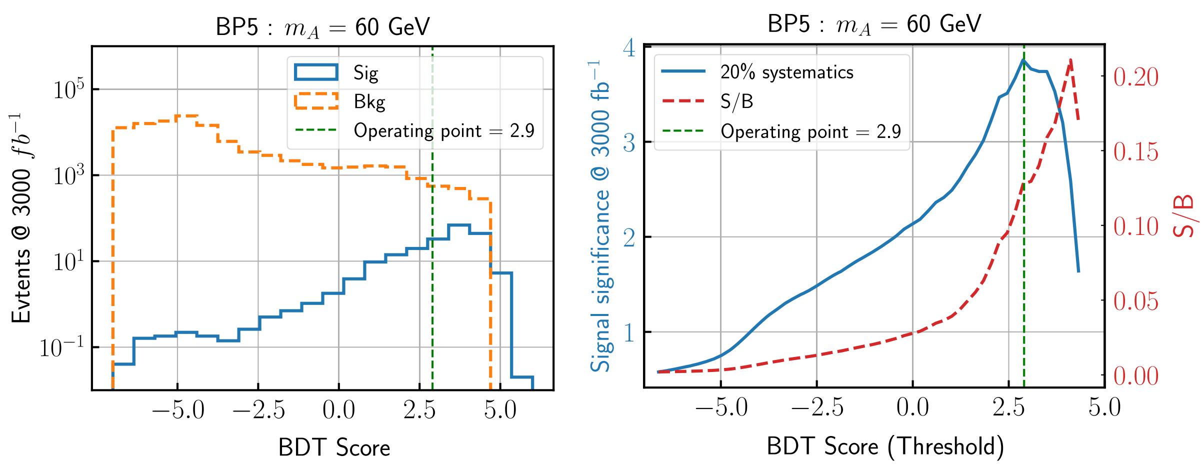

FIG. 9. For BP5: $m_A = 60$ GeV, the right panel represents the distribution of the number of signals and backgrounds at 3000 fb^{-1} luminosity with BDT Score. The left panel shows the signal significance taking 20% systematic uncertainty and the signal over background ratio at 3000 fb^{-1} luminosity with BDT score. The green vertical line in both panels indicates the operating point of BP5, which is set at 2.9.

importance of the variables BDT used to classify the signal and backgrounds. The three most significant variables turn out to be $m_{\ell\ell}$, $m_{b\bar{b}\ell\ell}$, and E_T^{miss}. The right panel of Fig. 8 shows the receiver operating characteristic (ROC) curve, with an area under the curve (AUC) to be 0.99. This indicates the classification power of the trained BDT model is high, i.e., one can have a strong background rejection without losing much of the signal efficiency.

Using the trained BDT model, we evaluate the classification performance of the signal and background events by testing them separately. BDT predicts a score (BDT score) for each event (both signal and background). The left panel of Fig. 9 shows the BDT scores distribution for the signal and backgrounds, indicated with blue and orange lines, respectively. As expected, the signal events predominantly clustered at the higher BDT scores, while the backgrounds accumulated at the lower. Then, we evaluate the signal significance by varying the BDT score as a threshold at an integrated luminosity of 3000 fb^{-1}, considering 20% systematic uncertainties. The right panel of Fig. 9 illustrates the signal significance variation with the BDT score threshold choice for the BP5, $m_A = 60$ GeV indicated by the blue line, while the signal-to-background ratio is plotted in the red line. In the cut-based approach, the signal significance is initially around 1.6σ, taking 20% uncertainty, but this BDT gives a better result with nearly 3.8σ deviation, if we choose the threshold of the BDT score around 2.9. In Table VI, we tabulate signal significance given by BDT models, trained and tested individually for each benchmark point at 20% systematic uncertainty. We then point out the luminosity that is required to reach the discovery level (5σ). For BP2 and BP3, the required luminosity is quite low, 65 and 111 fb^{-1}, respectively, and has already been achieved in the LHC. Therefore, offline analysis of the already existing data can enable one to probe the corresponding regions of flipped 2HDM.

