

# Search for mono-Higgs signals in $b\bar{b}$ final states using deep neural networks

A. Hammad,<sup>1</sup> S. Khalil,<sup>2</sup> and S. Moretti<sup>3,4</sup>

<sup>1</sup>*Institute of Convergence Fundamental Studies, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea*

<sup>2</sup>*Center for Fundamental Physics, Zewail City of Science and Technology, 6 October City, Giza 12588, Egypt*

<sup>3</sup>*School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom*

<sup>4</sup>*Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden*



(Received 16 October 2022; accepted 6 April 2023; published 24 April 2023)

We study mono-Higgs signatures emerging in an illustrative new physics scenario involving Standard Model Higgs boson decays to bottom quark pairs using hybrid deep neural networks. We use a multilayer perceptron to analyze the kinematic observables and optimize the signal-to-background discrimination. The global color flow structure of hard jets emerging from the decay of heavy particles with different color charges is crucial to single out the mono-Higgs signature. Upon embedding the different color flow structures for signal and backgrounds into constructed images, we use a convolution neural network to analyze the latter. Specifically, the approach takes initially mono-type data as input, frittering away invaluable multisource and multiscale information. We then discuss a general architecture of hybrid deep neural networks that supports instead mixed input data. In comparison with single input deep neural networks, like multilayer perceptron or convolution neural networks, the hybrid deep neural networks provide higher capacity in feature extraction and thus in signal versus background classification performance. We provide reference results for the case of the high-luminosity Large Hadron Collider.

DOI: [10.1103/PhysRevD.107.075027](https://doi.org/10.1103/PhysRevD.107.075027)

## I. INTRODUCTION

Mono- $X$  is a signal of a single particle produced alongside missing  $E_T$  (MET), where  $E_T$  represents transverse energy missing (transverse) energy, emerging at hadron colliders, where  $X$  stands generally for a (light) jet,  $W^\pm$ ,  $Z$ , or photon. In recent years, these processes have received a lot of attention as a way to investigate new physics beyond the Standard Model (BSM). In particular, mono- $X$  signatures of the above type would provide an indirect hint for dark matter (DM), which would indeed manifest itself as MET. A common question is which mono- $X$  channel works best. The monojet one has a large cross section relative to the other modes, but the probe (a gluon or a quark) does not typically couple to DM or the mediator. Also, such a channel can suffer from large quantum chromodynamics (QCD) backgrounds. Conversely, the other probes mentioned above could *also* couple to DM and/or the mediator in the presence of a smaller background, but they are very

suppressed as they are electroweak (EW) processes in nature. Altogether, to date, most sensitivity is acquired via monojet processes.

In the above list, we have purposely left aside another possible mono- $X$  probe, where  $X$  is a Higgs boson. This could be the SM-like one discovered at the Large Hadron Collider (LHC) or else a companion state emerging in many BSM scenarios. We are interested in this paper in establishing a mono-Higgs signal, wherein such a state decays into a  $b\bar{b}$  pair. The latter is the most frequent decay channel for a light (neutral) Higgs boson, including the SM-like one discovered in 2012 (hereafter denoted by  $h$ ), so it is of phenomenological relevance. Also this decay, however, requires one to deal with an overwhelming QCD background, so that attempts to extract it in mono-Higgs events have been not very fruitful to date, including in terms of eventually establishing sensitivity to DM and/or the mediator properties, in comparison to what can be established through the other aforementioned mono- $X$  channels. This is clearly a pity, as mono-Higgs has a distinctive advantage with respect to all other mono- $X$  channels, i.e., in proton-proton collisions, for the cases  $X = \text{jet}, W^\pm, Z$ , and  $\gamma$ , the probe can always be emitted directly from a light quark as initial state radiation (ISR) through the usual SM gauge interactions, even when the  $X$

---

*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 SCOAP<sup>3</sup>.*

object interacts with the remainder of the process. In contrast, ISR induced by Higgs-strahlung (i.e., mono-Higgs) is highly suppressed due to the small coupling of the  $h$  state to light quarks. Hence, unlike other mono- $X$  signatures, a mono-Higgs one would probe exclusively the properties of DM and/or the mediator.

Therefore, here, we take a step in the direction of improving the scope of such a signature, i.e., the mono-Higgs one with  $h \rightarrow b\bar{b}$ , by looking at signal events that have a peculiar kinematics with respect to the background. The opportunity is offered by the  $B-L$  supersymmetric Standard Model (BLSSM) of Ref. [1], wherein heavy BSM Higgs states (with masses up to around TeV) can mediate final states involving invisible DM and a Higgs boson [2]. In this case, the final state contains a large amount of MET, allowing the BLSSM signature to be accessed above the SM background. However, in [2], only the  $\gamma\gamma$  and  $ZZ^* \rightarrow 4l$  ( $l = e, \mu$ ) decays of the SM-like Higgs were established, as the  $b\bar{b}$  one was found to be overwhelmed by QCD radiation yielding  $b\bar{b}$  pairs, upon adopting a signal-to-background analysis solely based on kinematic cuts.

It is the purpose of this paper to revisit such an analysis, with the intent of extracting the mono-Higgs signal  $h \rightarrow b\bar{b}$  in the presence of machine learning (ML) techniques. The reason for doing this is twofold. On the one hand, since the appearance of [2], the  $h \rightarrow b\bar{b}$  decay has now been fully established (and found to be compliant with SM predictions [3,4]), so that one can enforce the  $h$  mass reconstruction from the  $b\bar{b}$  system. On the other hand, actual experimental searches based on standard approaches have primarily exploited alternative  $h$  decays, for the aforementioned reasons: e.g., the recent ATLAS [5,6] and CMS [7,8] analyses covered the mono-Higgs channels with  $h$  decays to  $ZZ + \text{MET}$ ,  $W^+W^- + \text{MET}$ ,  $\gamma\gamma + \text{MET}$ , and  $\tau^+\tau^- + \text{MET}$ , all of which have better sensitivity than the  $b\bar{b} + \text{MET}$  channel (and reporting no significant excess over the expected SM background).

Here, we surpass the state of the art in mono-Higgs searches with  $h \rightarrow b\bar{b}$  as we exploit three independent deep neural networks (DNNs). First, we invoke multilayer perceptron (MLP) systems, which analyze the constructed kinematic distributions of the final state particles. Second, we use convolution neural networks (CNNs), which analyze the jet images that can be constructed by embedding the  $p_T$  distribution of the final state jets into visual

representations, wherein the different color flow structure of the signal and background processes is accounted for. Third, we deploy hybrid deep neural networks (HDNNs), which are a two-stream input framework that can analyze the kinematic distributions and constructed jet images at the same time, thereby being able to enhance the signal-to-background classification performance better than the previous approaches.

This paper is organized as follow. In Sec. II we discuss the mono-Higgs signals via  $h \rightarrow b\bar{b}$  arising in the BLSSM, after a very brief discussion of the latter. In Sec. III we discuss our analysis strategy. Section IV shows the described network architectures and their data preprocessing stages. The results of our analysis, exemplified for the high-luminosity LHC (HL-LHC) [9], are given in Sec. V. Finally, we summarize and conclude in Sec. VI.

