

Tagging fully hadronic exotic decays of the vectorlike B quark using a graph neural network

Jai Bardhan^{1,*}, Tanumoy Mandal^{2,†}, Subhadip Mitra^{1,3,‡}, Cyrin Neeraj^{1,§} and Mihir Rawat^{1,||}

¹Center for Computational Natural Sciences and Bioinformatics,

International Institute of Information Technology, Hyderabad 500 032, India

²Indian Institute of Science Education and Research Thiruvananthapuram, Vithura, Kerala, 695 551, India

³Center for Quantum Science and Technology, International Institute of Information Technology, Hyderabad 500 032, India



(Received 25 May 2025; accepted 27 June 2025; published 15 July 2025)

Following up on our earlier study [J. Bardhan *et al.*, *Phys. Rev. D* **107**, 115001 (2023)], we investigate the LHC prospects of pair-produced vectorlike B quarks decaying exotically to a new gauge-singlet (pseudo)scalar field Φ and a b quark. After the electroweak symmetry breaking, the Φ decays predominantly to gg/bb final states, leading to a fully hadronic $2b + 4j$ or $6b$ signature. Because of the large Standard Model background and the lack of leptonic handles, it is a difficult channel to probe. To overcome the challenge, we employ a hybrid deep learning model containing a graph neural network followed by a deep neural network. We estimate that such a state-of-the-art deep learning analysis pipeline can lead to a performance comparable to that in the semileptonic mode, taking the discovery (exclusion) reach up to about $M_B = 1.8(2.4)$ TeV at HL-LHC when B decays fully exotically, i.e., $\text{BR}(B \rightarrow b\Phi) = 100\%$.

DOI: 10.1103/hk9n-37kx

I. INTRODUCTION

New fermions are usually considered vectorlike in scenarios beyond the Standard Model (SM) to avoid introducing gauge anomalies [1–3]. The LHC has been searching for vectorlike quarks (VLQs) that mix with the third-generation ones at the ATLAS and CMS detectors (see, e.g., Ref. [4]). So far, none of these detectors has observed any VLQ signatures, pushing the mass exclusion limits on them to about 2 TeV. Recently, there has been considerable interest in nonstandard decays of VLQs [5–21]. While there could be various motivations to add extra particles to the new particle spectrum from a model-building perspective (consider, e.g., the Maverick Top-partner model [22], where the VLQ can decay to a dark photon [23,24]), any extra decay mode would reduce the branching ratios (BRs) in the standard modes. Since the

experiments considered only the standard modes, the limits on VLQs could be actually weaker.

This paper is part of a series (see Refs. [25–27]) where we study the prospects in looking for VLQs decaying through an exotic decay mode, $q\Phi$ (where q is a third-generation quark, and Φ is a new spinless boson singlet under the SM gauge group) at the LHC. In Ref. [25], we charted out the possibilities at the LHC. Since a spinless gauge-singlet particle does not couple to SM particles directly (except, possibly, the Higgs boson), it is difficult to probe experimentally. Hence, *a priori*, it might seem that a VLQ dominantly decaying to Φ would have poor prospects at the LHC; however, in Refs. [26,27], we showed that this is not the case. Even if, for a conservative estimation, we assume Φ has no direct coupling to particles other than a VLQ, it gains such couplings once the electroweak symmetry (EWS) breaks and the VLQ mixes with SM quarks. Depending on the VLQ(s), a Φ could then decay to a pair of b or t quarks at the tree level and to a pair of SM bosons (like gg , $\gamma\gamma$, $Z\gamma$, or a pair of heavy SM bosons if kinematically allowed) at the loop level. Interestingly, the loop-mediated decay of the Φ to a pair of gluons can dominate over all other decay modes, including the tree-level ones, in a large part of the parameter space. We analyzed the $pp \rightarrow TT \rightarrow (t\Phi)(t\Phi) \rightarrow (tgg)(tgg)$ process to estimate the discover/exclusion prospects of a singlet T quark at the LHC in Ref. [26]. We studied the B quark prospects in Ref. [27] through a mixed-decay channel: $pp \rightarrow BB \rightarrow (b\Phi)(tW) \rightarrow (bgg/bbb)(tW)$. Our analyses

*Contact author: jai.bardhan@research.iiit.ac.in

†Contact author: tanumoy@iisertvm.ac.in

‡Contact author: subhadip.mitra@iiit.ac.in

§Contact author: cyrin.neeraj@research.iiit.ac.in

||Contact author: mihir.r@research.iiit.ac.in

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

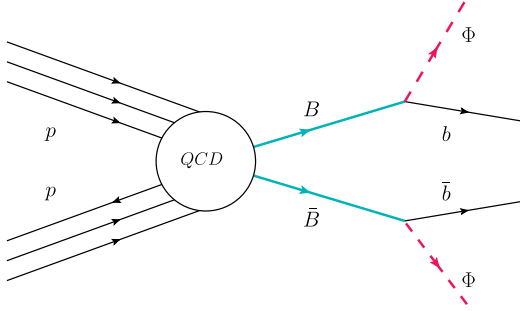


FIG. 1. Signal topology.

showed good prospects for both cases, especially with machine learning techniques.