TABLE VI. This table summarizes the signal significance using pretrained BDT models individually for each benchmark. The last row is the scaled luminosity for each benchmark point at the discovery level (5σ).

	Significance ($\mathcal{S}$) at 20% systematics				
	BP1	BP2	BP3	BP4	BP5
Signal significance	6.8	34	26	12	3.8
	Required luminosity in fb^{-1} for 5σ significance				
	1622	65	111	520	5194

VI. SUMMARY AND CONCLUSION

We have explored the detectability of the light neutral pseudoscalar in flipped 2HDM at the LHC, which can lie in the mass range of 20–60 GeV, consistently with all current theoretical and experimental limits. On scanning the regions of the parameter space, which are consistent with these constraints, we identify a set of benchmark points for our analysis. Since the pseudoscalar in such a scenario mostly decays dominantly in the channel $A \to b\bar{b}$, a channel appropriate for LHC search turns out to be $pp \to h \to b\bar{b}\ell^+\ell^-$, with the opposite-sign same-flavor leptons coming from a real or virtual Z boson. Furthermore, the strain on the energy budget of the lepton pair created by the hardness requirement of the b pair, showing an m_A-peak, mostly requires the leptons to come from a virtual Z.

We have performed a cut-based analysis first, identifying suitable event selection criteria that succeed in suppressing the backgrounds rather effectively. Our analysis has demonstrated that $m_A \simeq 20$ GeV has a somewhat poor signal significance in such an analysis, mostly because of the softness of b jets. For $m_A \gtrsim 40$ GeV, on the other hand, kinematic suppression tends to reduce the significance. Nonetheless, one expects signals across the aforementioned range, from the marginal level to the discovery level at the high-luminosity LHC. Our analysis has identified $m_A \approx 30$ GeV as the most optimistic from the signal point of view.

In the second part of our analysis, we have shown that a BDT algorithm improves our results significantly for all benchmark points. We have pointed out that one could reach the discovery level for certain benchmarks at Run II itself. This has motivated us to recommend dedicated searches in this channel using the existing and impending data, while the remaining parts of the light pseudoscalar parameter space in flipped 2HDM retain their promise of discovery when the luminosity reaches the ab^{-1} level.

ACKNOWLEDGMENTS

The authors are thankful to the High Performance Computing (HPC) Cluster (Kepler) facility provided by the Department of Physical Sciences at IISER Kolkata. S. S. acknowledges financial support from the Council of Scientific and Industrial Research (CSIR), India. We thank Rituparna Ghosh and Anirban Kundu for useful discussions. B. M., S. S., and R. S. acknowledge the hospitality of the Indian Association for the Cultivation of Science, where part of the project has been discussed and a portion of this paper was written.

DATA AVAILABILITY

The data that support the findings of this article are openly available [50].

[1] S. F. Novaes, Standard model: An introduction, in *10th Jorge Andre Swieca Summer School: Particle and Fields* (1999), pp. 5–102, arXiv:hep-ph/0001283.
[2] S. M. Bilenky and J. Hosek, Glashow-Weinberg-Salam theory of electroweak interactions and the neutral currents, Phys. Rep. **90**, 73 (1982).
[3] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, Theory and phenomenology of two-Higgs-doublet models, Phys. Rep. **516**, 1 (2012).
[4] G. Bhattacharyya and D. Das, Scalar sector of two-Higgs-doublet models: A minireview, Pramana **87**, 40 (2016).
[5] R. Aggleton, D. Barducci, N.-E. Bomark, S. Moretti, and C. Shepherd-Themistocleous, Review of LHC experimental results on low mass bosons in multi Higgs models, J. High Energy Phys. 02 (2017) 035.
[6] H. Davoudiasl, I. M. Lewis, and M. Sullivan, Good things to do with extra Higgs doublets, in *Snowmass 2021* (2022), arXiv:2203.01396.
[7] I. P. Ivanov, Building and testing models with extended Higgs sectors, Prog. Part. Nucl. Phys. **95**, 160 (2017).
[8] T. Robens, The THDMa revisited, Symmetry **13**, 2341 (2021).
[9] M. Sher, Flavor-changing neutral currents in the Higgs sector, Mod. Phys. Lett. A **37**, 2230011 (2022).
[10] H. E. Logan and D. MacLennan, Charged Higgs phenomenology in the flipped two Higgs doublet model, Phys. Rev. D **81**, 075016 (2010).
[11] M. Hashemi and G. Haghighat, Search for neutral Higgs bosons decaying to $b\bar{b}$ in the flipped 2HDM at future e^-e^+ linear colliders, Phys. Rev. D **100**, 015047 (2019).