## II. MONO-HIGGS IN THE BLSSM

In addition to the (s)particles of the minimal supersymmetric standard model (MSSM), the BLSSM includes three chiral right-handed superfields ( $\hat{N}_i$ ), a vector superfield associated with  $U(1)_{B-L}$  ( $\hat{Z}'$ ), and two chiral SM singlet Higgs superfields ( $\hat{\eta}_1, \hat{\eta}_2$ ), as discussed in detail, with the following superpotential:

$$\hat{W} = Y_u \hat{Q} \hat{H}_2 \hat{U}^c + Y_d \hat{Q} \hat{H}_1 \hat{D}^c + Y_e \hat{L} \hat{H}_1 \hat{E}^c + Y_\nu \hat{L} \hat{H}_2 \hat{N}^c + Y_N \hat{N}^c \hat{\eta}_1 \hat{N}^c + \mu \hat{H}_1 \hat{H}_2 + \mu' \hat{\eta}_1 \hat{\eta}_2.$$

The quantum numbers for the particles content in the BLSSM are given in Table I.

After  $B-L$  and EW symmetry breaking, the following mass for the BLSSM-like  $CP$ -odd Higgs  $A'$  is obtained:

$$m_{A'}^2 = \frac{2B\mu'}{\sin 2\beta'} \sim \mathcal{O}(1 \text{ TeV}), \quad (1)$$

where  $\tan \beta' = v'_1/v'_2$ . The BLSSM  $CP$ -even neutral Higgs fields have the following masses at tree level:

$$m_{H', H'}^2 = \frac{1}{2} \left[ (m_{A'}^2 + M_{Z'}^2) \mp \sqrt{(m_{A'}^2 + M_{Z'}^2)^2 - 4m_{A'}^2 M_{Z'}^2 \cos^2 2\beta'} \right]. \quad (2)$$

TABLE I. Particle content of the BLSSM.

	$\hat{Q}_i$	$\hat{U}_i^c$	$\hat{D}_i^c$	$\hat{\ell}_i$	$\hat{E}_i^c$	$\hat{N}_i^c$	$\hat{H}_1$	$\hat{H}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$
$SU(3)_C$	3	$\bar{3}$	$\bar{3}$	1	1	1	1	1	1	1
$SU(2)_L$	2	1	1	2	1	1	2	2	1	1
$U(1)_Y$	1/6	-2/3	1/3	-1/2	1	0	-1/2	1/2	0	0
$U(1)_{B-L}$	1/6	1/6	1/6	-1/2	1/2	1/2	0	0	-1	1

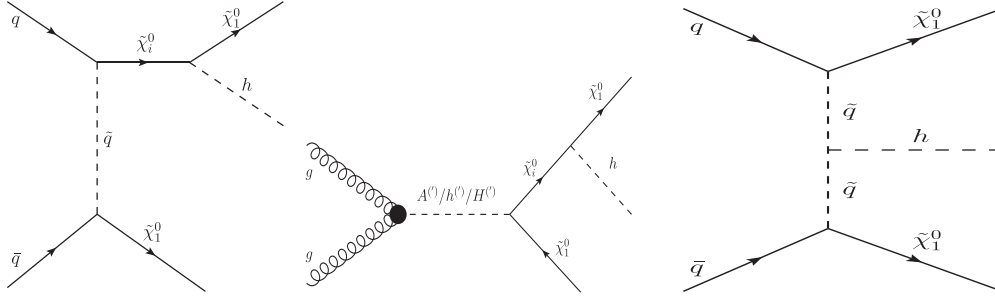


FIG. 1. Mono-Higgs Feynman diagrams  $q\bar{q} \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 h$  with  $\tilde{q}$  exchange (left),  $gg \rightarrow A^{(0)}/h^{(0)}/H^{(0)} \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 h$  (middle), and  $q\bar{q} \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 h$  with  $\tilde{q}\tilde{q}$  exchange (right).

If  $\cos^2 2\beta' \ll 1$ , one finds that the lightest  $B-L$  neutral Higgs mass is given by

$$m_{h'} \simeq \left( \frac{m_{A'}^2 M_{Z'}^2 \cos^2 2\beta'}{m_{A'}^2 + M_{Z'}^2} \right)^{\frac{1}{2}} \simeq \mathcal{O}(100 \text{ GeV}). \quad (3)$$

Another important sector of the BLSSM spectrum is the neutralino one. The neutral gaugino-Higgsino mass matrix can be written as [1]

$$\mathcal{M}_7(\tilde{B}, \tilde{W}^3, \tilde{H}_1^0, \tilde{H}_2^0, \tilde{B}', \tilde{\eta}_1, \tilde{\eta}_2) \equiv \begin{pmatrix} \mathcal{M}_4 & \mathcal{O} \\ \mathcal{O}^T & \mathcal{M}_3 \end{pmatrix},$$

where  $\mathcal{M}_4$  is the MSSM-type neutralino mass matrix [10–13] and  $\mathcal{M}_3$  is a  $3 \times 3$  additional neutralino mass matrix, which is given by

$$\mathcal{M}_3 = \begin{pmatrix} M_{B'} & -g_{B-L} v'_1 & g_{B-L} v'_2 \\ -g_{B-L} v'_1 & 0 & -\mu' \\ g_{B-L} v'_2 & -\mu' & 0 \end{pmatrix}. \quad (4)$$

In addition, the off-diagonal matrix  $\mathcal{O}$  is given by

$$\mathcal{O} = \begin{pmatrix} \frac{1}{2} M_{BB'} & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2} \tilde{g} v_1 & 0 & 0 \\ \frac{1}{2} \tilde{g} v_2 & 0 & 0 \end{pmatrix}. \quad (5)$$

Note that the off-diagonal matrix elements herein vanish identically if  $\tilde{g} = 0$ . In this case, one diagonalizes the real matrix  $\mathcal{M}_7$  with a symmetric mixing matrix  $V$  such that

$$V \mathcal{M}_7 V^T = \text{diag}(m_{\tilde{\chi}_k^0}), \quad k = 1, \dots, 7. \quad (6)$$

In these conditions, the lightest supersymmetric particle (LSP), i.e., the DM candidate of the BLSSM, has the following decomposition:

$$\tilde{\chi}_1^0 = V_{11} \tilde{B} + V_{12} \tilde{W}^3 + V_{13} \tilde{H}_1^0 + V_{14} \tilde{H}_2^0 + V_{15} \tilde{B}' + V_{16} \tilde{\eta}_1 + V_{17} \tilde{\eta}_2. \quad (7)$$

The LSP is called pure  $\tilde{B}'$  if  $V_{15} \sim 1$  and  $V_{1i} \sim 0$  for  $i \neq 5$  and pure  $\tilde{\eta}_{1(2)}$  if  $V_{16(7)} \sim 1$  and all the other coefficients are close to zero.

Now, we can investigate the Mono-Higgs signal of concern in such a BLSSM model. This signal can be produced as initial, intermediate, or final state  $h$  radiation associated with DM pair production (i.e.,  $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ ) [2]. Representative Feynman diagrams for the considered signal processes are shown in Fig. 1. (Notice that we ignore herein  $Z'$  mediated topologies, considered in [2], as they are presently negligible, given the latest limits on  $Z'$  masses and couplings in the BLSSM [14].) The considered mono-Higgs process benefits then from three subprocesses, two involve  $t$ ,  $u$ -channel exchange of squarks, with the  $h$  emerging from either a  $\tilde{q}$  or a  $\tilde{\chi}_i^0$ , and one includes  $s$ -channel propagation of all neutral Higgs boson states of the BLSSM, with the  $h$  emerging from a  $\tilde{\chi}_i^0$ .