In this paper, we investigate the collider prospects of pair-produced B quarks that decay predominantly to the singlet Φ in fully hadronic final states (see Fig. 1). Earlier, we selected those channels because a top quark can decay to produce a (high- p_T , in these cases) lepton in the final states, simplifying the analyses. Because of the large QCD background in a hadron collider, fully hadronic final states are highly complicated to isolate. With sophisticated tagging algorithms, one can still tag a highly boosted hadronic top quark. However, without identifiable known structures, probing a B quark becomes formidable with a simple analysis when the Φ decays to $b\bar{b}$ or $g\bar{g}$ final states. We use a graph neural network (GNN) model to separate the signal events from those coming from the significant background processes after passing them through some basic signal selection criteria. A key advantage of representing event data as graphs is that GNNs are permutation invariant, allowing us to train the model with variable numbers of nodes and edges. Therefore, we can use all the well-defined reconstructed objects in events passing, even those playing no role in the basic selection.

The plan of the paper is as follows. In Sec. II, we sketch an outline of the extended VLQ models. Section III describes the event-selection criteria and background processes considered in our analysis. Section IV presents the GNN model in detail, along with the design choices we made and the training details. In Sec. V, we present the HL-LHC reach of the singlet and doublet VLQ models. We conclude the paper in Sec. VI.

II. PHENOMENOLOGICAL MODELS

We are interested in the following signatures:

$$pp \rightarrow B\bar{B} \rightarrow (b\Phi)(\bar{b}\Phi) \rightarrow \begin{cases} (bgg)(\bar{b}gg) \\ (bb\bar{b})(\bar{b}b\bar{b}) \end{cases}. \quad (1)$$

Below, we discuss the essential phenomenological parameters in the singlet and doublet $B + \Phi$ models that can produce these processes.

A. Singlet B model

In the notation of Ref. [25], the mass terms relevant to the bottom sector can be written as

$$\mathcal{L} \supset -\{\lambda_b(\bar{Q}_L H)b_R + \omega_B(\bar{Q}_L H)B_R + M_B\bar{B}_L B_R + \text{H.c.}\}, \quad (2)$$

where $H = 1/\sqrt{2}(0 \ \nu)^T$, with ν being the vacuum expectation value of the Higgs field, λ_b is the b -quark Yukawa coupling, the off diagonal ω_B parametrizes the mixing between the b and B quarks, and M_B is the mass of B . Once EWS breaks, we get the mass terms as

$$\mathcal{L}_{\text{mass}} = \begin{pmatrix} \bar{b}_L & \bar{B}_L \end{pmatrix} \begin{pmatrix} \lambda_b \frac{v}{\sqrt{2}} & \mu_{B1} \\ 0 & M_B \end{pmatrix} \begin{pmatrix} b_R \\ B_R \end{pmatrix} + \text{H.c.}, \quad (3)$$

where $\mu_{B1} = \omega_B v/\sqrt{2}$. The matrix is diagonalized by a biorthogonal rotation with two mixing angles θ_L and θ_R ,

$$\begin{pmatrix} b_{L/R} \\ B_{L/R} \end{pmatrix} = \begin{pmatrix} c_{L/R} & s_{L/R} \\ -s_{L/R} & c_{L/R} \end{pmatrix} \begin{pmatrix} b_{1L/R} \\ b_{2L/R} \end{pmatrix}, \quad (4)$$

where $s_P = \sin \theta_P$ and $c_P = \cos \theta_P$ for the two chirality projections, and b_1 and b_2 are the mass eigenstates. We identify b_1 as the physical bottom quark and b_2 is mostly the B quark. Hence, we use the notations B and b_2 interchangeably.