[12] L. Dhargyal, $R(D^{(*)})$ and $\mathcal{B}r(B \to \tau\nu_\tau)$ in a flipped/lepton-specific 2HDM with anomalously enhanced charged Higgs coupling to τ/b, Phys. Rev. D **93**, 115009 (2016).

[13] T. Mondal, S. Moretti, S. Munir, and P. Sanyal, Electroweak multi-Higgs production: A smoking gun for the type-I two-Higgs-doublet model, Phys. Rev. Lett. **131**, 231801 (2023).

[14] O. Atkinson, M. Black, A. Lenz, A. Rusov, and J. Wynne, Cornering the two Higgs doublet model type II, J. High Energy Phys. 04 (2022) 172.

[15] E. J. Chun, S. Dwivedi, T. Mondal, B. Mukhopadhyaya, and S. K. Rai, Reconstructing heavy Higgs boson masses in a type X two-Higgs-doublet model with a light pseudoscalar particle, Phys. Rev. D **98**, 075008 (2018).

[16] N. Ghosh and J. Lahiri, Revisiting a generalized two-Higgs-doublet model in light of the muon anomaly and lepton flavor violating decays at the HL-LHC, Phys. Rev. D **103**, 055009 (2021).

[17] A. Dey, J. Lahiri, and B. Mukhopadhyaya, Muon g-2 and a type-X two-Higgs-doublet scenario: Some studies in high-scale validity, Phys. Rev. D **106**, 055023 (2022).

[18] B. Mukhopadhyaya, S. Samanta, T. Samui, and R. K. Singh, Novel signals for the type-X two-Higgs-doublet scenario at the Large Hadron Collider, Phys. Rev. D **108**, 075004 (2023).

[19] B. Mukhopadhyaya, S. Samanta, T. Samui, and R. K. Singh, Improving probes of hAA coupling in the type-X two Higgs doublet model scenario: The crucial role of τ-jet charge identification, Phys. Lett. B **856**, 138949 (2024).

[20] N. Das and N. Ghosh, Unveiling the CP-odd Higgs boson in a generalized 2HDM at a muon collider, Phys. Rev. D **111**, 015035 (2025).

[21] A. Djouadi, P. Zerwas, and J. Zunft, Search for light pseudoscalar Higgs bosons in z decays, Phys. Lett. B **259**, 175 (1991).

[22] S. Chakdar, D. K. Ghosh, P. Q. Hung, N. Khan, and D. Nanda, A model of light pseudoscalar dark matter, Nucl. Phys. **B1008**, 116721 (2024).

[23] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Boosted decision trees, an alternative to artificial neural networks, Nucl. Instrum. Methods Phys. Res., Sect. A **543**, 577 (2005).

[24] A. Pich, Electroweak symmetry breaking and the Higgs boson, Acta Phys. Pol. B **47**, 151 (2016).