In order to be consistent with the current experimental bounds, we adopt a benchmark point (BP) from the scan performed in [2] with the following features (as mentioned, we take  $m_h = 125 \text{ GeV}$ ):

$$\begin{aligned} m_{\tilde{\chi}_1^0} &= 48.4 \text{ GeV}, & m_{h'} &= 263.8 \text{ GeV}, \\ m_{\tilde{\chi}_1^{\pm 1}} &= 712 \text{ GeV}, & m_{A'} \sim m_{H'} &= 800 \text{ GeV}, \\ m_{\tilde{\chi}_{2,3}^0} &= \mathcal{O}(100 \text{ GeV}), & m_{\tilde{\chi}_{4,5,6,7}^0} &= \mathcal{O}(\text{TeV}), \\ m_{\tilde{q}} &= \mathcal{O}(\text{TeV}), & m_A \sim m_H &= \mathcal{O}(\text{TeV}). \end{aligned}$$

Thus, for our BP, the DM is significantly lighter than heavier neutralinos and the neutral Higgs states, which will have implications for the event kinematics of the signal. We have checked that the selected BP is consistent with current LHC run-2 bounds and allowed by all the Higgs searches and measurements at the 95% CL as checked by HiggsBounds v5.3.2 [15–17] and HiggsSignals v2.2.3 [18–20], respectively. Moreover, the LSP satisfies the latest LUX bounds on the direct detection search for DM therein and

all other experimental limits [21]. However, the relic abundance depends on the details of the underlying cosmology (thermal or nonthermal abundance), so its constraints will not be considered here [22–25].

The total cross section for the combined signal topologies including SM-like Higgs decays to bottom quark pairs at  $\sqrt{s} = 14$  TeV is  $9 \times 10^{-3}$  pb. Assuming  $1000 \text{ fb}^{-1}$  of luminosity at the HL-LHC, this represents a sizable signal event sample, thereby affording one with the possibility of efficiently exploiting ML techniques.

### III. ANALYSIS

In this section, we describe the strategy behind our numerical analysis as well as the constituent elements of it, i.e., the event generation and detector simulation procedure, as well as the signal and background properties in terms of kinematics and color dynamics of jets.

#### A. Strategy

Having clarified the theoretical setup, we now showcase a phenomenological study of mono-Higgs signals in  $b\bar{b} + \text{MET}$  final states using ML methods based on three independent DNNs.

To start, global event information is extracted from constructing all relevant kinematic distributions for signal and background events. A MLP model is then adopted to optimize the separation power between the two by analyzing such distributions. The fact that some background processes have similar kinematic structures as the signal hinders the classification efficiency of this network, though. Moreover, the huge background cross section leads to smaller signal significance even after optimizing the cut on the output score. To improve the MLP performance, one could then apply initial cuts on some variables before feeding the distributions to the MLP to enhance the signal and suppress the backgrounds. Because of the correlations among the constructed kinematic variables, applying a cut on some variables will affect all others, though, which in turn may undermine the network classification

performance. To improve the impact of the initial cuts, one has then to decorrelate such a dependence across the kinematic variables via the square root of the covariance matrix as in [26]. However, in our analysis, we opted not to apply any such cuts on the constructed kinematic variables.

Instead, we adopted a second approach following the fact that  $SU(3)$  of color is conserved in the interaction processes and provides different color flow structures for different processes. This depends on the color nature of the interacting particles, e.g., the radiation pattern within and around the bottom quark pair from Higgs boson decays is expected to be different from the radiation pattern of bottom quark pairs from  $t\bar{t}$  production and decay or prompt QCD processes. To quantify the impact on the signal-to-background rate of different color structures, one can think of a LHC detector as a giant camera and the streams of hadrons as an image. Jet images can then be constructed as two-dimensional arrays in the  $(\eta, \phi)$  plane and the pixels can be weighted by the sum of the total transverse momentum deposited in the corresponding calorimetric region of the detector [27–31]. We then adopt a CNN model to analyze the constructed jet images in order to maximize the signal-to-background classification efficiency.

Finally, to incorporate the different data structures as inputs, whether kinematics or color flows, a two-stream DNN is constructed [32–34]. The first stream, which processes the input jet images, consists of convolution, max pooling, and dropout layers plus one flattened layer. The second stream, which processes the kinematic distributions, consists of fully connected and dropout layers. Both streams are then concatenated to one fully connected layer and one output layer with two neurons for predictions, hence, a hybrid DNN (HDNN).

Figure 2 shows the schematic architecture of the three DNNs: MLP (left), CNN (middle), and HDNN (right). Feeding the HDNN model with different kinds of information enables it to extract the characteristic features of the signal and background events without requiring initial cuts or manual feature optimization. Moreover, combining the global color structure and the events kinematics boosts the

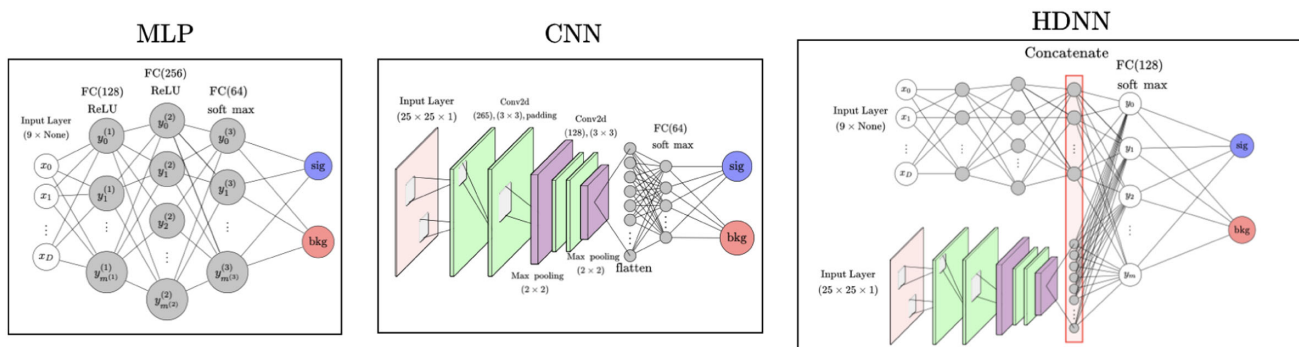


FIG. 2. A schematic architecture for the used DNNs: MLP (left), CNN (middle), and HDNN (right). Unlabeled layers of HDNNs are similar to the ones in MLP and CNN. The terms “FC” represents a fully connected neural network layer and “Conv2d” represents a two dimensional convolution neural network layer.



HDNN model performance above and beyond that of the single stream networks, whether MLP or CNN, where individual streams are used. In the following, we discuss the structure of the three DNNs in detail and illustrate how to prepare the input data for the networks. Before doing so, though, we describe how we have generated our Monte Carlo (MC) events.

### B. Events generation and detector simulation

Both signal and background are simulated with MadGraph5 [35], which is used to estimate multiparton amplitudes and to generate events for subsequent processing. All processes in Fig. 1 are computed at leading order (LO) except Higgs production from gluon-gluon fusion, which is calculated at next to LO in QCD using an effective coupling calculated by SPheno [36,37]. PYTHIA [38] is used for parton showering, hadronization, heavy flavor decays, and for adding the soft underlying event. The simulation of the response of, e.g., the ATLAS detector was done with the Delphes package [39]. We slightly modified the standard ATLAS card therein to allow for the extraction of the track and energy deposit information for the final state hadrons. In fact, reconstructed objects are simulated from the parametrized detector response and include tracks, calorimeter deposits, as well as high level objects such as isolated electrons, jets, and taus (both leptonic and hadronic), as well as MET.