If we express the Φ interactions before the breaking of EWS as

$$\mathcal{L}_{\text{int}}^{B\Phi} = -\lambda_\Phi^a \Phi \bar{B}_L \Gamma B_R - \lambda_\Phi^b \Phi \bar{B}_L \Gamma b_R + \text{H.c.}, \quad (5)$$

where $\Gamma = \{1, i\gamma_5\}$ for $\Phi = \{\text{scalar } \phi, \text{pseudo scalar } \eta\}$, expanding in terms of b_1, b_2 , we get

$$\begin{aligned} \mathcal{L}_{\text{int}}^{b_2\Phi} &= -\lambda_\Phi^a \Phi (c_L \bar{b}_{2L} - s_L \bar{b}_L) \Gamma (c_R b_{2R} - s_R b_R) \\ &\quad - \lambda_\Phi^b \Phi (c_L \bar{b}_{2L} - s_L \bar{b}_L) \Gamma (c_R b_R + s_R b_{2R}) \\ &\quad + \text{H.c.}, \end{aligned} \quad (6)$$

after the symmetry breaking. Thus, apart from the masses, the couplings λ_a and λ_b and the mixing angles form the set of parameters of the singlet B model. As explained in Ref. [25], we can draw two types of limits on the VLQ + Φ models from the current LHC data: (1) We can rescale the current mass exclusion limits on VLQs from their pair-production searches to account for the Φ decay mode, and (2) recast the heavy resonance searches to put constraints on Φ production at hadron colliders. The strongest limit on Φ comes from the clean two-photon search data. (See Refs. [25,27] for a detailed discussion on the constraints on the VLQ + Φ models from the LHC data.) Fig. 3 shows different projections of the allowed parameter region after these constraints are considered. The colored areas are