[25] J. Bernon, J. F. Gunion, H. E. Haber, Y. Jiang, and S. Kraml, Scrutinizing the alignment limit in two-Higgs-doublet models: $m_h = 125$ GeV, Phys. Rev. D **92**, 075004 (2015).

[26] A. Dey, J. Lahiri, and B. Mukhopadhyaya, LHC signals of a heavy doublet Higgs as dark matter portal: Cut-based approach and improvement with gradient boosting and neural networks, J. High Energy Phys. 09 (2019) 004.

[27] S. Nie and M. Sher, Vacuum stability bounds in the two Higgs doublet model, Phys. Lett. B **449**, 89 (1999).

[28] B. W. Lee, C. Quigg, and H. B. Thacker, Weak interactions at very high energies: The role of the Higgs-boson mass, Phys. Rev. D **16**, 1519 (1977).

[29] D. A. Dicus and V. S. Mathur, Upper bounds on the values of masses in unified gauge theories, Phys. Rev. D **7**, 3111 (1973).

[30] P. Bechtle, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein, and J. Wittbrodt, HiggsSignals-2: Probing new physics with precision Higgs measurements in the LHC 13 TeV era, Eur. Phys. J. C **81**, 145 (2021).

[31] H. Bahl, T. Biekötter, S. Heinemeyer, C. Li, S. Paasch, G. Weiglein, and J. Wittbrodt, HiggsTools: BSM scalar phenomenology with new versions of HiggsBounds and HiggsSignals, Comput. Phys. Commun. **291**, 108803 (2023).

[32] P. Bechtle, D. Dercks, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein, and J. Wittbrodt, HiggsBounds-5: Testing Higgs sectors in the LHC 13 TeV era, Eur. Phys. J. C **80**, 1211 (2020).

[33] M. E. Peskin and T. Takeuchi, Estimation of oblique electroweak corrections, Phys. Rev. D **46**, 381 (1992).

[34] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaborations, Precision electroweak measurements on the Z resonance, Phys. Rep. **427**, 257 (2006).

[35] S. Navas *et al.* (Particle Data Group)Review of particle physics, Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).

[36] HFLAV Collaboration, Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016, Eur. Phys. J. C **77**, 895 (2017).

[37] CMS, LHCb Collaborations, Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data, Nature (London) **522**, 68 (2015).

[38] CMS Collaboration, Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. D **99**, 012005 (2019).

[39] DELPHES 3 Collaboration, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.

[40] C. Bierlich *et al.*, A comprehensive guide to the physics and usage of PYTHIA 8.3, SciPost Phys. Codebases **2022**, 8 (2022).

[41] CMS Collaboration, Identification of b-quark jets with the CMS experiment, J. Instrum. **8**, P04013 (2013).

[42] F. Staub, SARAH 4: A tool for (not only SUSY) model builders, Comput. Phys. Commun. **185**, 1773 (2014).

[43] W. Porod and F. Staub, SPheno 3.1: Extensions including flavour, CP-phases and models beyond the MSSM, Comput. Phys. Commun. **183**, 2458 (2012).

[44] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.

[45] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.

[46] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, and P. Torrielli, W and $Z/\gamma*$ boson production in association with a bottom-antibottom pair, J. High Energy Phys. 09 (2011) 061.

[47] D. Kim, S. Lee, H. Jung, D. Kim, J. Kim, and J. Song, A panoramic study of K-factors for 111 processes at the 14 TeV LHC, J. Korean Phys. Soc. **84**, 914 (2024).

[48] M. Czakon, P. Fiedler, and A. Mitov, Total top-quark pair-production cross section at hadron colliders through $O(\alpha_S^4)$, Phys. Rev. Lett. **110**, 252004 (2013).

[49] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), arXiv:1603.02754.

[50] S. Samanta, D. K. Ghosh, B. Mukhopadhyaya, and R. K. Singh, Data for 'Probing the low mass pseudoscalar in the flipped two-Higgs-doublet model' (Zenodo, 2025), 10.5281/zenodo.17189606.