### C. Signal and background kinematics

Considering the discussed BP and the signal topologies shown in Fig. 1, the dominant SM background contributions arise from the semileptonic decay of  $t\bar{t}$  pairs, digauge boson production  $pp \rightarrow hZ, ZZ$ , and, finally, vector boson production with bottom quark pairs stemming from QCD radiation,  $pp \rightarrow Vb\bar{b}$  with  $V = W$  or  $Z$ . We omit here trigauge boson production, which has a small cross section in comparison. Also, we neglect the processes  $pp \rightarrow b\bar{b}$  and  $pp \rightarrow b\bar{b} + j$  with MET from the jet misreconstruction or from mesons decay, by assuming large missing energy cut,  $\text{MET} \geq 250$  GeV [40,41].

Based on the kinematics of signal and backgrounds, in order to generate the events more effectively, we apply the following cuts at the simulation (i.e., MadGraph level):  $p_T(\text{jet}) \geq 20$  GeV,  $p_T(l) \geq 20$  GeV,  $p_T(b) \geq 20$ ,  $\text{MET} \geq 30$  GeV, and  $|\eta(l, \text{jet}, b)| \leq 2.5$  (hereafter,  $j = e, \mu$ ). Moreover, we require the events to have at least two  $b$  jets with cone radius  $R = 0.4$ <sup>1</sup> using a flat tagging efficiency of 70%. (As for the mistagging efficiency of gluon- and light-quark jets as  $b$  ones, we adopt a flat rate of  $10^{-3}$ .) After generating the events, we reconstruct nine

kinematic distributions for both signal and backgrounds as detailed below. To collect the two  $b$  jets from the Higgs boson decay, we do not require any mass window cut; rather we collect the  $b$ -jet pair that has the closest invariant mass to the SM-like Higgs boson mass.

Once events have passed the selection criteria, we exploit the following kinematic variables, which we have plotted in Fig. 3.

- (i) MET: Defined as  $\text{MET} = |-\sum v_i \vec{p}_T(v_i)|$ . Mono-Higgs signals with neutralino DM with mass about 50 GeV allow for large MET that can help to suppress the  $t\bar{t}$  and digauge boson backgrounds.
- (ii)  $p_T(b_1)$ : Transverse momentum of the leading  $b$  jet originating from SM-like Higgs boson decays. The rest mass of the (rather) heavy mediator Higgs boson [with mass of  $\mathcal{O}(100\text{--}800)$  GeV, see Fig. 1 (middle)], in comparison to the DM mass, boosts the final state  $b$  jets showing a similar distribution to that of  $t\bar{t}$  production and decay (wherein  $b$  jets originate from top quark decays), while this kinematics is notably different for the other backgrounds.
- (iii)  $p_T(b_2)$ : Transverse momentum of the second leading  $b$  jet, which shows a similar behavior as  $p_T(b_1)$  in both signal and background events.
- (iv)  $\eta(b_1)$ : Pseudorapidity of the leading  $b$  jet, which shows a similar behavior in both signal and background events except for the process  $pp \rightarrow Zb\bar{b}$ , showing a broader peak.
- (v)  $\eta(b_2)$ : Pseudorapidity of the second leading  $b$  jet, which shows a similar behavior as  $\eta(b_1)$  for all processes.
- (vi)  $p_T(b\bar{b})$ : Transverse momentum of the two  $b$  jets best reconstructing  $m_h$  which exemplifies the Higgs boson boost in signal events (again, signal and  $t\bar{t}$  events have similar distributions, which are in turn different from all other background processes).
- (vii)  $\eta(b\bar{b})$ : Pseudorapidity of the two  $b$  jets above which exhibits a deeper dip around 0 for background events, where  $b$  jets are scattered back to back [signal events with a larger boost show a somewhat different behavior at  $\eta(b\bar{b}) = 0$ ].
- (viii)  $\Delta R(b, \bar{b})$ : Angular distance separation between the two  $b$  jets reconstructing the Higgs boson, with  $\Delta R(b\bar{b}) = \sqrt{(\Delta\eta(b\bar{b}))^2 + (\Delta\phi(b\bar{b}))^2}$ . For background events via  $ZZ$  or  $hZ$  production, the two  $b$  jets are from the  $Z$  or  $h$  boson decay. When the  $Z/h$  boson is produced near its peak in the running of  $\sqrt{s}$ , most of the gauge/Higgs boson events have small boost factors. Then the pair of  $b$  jets from  $Z/h$  decay are expected to fly back to back. A similar behavior also applies to the  $b$  jets emerging from top quark decays, for which  $\Delta R(b, \bar{b})$  peaks around 3. For the QCD final state  $Zb\bar{b}$ , the angular separation between the two  $b$  jets has a broader peak, as the  $b$  jets come

<sup>1</sup>The analysis is insensitive to whether we use the  $k_T$  [42], anti- $k_T$  [43,44], or Cambridge-Aachen [45,46] jet clustering algorithm.