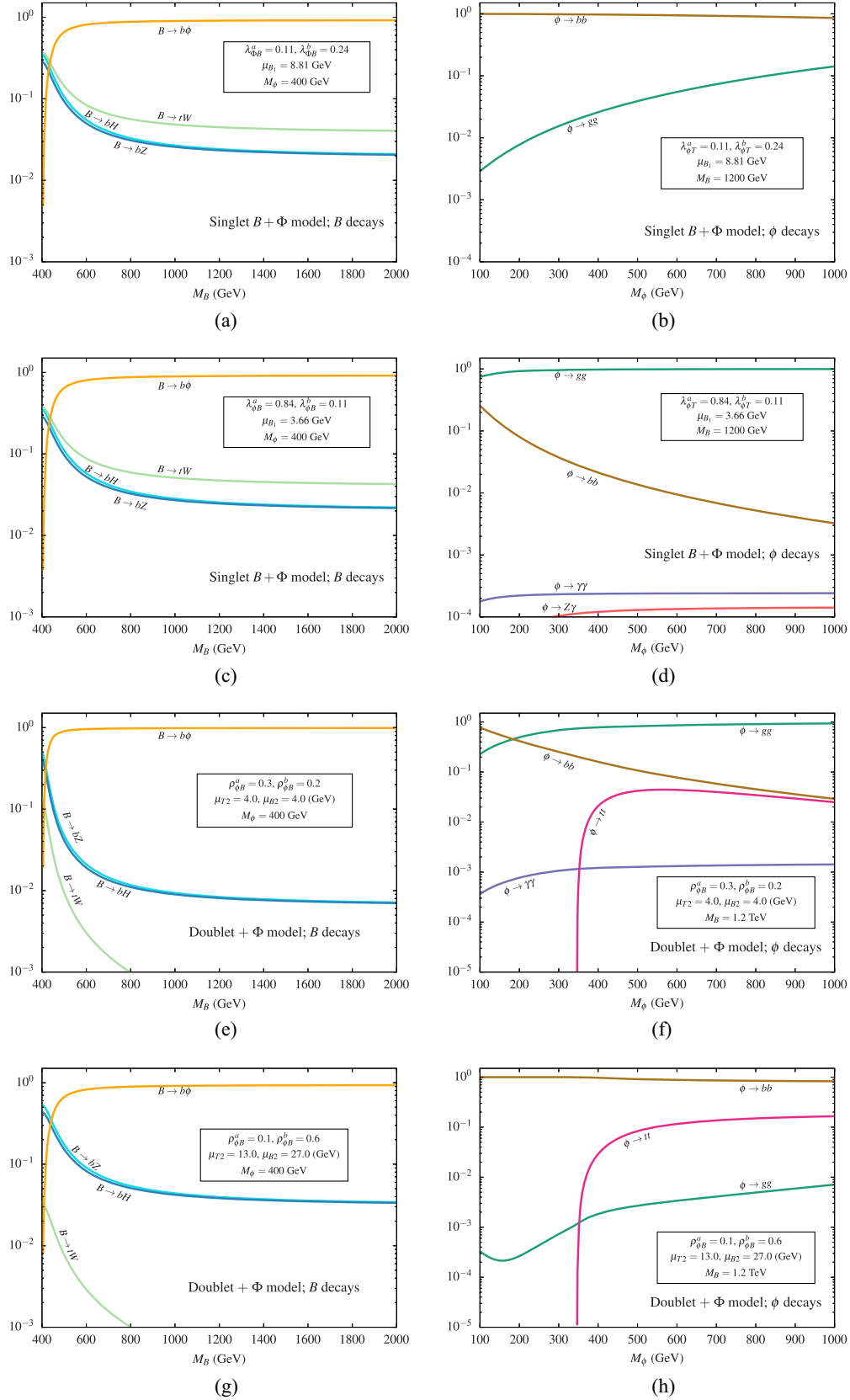


FIG. 2. Decays of B and ϕ in the singlet B [(a)–(d)] and doublet models [(e)–(h)] for some benchmark parameters, where $B \rightarrow b\phi$ decay dominates over all the other decay modes. On the right panel, we show that either $\phi \rightarrow gg$ or $\phi \rightarrow b\bar{b}$ can dominate in both scenarios. For more benchmarks and a detailed discussion on the final states from extended VLQ models, see Ref. [25].