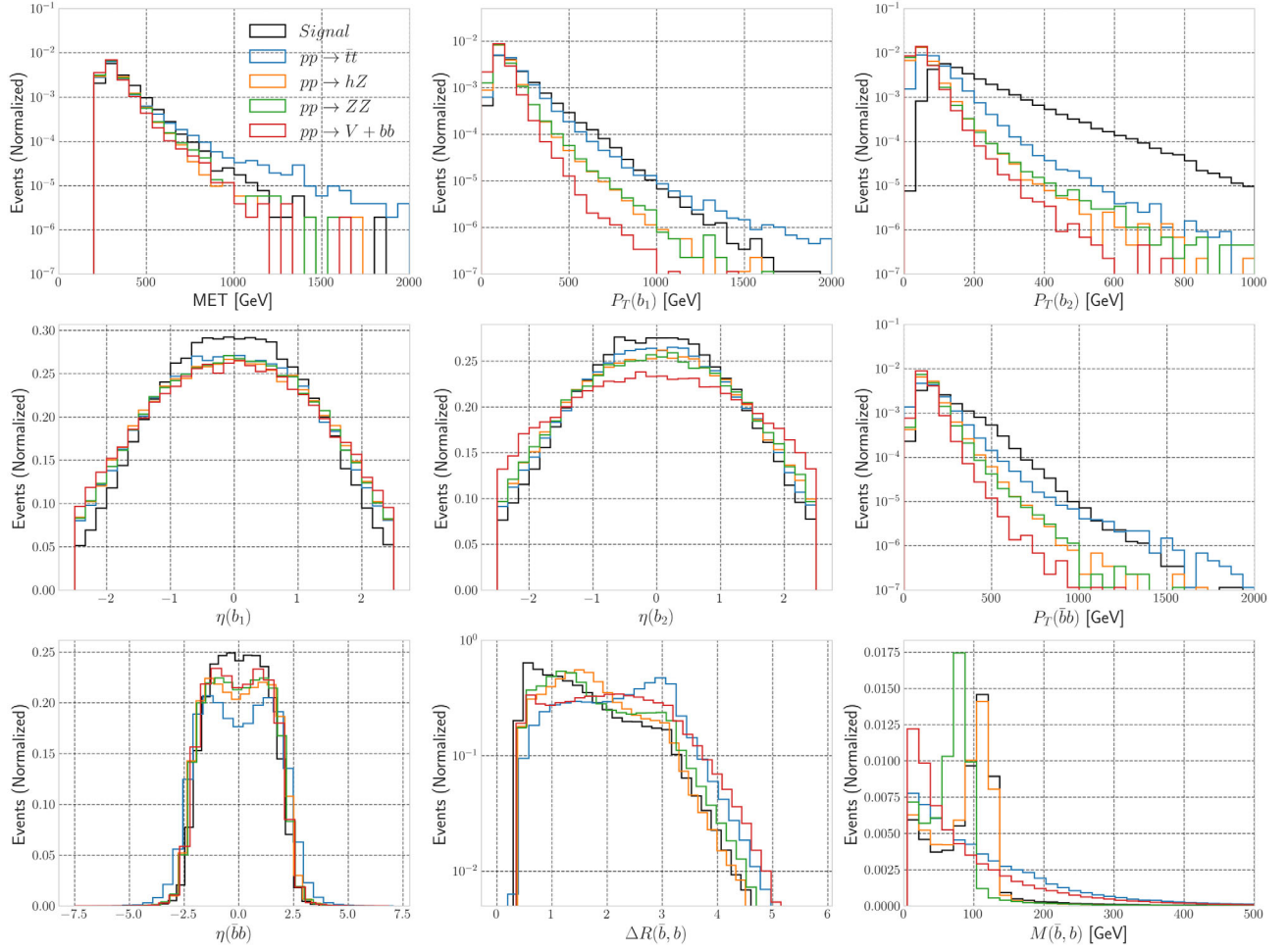


FIG. 3. The kinematic distributions discussed in the text for signal and background events superimposed and normalized to 1. The color codes apply for all distributions as follows: signal (black),  $pp \rightarrow t\bar{t}$  (blue),  $pp \rightarrow hZ$  (orange),  $pp \rightarrow ZZ$  (green), and  $pp \rightarrow Zb\bar{b}$  (red).

from QCD radiation. The signal events, with a larger boost of the  $b$  jets, allow for  $\Delta R(b, \bar{b})$  to peak around 1.

- (ix)  $M(b, \bar{b})$ : Invariant mass of the two  $b$  jets that is closest to the SM-like Higgs boson mass. For signal events,  $M(b, \bar{b})$  peaks at the Higgs boson mass, while, for the processes  $ZZ$  and  $hZ$ ,  $M(b, \bar{b})$  peaks at the  $Z$  and (somewhat more broadly than the signal) Higgs boson mass, respectively, while, for the  $t\bar{t}$  and  $Zb\bar{b}$  processes,  $M(b, \bar{b})$  does not show any obvious peak.

To adjust the reconstructed distributions as an input to the DNNs, we stack all background and signal events separately such that each dataset has dimensions of  $d_{\text{distribution}} = (N, 9)$ , with  $N$  being the total number of events. In our analysis, we use equal size datasets for signal and background events,  $N = 200,000$  each. As the network does not understand the meaning of signal or background, we assign a numeric label  $Y = 1$  to the former and  $Y = 0$  to (the whole of) the latter. Once the labels are

adjusted during the training of the network, the model tries to minimize the error between its predictions and the assigned labels for each dataset of signal or background events. The network then repeats this process until it reaches the desired classification accuracy. Finally, it is important to mention that a DNN convergence to a global minimum is sensitive to the ranges of the input datasets; this is why we normalize all distributions to one.

#### D. Color flow in signal and background

$SU(3)$  color symmetry implies that color is conserved in the jet interactions providing different color flows for different processes. The structure of the color flow depends on how the group indices are contracted at each vertex, e.g., jets produced from a color singlet have their indices contracted with each other, thereby forming a color dipole, while jets produced from the decay of a colored particle have their indices contracted with those of the parent particle, thereby forming isolated poles connected to the parent particle [47–49]. The global structure of the color

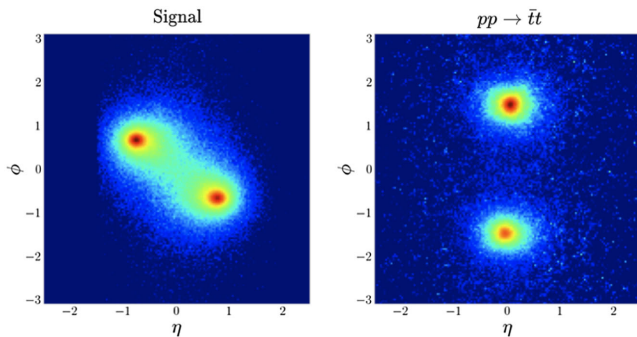


FIG. 4. Cumulative  $p_T$  distributions resulting from showering 10,000 times a single event at parton level for the signal (left) and  $t\bar{t}$  production (right).

flow can be seen in LHC events as a color string from the soft hadrons that stretches between two color-connected jets [33,50]. This dynamics thus potentially provides one with color observables that can be used to aid the search for new physics. Specific to our case, the two  $b$  quarks from the hadronic decay of the Higgs boson form a color dipole whose radiation pattern is contained primarily within a pair of cones around the two  $b$ -quark directions, with a tendency for more radiation to occur in the region between the two  $b$  quarks. In contrast, the two  $b$  quarks in the hadronic decay of a colored particle, e.g., a top (anti)quark, are color connected to the incoming proton beams, forming two isolated cones with less radiation in the region between the two  $b$  quarks. Since color flow is physical, it may be possible to extract the color connections of an event. In Fig. 4 we show, for illustrative purpose only, the cumulative transverse momentum distribution from showering a parton level signal and  $t\bar{t}$  event 10,000 times each. As can be seen, indeed, the two  $b$  quarks from the Higgs boson decay tend to radiate toward each other, allowing for soft radiation to fill the region between them, while the  $b$  quarks from (anti)top quark decay are color connected to the incoming beam and thus the soft radiation tends to fill the outer region between the two  $b$  quarks.

In order to explore the global color structure of actual signal and background events entering our MC analysis, we can use the energy deposit in the LHC calorimeters to

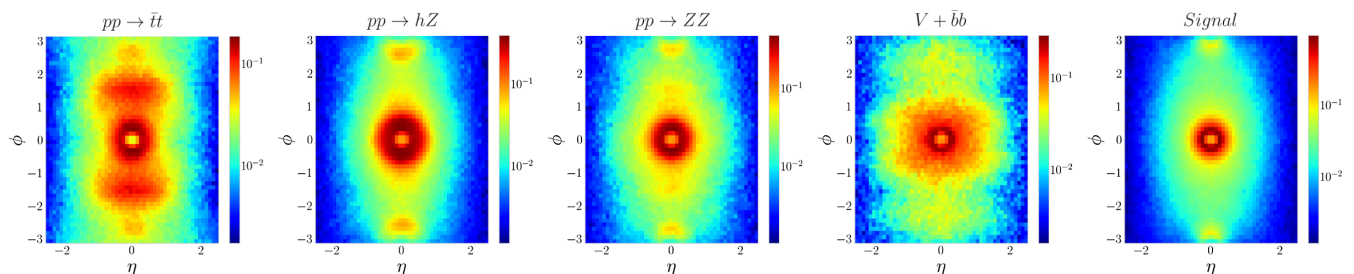


FIG. 5. Accumulated average of the discussed  $p_T$  distribution for 50,000 events after a Delphes simulation for signal and backgrounds events.

depict the information of the final state hadrons into images. Indeed, the calorimeters already provide the requisite pixelization of the constructed image. The pixel intensity of the image can be weighted according to the total transverse momentum of the final state hadrons, while the image dimensions will be the pseudorapidity and the azimuthal angle of the calorimetric element. The results in Fig. 4 are only valid in a statistical sense, in fact, since we use the same parton level event and shower it multiple times. In reality, events are independent and we remark that only one  $b$ -quark pair is embedded into an image with different locations for it in each event. Thus, in Fig. 5 we show the accumulated average of the described  $p_T$  distribution for 50,000 events (with image preprocessing steps discussed in the next section) after a Delphes simulation for signal and background events.