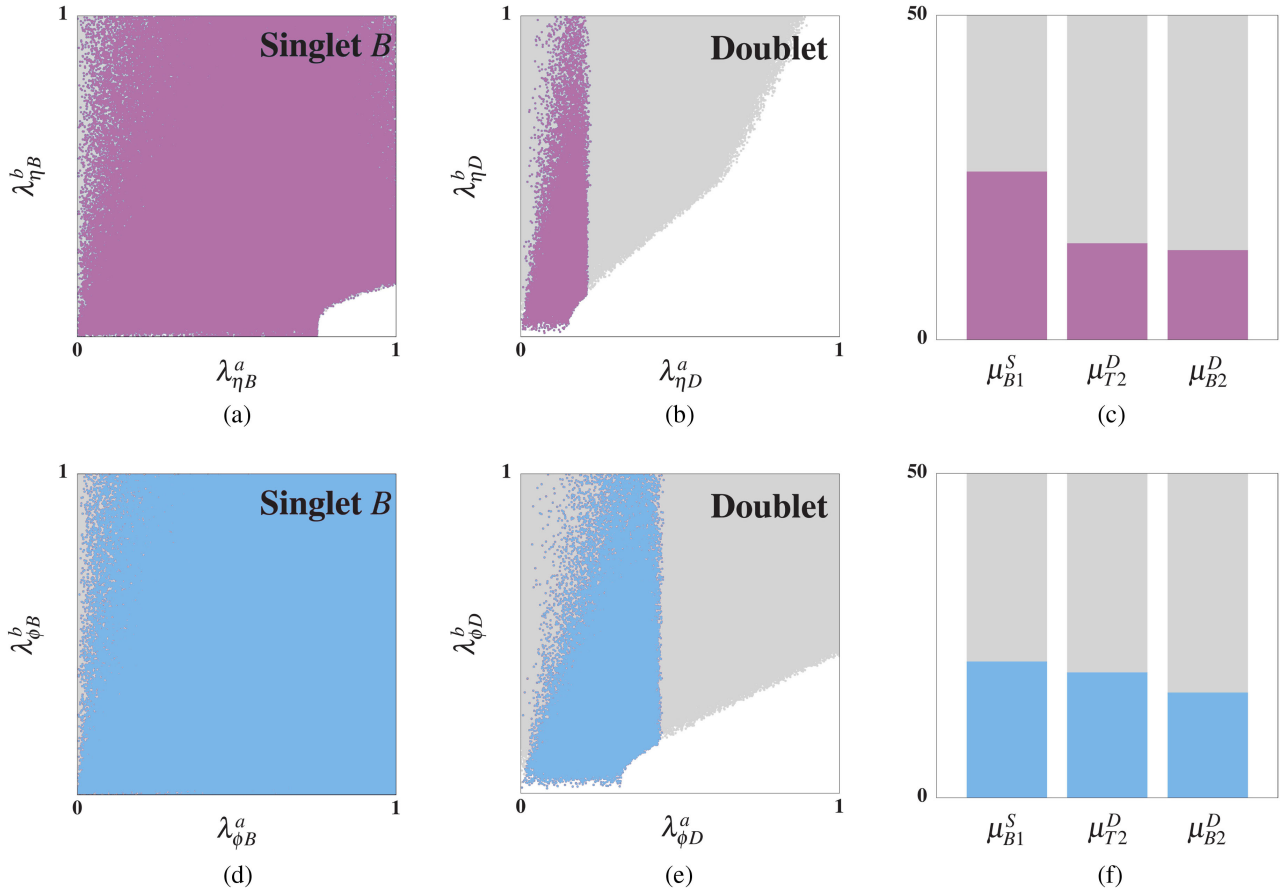


FIG. 3. Parameter scans for VLQ $B + \Phi$ models for benchmark point $(M_B, M_\Phi) = (1.2, 0.4)$ TeV. The top row (a)–(c) shows the scans for the model where the new singlet is a pseudoscalar boson (η). The bottom row shows (d)–(f) shows the scans for the model where the new singlet is a scalar boson ϕ . The gray-colored regions are allowed by reinterpreted LHC limits. Colored in purple [blue] are the regions where $\text{BR}(\eta \rightarrow gg) \geq 50\%$ [$\text{BR}(\phi \rightarrow gg) \geq 50\%$].

allowed. The dark region is where the loop-mediated $\Phi \rightarrow gg$ decay dominates, i.e., $\text{BR}(\Phi \rightarrow gg) \geq 50\%$, giving a dijet signature at the colliders. In this region, a dedicated collider analysis in fully hadronic modes is necessary to look for VLQ + Φ models.

B. Doublet VLQ

When the T and B quarks together forms a weak doublet, $\mathcal{F} = (TB)^T = (\mathbf{3}, \mathbf{2}, 1/6)$, we can write the terms relevant for the quark masses as

$$\begin{aligned} \mathcal{L} \supset & -\lambda_t(\bar{Q}_L \tilde{H})t_R + \rho_T(\bar{\mathcal{F}}_L \tilde{H})t_R + \lambda_b(\bar{Q}_L H)b_R \\ & + \rho_B(\bar{\mathcal{F}}_L H)b_R + M_F \bar{\mathcal{F}}_L \mathcal{F}_R + \text{H.c.} \end{aligned} \quad (7)$$

From this, we get the following mass matrices:

$$\begin{aligned} \mathcal{L}_{\text{mass}}^{\mathcal{F}} = & \begin{pmatrix} \bar{t}_L & \bar{T}_L \end{pmatrix} \begin{pmatrix} \lambda_t \frac{v}{\sqrt{2}} & 0 \\ \rho_T \frac{v}{\sqrt{2}} & M_T \end{pmatrix} \begin{pmatrix} t_R \\ T_R \end{pmatrix} \\ & + \begin{pmatrix} \bar{b}_L & \bar{B}_L \end{pmatrix} \begin{pmatrix} \lambda_b \frac{v}{\sqrt{2}} & 0 \\ \rho_B \frac{v}{\sqrt{2}} & M_B \end{pmatrix} \begin{pmatrix} b_R \\ B_R \end{pmatrix} + \text{H.c.} \end{aligned} \quad (8)$$

The interactions between Φ and the doublet $\mathcal{F}$ can be written as

$$\mathcal{L}_{\text{int}}^{\Phi \mathcal{F}} = -\lambda_{\Phi D}^a \Phi \bar{\mathcal{F}}_L \Gamma \mathcal{F}_R - \lambda_{\Phi D}^b \Phi \bar{\mathcal{F}}_R \Gamma Q_L + \text{H.c.} \quad (9)$$

Expanding, we get the interactions between Φ and B ,

$$\begin{aligned} \mathcal{L} \supset & -\lambda_{\Phi D}^a \Phi (c_L^B \bar{b}_{2L} - s_L^B \bar{b}_L) \Gamma (c_R^B \bar{b}_{2R} - s_R^B \bar{b}_R) \\ & + \lambda_{\Phi D}^b \Phi (c_R^B \bar{b}_{2R} - s_R^B \bar{b}_R) \Gamma (c_L^B \bar{b}_L + s_L^B \bar{b}_{2L}) + \text{H.c.} \end{aligned} \quad (10)$$

In Fig. 2, we show the decays of the B quark and the scalar, ϕ , for some benchmark choices of parameters where the $B \rightarrow b\phi$ decay dominates over all other decay modes and ϕ (essentially) decays hadronically. Note that since no fine-tuning is needed to find such parameter points, one can easily find similar benchmarks for the pseudoscalar, η . Figure 3 illustrates the dependence of the dominant decay mode of Φ on the value of the off diagonal mass term: $\mu_{B1} = \omega_B v / \sqrt{2}$, for the singlet model, and $\mu_{T2}^D = \mu_{T2} = \rho_T v / \sqrt{2}$ and $\mu_{B2}^D = \mu_{B2} = \rho_B v / \sqrt{2}$, for the doublet model.

The $\Phi \rightarrow gg$ mode is enhanced for small mixing values, and the tree-level $\Phi \rightarrow b\bar{b}$ mode starts dominating with increasing μ_{B1} . In the doublet mode, Φ can also decay to a $t\bar{t}$ pair (if kinematically allowed). However, the combined $\Phi \rightarrow gg/bb$ mode dominates.

III. COLLIDER PHENOMENOLOGY

We use FeynRulesv2.3 [28] to obtain the Universal FeynRules Output [29] model files, which we use in MadGraph5v3.3 [30] to simulate the hard scatterings at the leading order. We use Pythia8 [31] for showering and hadronization, and Delphes3 [32] for simulating a generic LHC detector environment. The events are generated at $\sqrt{s} = 14$ TeV. The b -tagging efficiency and the mistag rate for the lighter quarks were updated to reflect the medium working point of the DeepCSV tagger from Ref. [33]. Our analysis relies on two types of jets clustered using the anti- k_T algorithm [34]—one with $R = 0.4$ (AK4) and the other with $R = 1.2$ (fatjet). The AK4 objects are required to pass a minimum- p_T cut: $p_T > 20$ GeV and have $|\eta| < 5$. We use the next-to-next-to-leading order (NNLO) signal cross sections from Ref. [35] (Fig. 4) and for the background processes, we use the highest-order cross sections available in the literature (Table I). We scan over a wide kinematic range—we take $M_\Phi \in [300, 1000]$ GeV and $M_B \in [1.0, 3.0]$ TeV.

A. The background and selection criteria

Our signal has at least two b jets and two Φ (fat)jets originating in heavy particles, but no isolated leptons. We collect all the major background processes that could lead to similar final states in Table I. With the signal topology in

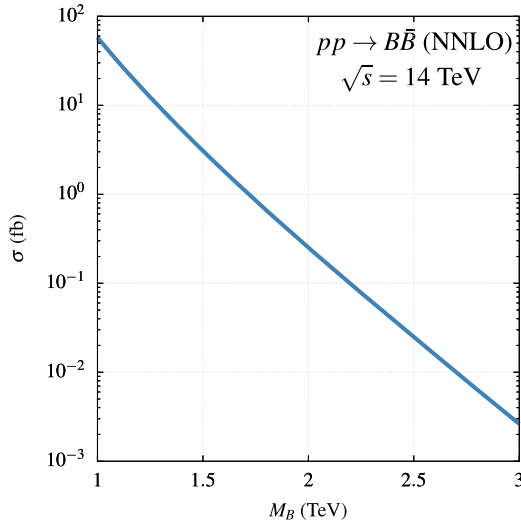


FIG. 4. The signal cross section at the 14 TeV LHC calculated from the leading order cross section as $\sigma(\text{LO}) \times K_{\text{NNLO}}$. We estimate the NNLO QCD K factor to be about 1.3 from Ref. [35].

TABLE I. Higher-order cross sections of the SM backgrounds considered in our analysis.

Background processes	σ (pb)	QCD order
$V + \text{jets}$ [36,37]	$W + \text{jets}$ 1.95×10^5	NLO
	$Z + \text{jets}$ 6.33×10^4	NNLO
$t\bar{t}$ [38]	$t\bar{t} + \text{jets}$ 988.57	N ³ LO
Single t [39]	tW 83.10	N ² LO
	$t\bar{b}$ 248.00	N ² LO
	$t + \text{jets}$ 12.35	N ² LO
$VV + \text{jets}$ [40]	$WW + \text{jets}$ 124.31	NLO
	$WZ + \text{jets}$ 51.82	NLO
VH [41]	WH 1.50	NNLO (NLO EW)
	ZH 0.98	NNLO (NLO EW)
$t\bar{t}V$ [42,43]	$t\bar{t}Z$ 1.05	NLO + NNLL
	$t\bar{t}W$ 0.65	NLO + NNLL
	$t\bar{t}H$ 0.61	NLO

mind, we first pass the events through the following selection criteria:

$\mathfrak{C}_1$: ℓ -veto, where $\ell \in \{e, \mu\}$ and $H_T > 1$ TeV.

We require an event to have no electrons or muons, and the scalar sum of hadronic transverse momenta (H_T) should be at least 1 TeV.