To adjust the reconstructed images as an input to the DNNs, we construct image datasets by embedding each event into an image with dimensions  $25 \times 25 \times 1$ . Specifically, we stack all signal and background images separately such that the dimensions of each set are  $d_{\text{images}} = (N, 25, 25, 1)$ , where  $N$  is the number of constructed images for signal and backgrounds, which we take to be  $N = 200,000$ . In order to improve the network performance and to minimize the ensuing error, it is important to properly process signal and background images before feeding them into the DNNs, as we shall discuss later in the text.

#### IV. DEEP LEARNING ANALYSIS

In this section, we discuss how to construct the aforementioned three independent DNNs to classify the mono-Higgs signal events from the various backgrounds taken together. We thus discuss the model architecture, preprocessing of input data, model training, and classification efficiency. Finally, we compare the classification performance for the three networks.

##### A. Image preprocessing

For each event we construct an image as a square array in the  $(\eta, \phi)$  plane with each pixel given by the total hadrons  $p_T$  deposited in the associated region in the calorimeter.



The rectangular region between  $-2.5 \leq \eta \leq 2.5$  and  $-\pi \leq \phi \leq \pi$  is discretized into a  $25 \times 25$  pixel grid. To ensure that the DNN is not learning spacetime symmetries, the jet images are preprocessed as follows.

- (1) Image cleansing: We consider only particles that have track information. Also, we remove leptons and photons from the constructed images.
- (2) Pixelization: We discretize the region in the  $(\eta, \phi)$  plane into a  $25 \times 25$  grid with each pixel weighted by the sum of the transverse momentum in it.
- (3) Centering: We shift the center of the image from  $(0,0)$  to  $(\frac{\eta_b + \eta_{\bar{b}}}{2}, \frac{\phi_b + \phi_{\bar{b}}}{2})$ .
- (4) Momentum smearing: We smear the transverse momentum using a Gaussian function with 3 standard deviations to reduce the number of sparse pixels in the images [51].
- (5) Normalization: We normalize the pixel intensity by dividing each pixel in the image by the maximum pixel intensity value.

After these image preprocessing steps, we end up with signal and background image datasets ready to be analyzed by the networks (herein, with image dimension of  $25 \times 25 \times 1$ ).

### B. Network architectures

As intimated, in this paper, we compare the performance of three DNNs in distinguishing mono-Higgs signals via  $h \rightarrow b\bar{b}$  from various backgrounds. We first use the MLP model that processes the nine constructed kinematic distributions. The model consists of three fully connected layers with a Rectified Linear Unit (ReLU) activation function and last output layer with two neurons and a soft-max activation function. The number of neurons in the first dense layer is 256, in the second layer is 128, and in the third layer is 64. To avoid overtraining, we insert a dropout layer after each dense layer with 20% dropout rate. An Adam optimizer, i.e., an algorithm for first order gradient-based optimization of stochastic objective functions [52], is used to minimize the loss function with learning rate  $\eta = 0.001$ .

In analyzing jet images, we use a CNN with four convolution layers, one dense layer, and one output layer. The first and second convolution layers have 256 kernels with kernel size 3, a ReLU activation function, and stride length of 1 plus; in order to keep the dimensions of the original input images, we use a padding layer. The third and fourth convolution layers both have 128 kernels with kernel size 3 and a ReLU activation function. After the second and fourth convolution layers, we use max-pooling layers with size  $2 \times 2$  with stride of length 2. After the pooling layer, we use a dropout layer with a 30% dropout rate. The last convolution layer is flattened and projected to one fully connected layer with 64 neurons and a ReLU activation function. The output layer has two neurons and a soft-max

activation function. Again, the Adam optimizer is used to minimize the loss function with learning rate  $\eta = 0.001$ .

Finally, a HDNN is constructed by combining the above models without the output layers; rather a layer is inserted to concatenate the two dense layers from the two models. A fully connected dense layer with 128 neurons is thus added with a ReLU activation function and a dropout layer with 30% dropout rate is also added. The last output layer consists of two neurons and a soft-max activation function. Finally, the Adam optimizer is used again to minimize the loss function with learning rate  $\eta = 0.001$ .

### C. Training the network

Once the datasets are prepared, we train the networks to learn the nonlinear relationships between the input data and their labels. For signal events, as mentioned, we assign the label  $Y = 1$  and for background events we assign the label  $Y = 0$ . In order to remove the network dependence on the position of the signal and background events, we stack the signal and background events in one dataset and shuffle it together with the assigned labels. During the network training stage, during each epoch (defined as number of passes of the entire datasets), the network updates the weights assigned to the neurons for each event via backward propagation of errors. The network then tries to minimize the error between its predictions and the true labels by reaching a global minimum of some loss function. Finally, the network repeats the process until it reaches the desired accuracy. Once the model is trained, we test it by using completely unseen new datasets to measure the network performance. For MLP and CNN models, we divide the datasets, both kinematic distributions and images capturing color dynamics, into 70% training and 30% testing samples with total size of 200,000. For the HDNN model, we use the same training and test datasets, with the same sizes now for each stream. Also, we train the model on equal size datasets for both signal and (whole) background. The network predictions error is quantified by using the categorical cross entropy loss, given as

$$\text{Loss} = -\sum_i Y_i \log(\hat{Y}_i), \quad (8)$$

with  $i = 0, 1$  for signal and background classes, respectively, and where  $Y_i, \hat{Y}_i$  are the true and predicted labels for each class. For all networks, we train the model with 20 epochs with batch size equaling a 500 sample. The dimension of the final output probability  $\hat{Y}$  is  $1 \times 2$ ,  $(\mathcal{P}_{sig}, \mathcal{P}_{bkg})$ , with  $\mathcal{P}$  ranging between  $[0, 1]$ . If  $\mathcal{P}_{sig} > 0.5$  ( $\mathcal{P}_{bkg} < 0.5$ ), the corresponding event is classified as most likely being a signal event, and if  $\mathcal{P}_{sig} < 0.5$  ( $\mathcal{P}_{bkg} > 0.5$ ), the corresponding event is classified as most likely being a background event.



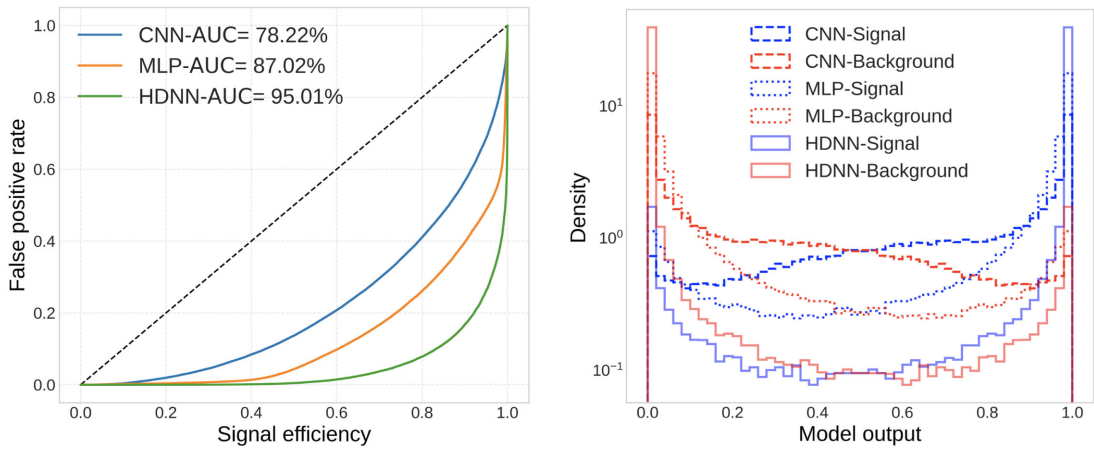


FIG. 6. Signal efficiency versus false positive rate with AUC distribution for the three ML types (left). Model output versus density for signal and background in the three ML types (right).

#### D. Hyperparameters optimization and cross validation

When creating a ML network, one has to define the model architecture that contains a set of parameters, for which values the network itself is unable to estimate, and thus the user has to fix these. Such parameters are called hyperparameters. The network performance is crucially dependent on the fixed values of these hyperparameters and thus we have to search for those that give the optimal network performance. A traditional way for hyperparameter optimization is the grid search. Grid search is arguably the most basic hyperparameter tuning method. With this technique, one simply builds a model for each possible combination of all hyperparameter values provided, evaluating each model and selecting the architecture that produces the best results. Although the grid search method can perfectly fix the hyperparameter values that optimize the network performance, it is very time consuming. Consequently, we use a more economical method called random grid search. Random grid search differs from grid search in that one no longer provides a discrete set of values to explore for each hyperparameter, rather, one provides a statistical distribution for each hyperparameter from which their values can be randomly sampled. To find the hyperparameter values for the networks, like done in Sec. IV B, we carried out a random grid search over the hyperparameters and fixed their values to the highest network performance. For the MLP, we randomly searched for the number of neurons in each layer and dropout rate, with number of neurons from 64 to 256 in each layer and dropout rate from 5% to 35%. For the CNN, we searched for the kernel numbers in each convolution layer, kernel, and pooling size as well as dropout rate, with number of kernels from 64 to 256 in each layer. For both convolution and pooling kernel size, we scanned over a list of symmetric kernels with size from (2,2) to (5,5), while we used the same MLP ranges to scan over the dropout rate. For the HDNN, in which we concatenated these two networks, thus leading to larger set

of hyperparameters, we just used the optimized hyperparameter values in each previous network.

The fact that the network performance can depend on the random partitioning of the available data into three sets, e.g., training, test, and validation set, which can change the network performance if we repeat the training and test steps with new splitting. A solution to this problem is a procedure called cross validation. In this approach, the dataset is split into  $k$  smaller sets and the network is trained on  $k - 1$  sets, while the remaining set is used for validation purposes. The performance measure reported by the cross validation method is then the average of the values computed in the loop. In our training, we use  $k = 5$  for all networks.

#### V. RESULTS

We employ our DNNs to perform the described mono-Higgs analysis at the HL-LHC with  $\sqrt{s} = 14$  TeV in energy and  $1000 \text{ fb}^{-1}$  of integrated luminosity. The discriminating power of each of the networks will be a measure of how well the signal and background may be characterized through their different features, all entangled together into several kinematic distributions and color flows.

Figure 6 (left) shows the recursive operating curve score for the three networks: MLP using kinematic distributions as input, CNN trained on jet images, and HDNN, which processes both kinematic distributions and jet images. Herein, the area under the curve (AUC) quantifies the ability of the network to correctly predict the event class, i.e., signal or background. The MLP gives better performance than the CNN network, while the HDNN has the best performance overall for separating the mono-Higgs signal events from the various backgrounds combined. Specifically, the AUC for the HDNN model (green curve) is 96.87%, while we find 88.84% and 78.81% for the MLP (orange curve) and CNN (blue curve) network, respectively. In Fig. 6 (right) we show the classifier distribution

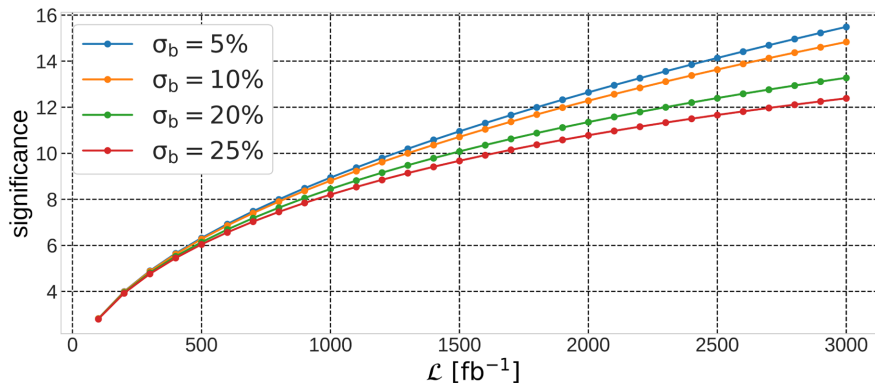


FIG. 7. Signal significance in terms of integrated luminosity with background uncertainty estimate of 5% (blue), 10% (orange), 20% (green), and 25% (red).

for both signal (blue histograms) and background (red histograms) in the case of the three networks: MLP (short dashed), CNN (long dashed), and HDNN (solid), each with the usual inputs. The DNN model output predictions ( $\mathcal{P}_{sig}, 1 - \mathcal{P}_{bkg}$ ) range from 0 to 1: the events with discriminant value near 1 are classified as signal-like events, while those near 0 are considered as backgroundlike events. The intersection area between the two distributions (signal and background) indicates the misclassified events by each network. Again, it is clear that HDNN is the best performer, ahead of MLP and CNN, in turn.

The optimization of the signal-to-background significance, as a function of the network output, has been performed using the following formula [53–55]:

$$\sigma = \left[ 2 \left( (N_s + N_b) \ln \frac{(N_s + N_b)(N_b + \sigma_b^2)}{N_b^2 + (N_s + N_b)\sigma_b^2} - \frac{N_b^2}{\sigma_b^2} \ln \left( 1 + \frac{\sigma_b^2 N_s}{N_b(N_b + \sigma_b^2)} \right) \right) \right]^{1/2}, \quad (9)$$

with  $N_s$  and  $N_b$  being the number of signal and background events, respectively, and  $\sigma_b$  parametrizing the systematic uncertainty on the latter. Upon adjusting the cut to maximize the signal-to-background yield, the CNN shows a best significance of  $4.9\sigma$ , MLP gives  $6.1\sigma$ , and HDNN yields  $8.9\sigma$ . Figure 7 shows the HDNN signal significance in terms of the integrated luminosity for different background uncertainties.

## VI. SUMMARY AND CONCLUSIONS

The discovery of a SM-like Higgs boson  $h$  with a mass of 125 GeV measured to percent precision by the ATLAS and CMS Collaborations and the extraction of its  $b\bar{b}$  decays at the LHC has opened up the possibility of using this decay channel as an indirect probe of the existence of DM (and its mediator) in so-called mono- $X$  channels, wherein  $X = h$ . However, this specific probe suffers from significant QCD background, when standard analysis techniques (i.e., those

used in cut-and-count approaches) are used to extract such a signal, so that other  $h$  decay channels typically afford one with greater sensitivity in comparison.

In this paper, we have carried out an analysis that showed the possibility of establishing such a signature in a specific case where the kinematics of the signal is peculiarly different from that of the background, as the DM mediator is a rather heavy object, in the form of a companion Higgs boson existing, e.g., in the BLSSM, which we have used as a template scenario of a supersymmetry model (wherein the DM candidate has a mass of around 50 GeV). This has been made possible by ML techniques, specifically exploiting HDNNs, which afford one with the ability of simultaneously using kinematic distributions and images of jets, the latter obtained by embedding the  $p_T$  distribution of the final state particles into visual representations, wherein the different color flow structure of the signal and background processes is accounted for.