$\mathfrak{C}_2$: At least three AK4 jets with $p_{T1}, p_{T2} > 100$ GeV and $p_{T3}, \dots, p_{Tn} > 30$ GeV.

In our analysis, we only consider reconstructed jets with a minimum p_T of 30 GeV. We also demand the p_T of the top two leading- p_T jets should be at least 100 GeV.

$\mathfrak{C}_3$: At least two AK4 jets in the event must be b tagged.

$\mathfrak{C}_4$: At least one fatjet ($R = 1.2$) with $p_T > 200$ GeV.

We tag one of the Φ jets in the event as a fatjet of $R = 1.2$ with minimum $p_T > 200$ GeV. We also demand that the mass of fatjet, $M_J > 250$ GeV. To cover both $\Phi \rightarrow gg/bb$, our analysis relies only on the two-pronged nature of Φ .

$\mathfrak{C}_5$: At least two b -tagged jets with $\Delta R_{b_i, j} \geq 1.2$.

We demand that the event has at least two b -tagged jets that are well-separated from the fatjet; i.e., they do not come from the Φ .

The strong p_T cuts and the demand for b -tagged jets essentially eliminate the QCD multijet background. These topology-motivated cuts also reduce the total background from other processes significantly (roughly by an order of magnitude) while retaining a large part of the signal. We can see the reduction in Table II. Figure 5 shows the efficiency of the selection cuts at different parameter points. However, as expected, the set of background events remains overwhelmingly larger in this hadronic channel even after the cuts. We collect all the signal and background events that pass the selection cuts and feed them to a GNN model for further analysis.

TABLE II. Cut flow for the four benchmark choices of the signal and the relevant background processes. We use the MLM jet-parton shower matching technique [44] to generate the background samples as indicated by the additional jets in brackets. Wherever necessary, we simulate background events with generation-level cuts to reduce computational time. The events are estimated for luminosity $\mathcal{L} = 3 \text{ ab}^{-1}$.

	Selection criteria				
	$\mathfrak{C}_1$	$\mathfrak{C}_2$	$\mathfrak{C}_3$	$\mathfrak{C}_4$	$\mathfrak{C}_5$
Signal benchmarks					
$M_B = 1200 \text{ GeV}, M_\Phi = 400 \text{ GeV}$	4.90×10^4	4.90×10^4	4.76×10^4	3.76×10^4	3.22×10^4
$M_B = 1200 \text{ GeV}, M_\Phi = 900 \text{ GeV}$	4.91×10^4	4.91×10^4	4.78×10^4	3.90×10^4	3.35×10^4
$M_B = 2000 \text{ GeV}, M_\Phi = 700 \text{ GeV}$	754	754	725	664	571
$M_B = 2800 \text{ GeV}, M_\Phi = 400 \text{ GeV}$	19	19	18	18	14
Background processes					
$t_h t_h(+j)$	2.21×10^7	2.21×10^7	1.22×10^7	4.504×10^6	1.38×10^6
$Z_{bb}(+j)$	1.36×10^7	1.18×10^7	2.54×10^6	4.11×10^5	1.45×10^5
$tb(+j)$	4.28×10^5	4.14×10^5	2.20×10^5	6.21×10^4	1.44×10^4
$t_h t_h H(+j)$	5.50×10^4	5.50×10^4	4.74×10^4	2.40×10^4	1.13×10^4
$W_h(+j)$	8.34×10^6	7.09×10^6	1.85×10^5	3.98×10^4	8.74×10^3
$t_h t_h W_h(+j)$	7.15×10^4	7.15×10^4	3.97×10^4	1.84×10^4	6.67×10^3
$W_h Z_{bb}(+j)$	1.84×10^5	1.81×10^5	6.24×10^4	2.13×10^4	5.39×10^3
$t_h t_h Z_h(+j)$	4.58×10^4	4.57×10^4	2.83×10^4	1.28×10^4	5.17×10^3
$h_b W_h(+j)$	6.77×10^3	98	< 10	< 10	< 10
$h_b Z_h(+j)$	4.31×10^3	66	< 10	< 10	< 10
Total number of background events:					1.58×10^6