Using such a new computational framework, we have been able to prove the sensitivity of mono-Higgs signals involving  $h \rightarrow b\bar{b}$  decays emerging in the BLSSM in the presence of large MET over a region of its parameter space comparable to those that have been previously shown to be accessible in such a theoretical framework, specifically involving  $h \rightarrow \gamma\gamma$  and  $ZZ^* \rightarrow 4l$  ( $l = e, \mu$ ) decays at the HL-LHC. We regard this as an encouraging result, enabled by ML techniques which one would hope to be also exploitable for other DM production mechanisms where the mediator is lighter, thereby originating less MET in comparison, in turn plagued by a much larger SM background. In fact, eventually, we deem this ML informed approach to be potentially also useful in the search for new Higgs bosons decaying into  $b\bar{b}$  pairs.

## ACKNOWLEDGMENTS

A.H. is funded by the Grant No. NRF-2021R1A2C4002551. S.M. is supported in part through the NExT Institute and the STFC Consolidated Grant No. ST/L000296/1.

- [1] S. Khalil and A. Masiero, *Phys. Lett. B* **665**, 374 (2008); Z. M. Burell and N. Okada, *Phys. Rev. D* **85**, 055011 (2012); P. Fileviez Perez and S. Spinner, *Phys. Rev. D* **83**, 035004 (2011); P. Fileviez Perez, S. Spinner, and M. K. Trenkel, *Phys. Rev. D* **84**, 095028 (2011).
- [2] W. Abdallah, A. Hammad, S. Khalil, and S. Moretti, *Phys. Rev. D* **95**, 055019 (2017).
- [3] M. Aaboud *et al.* (ATLAS Collaboration), *Phys. Lett. B* **786**, 59 (2018).
- [4] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **121**, 121801 (2018).
- [5] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **11** (2021) 209.
- [6] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **10** (2021) 013.
- [7] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **03** (2020) 025.
- [8] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **09** (2018) 046.
- [9] F. Gianotti, M. L. Mangano, T. Virdee, S. Abdullin, G. Azuelos, A. Ball, D. Barberis, A. Belyaev, P. Bloch, M. Bosman *et al.*, *Eur. Phys. J. C* **39**, 293 (2005).
- [10] H. E. Haber and G. L. Kane, *Phys. Rep.* **117**, 75 (1985).
- [11] J. F. Gunion and H. E. Haber, *Nucl. Phys.* **B272**, 1 (1986); **B402**, 567(E) (1993).
- [12] M. M. El Kheishen, A. A. Aboshousha, and A. A. Shafik, *Phys. Rev. D* **45**, 4345 (1992).
- [13] M. Guchait, *Z. Phys. C* **57**, 157 (1993); **61**, 178(E) (1994).
- [14] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **07** (2021) 208.
- [15] P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, and K. E. Williams, *Comput. Phys. Commun.* **181**, 138 (2010).
- [16] P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, and K. E. Williams, *Comput. Phys. Commun.* **182**, 2605 (2011).
- [17] P. Bechtle, O. Brein, S. Heinemeyer, O. Stål, T. Stefaniak, G. Weiglein, and K. E. Williams, *Eur. Phys. J. C* **74**, 2693 (2014).
- [18] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak, and G. Weiglein, *Eur. Phys. J. C* **74**, 2711 (2014).
- [19] O. Stål and T. Stefaniak, *Proc. Sci. EPS-HEP2013* (2013) 314 [arXiv:1310.4039].
- [20] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak, and G. Weiglein, *J. High Energy Phys.* **11** (2014) 039.
- [21] D. S. Akerib *et al.* (LUX Collaboration), *Phys. Rev. Lett.* **116**, 161301 (2016).
- [22] E. Aprile *et al.* (XENON Collaboration), *J. Cosmol. Astropart. Phys.* **04** (2016) 027.
- [23] E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley, Redwood City, CA, 1988), p. 719.
- [24] G. F. Giudice, E. W. Kolb, and A. Riotto, *Phys. Rev. D* **64**, 023508 (2001).
- [25] T. Moroi and L. Randall, *Nucl. Phys.* **B570**, 455 (2000).
- [26] A. Hocker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov *et al.*, arXiv:physics/0703039.
- [27] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, *J. High Energy Phys.* **01** (2017) 110.
- [28] K. Fraser and M. D. Schwartz, *J. High Energy Phys.* **10** (2018) 093.
- [29] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, *J. High Energy Phys.* **02** (2015) 118.
- [30] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, and M. Perelstein, *J. High Energy Phys.* **07** (2015) 086.
- [31] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, *J. High Energy Phys.* **07** (2016) 069.
- [32] J. Lin, M. Freytsis, I. Moulton, and B. Nachman, *J. High Energy Phys.* **10** (2018) 101.
- [33] J. H. Kim, M. Kim, K. Kong, K. T. Matchev, and M. Park, *J. High Energy Phys.* **09** (2019) 047.
- [34] L. Huang, S. b. Kang, J. H. Kim, K. Kong, and J. S. Pi, *J. High Energy Phys.* **08** (2022) 114.
- [35] 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.
- [36] W. Porod, *Comput. Phys. Commun.* **153**, 275 (2003).
- [37] W. Porod and F. Staub, *Comput. Phys. Commun.* **183**, 2458 (2012).
- [38] T. Sjostrand, S. Mrenna, and P. Z. Skands, *J. High Energy Phys.* **05** (2006) 026.
- [39] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [40] L. Carpenter, A. DiFranzo, M. Mulhearn, C. Shimmmin, S. Tulin, and D. Whiteson, *Phys. Rev. D* **89**, 075017 (2014).
- [41] D. Bhowmik, J. Lahiri, S. Bhattacharya, B. Mukhopadhyaya, and R. K. Singh, *Eur. Phys. J. C* **82**, 914 (2022).
- [42] S. D. Ellis and D. E. Soper, *Phys. Rev. D* **48**, 3160 (1993).
- [43] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [44] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Nucl. Phys.* **B406**, 187 (1993).
- [45] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *J. High Energy Phys.* **08** (1997) 001.
- [46] M. Wobisch and T. Wengler, arXiv:hep-ph/9907280.
- [47] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, *Phys. Rev. D* **67**, 014026 (2003).
- [48] K. Hagiwara and Y. Takaesu, *Eur. Phys. J. C* **71**, 1668 (2011).
- [49] W. Kilian, T. Ohl, J. Reuter, and C. Speckner, *J. High Energy Phys.* **10** (2012) 022.
- [50] J. Gallicchio and M. D. Schwartz, *Phys. Rev. Lett.* **105**, 022001 (2010).
- [51] T. Buss, B. M. Dillon, T. Finke, M. Krämer, A. Morandini, A. Mück, I. Oleksiyuk, and T. Plehn, arXiv:2202.00686.
- [52] D. P. Kingma and J. Ba, arXiv:1412.6980.
- [53] T. Abe *et al.* (LHC Dark Matter Working Group), *Phys. Dark Universe* **27**, 100351 (2020).
- [54] S. Antusch, E. Cazzato, O. Fischer, A. Hammad, and K. Wang, *J. High Energy Phys.* **10** (2018) 067.
- [55] S. Antusch, A. Hammad, and A. Rashed, *Phys. Lett. B* **810**, 135796 (2020).