IV. THE GRAPH NEURAL NETWORK MODEL

GNNs are a class of permutation-invariant machine learning methods that learn functions on graphs. The functions may be at the node, edge, or graph levels. The workings of GNNs are better understood by thinking of them as message-passing frameworks: GNN modules (layers) pass messages (information) between nodes to let each node learn the embedding that depends on its

attributes and its connections to other nodes in the graph. It can be mathematically expressed as [45]

$$\mathbf{x}_i^{(k)} = \gamma^{(k)}(\mathbf{x}_i^{(k-1)}, \bigoplus_{j \in \mathcal{N}(i)} \phi^{(k)}(\mathbf{x}_i^{(k-1)}, \mathbf{x}_j^{(k-1)}, \mathbf{e}_{j,i})), \quad (11)$$

where $\mathbf{x}_i^{(k)}$ denotes the embedding of the i th node at the k th layer of the GNN, $\mathcal{N}(i)$ is the neighborhood of i , $\mathbf{e}_{j,i}$ is the edge (link) from node j to node i , $\bigoplus$ denotes a differentiable, permutation invariant function such as sum, max and min, and γ and ϕ are differentiable functions like multilayer perceptrons. For graph-level tasks (like our classification task), the latent embeddings of each node are aggregated through a permutation-invariant function to give a single prediction. GNNs can be helpful for tasks such as particle tracking and particle flow [46], jet identification [47,48], and, in particular, event identification [49–51]. GNNs are especially effective since they operate much lower than boosted decision trees and deep neural networks (DNNs), allowing them to model features that might be difficult for us to construct or provide. (See Ref. [45] for a comprehensive review of GNNs.)

A. Constructing graphs from a collision event

Even though our event selection criteria focus on three AK4 jets (of which at least two are b jets) and one fatjet, a typical signal event contains more well-defined objects due to the high jet multiplicity of the B -quark decay.

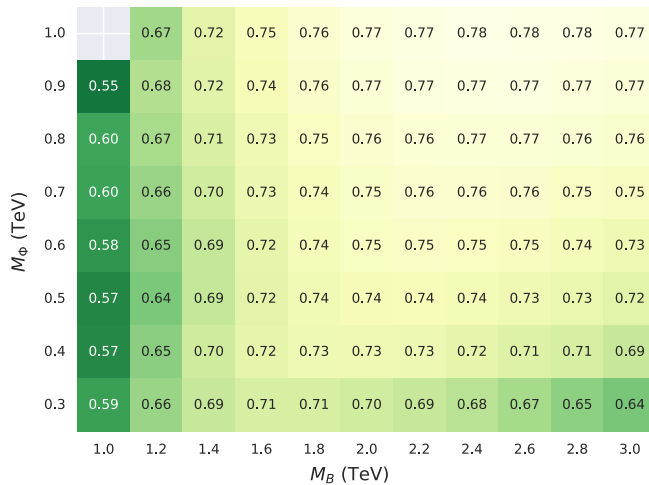


FIG. 5. Selection efficiency after $\mathfrak{C}_5$ at various parameter points on the $M_B - M_\Phi$ plane. On average, the efficiency exceeds 70%.

TABLE III. Node types and their respective input features. For the auxiliary fatjet node, τ denotes the N -subjettiness feature, where $i = 1, 2, 3, 4$, and τ_{ij} denotes the N -subjettiness ratios ($j = 1, 2, 3$).

Node-type	Input features
Shared attributes (SA) nodes	m, p_T, η, ϕ
Auxiliary jet	MeanSqDeltaR, b -tag, PTD
Auxiliary fatjet	Girth, $E, p_x, p_y, p_z, \tau_i, \tau_{ij}$
global	$H_T, \vec{E}_T$
CLS	...

In principle, once an event passes our selection criteria, all well-defined objects in that event can be used to classify the event using a GNN model. We can also make use of the event-level features. Therefore, to capture the information content in the events, we model them as heterogeneous graphs.

We represent the kinematic features of jets and fatjets with special abstract nodes—we call them “shared attributes” or SA nodes—essentially containing a set of attributes: η, ϕ, m , and p_T . By constructing an abstract common node type for these kinematic features, we enable the model to process these generic features in a unified way and learn functions to model them better. Each jet or fatjet in a selected event has one corresponding SA node. Since our event topology has two kinds of hadronic objects (jets and fatjets), we choose two kinds of auxiliary nodes to represent them. The auxiliary jet node contains features like b tag, MeanSqDeltaR (the mean of p_T^2 -weighted rms distance

between the constituent and the jet axis), and PTD (a measure of the p_T dispersion of the jet constituents). The auxiliary fatjet node has N -subjettiness ratios and girth (mean of the p_T -weighted distances of the constituents from the fatjet axis) along with their four-momenta as its attributes.

The graph also contains two additional nodes—global and CLS. The global node contains features of the whole event, such as H_T and $\vec{E}_T$. The CLS node extracts the embedding for the entire graph (i.e., event) as a vector. This vector is then fed as the input to a fully connected DNN that performs the signal vs background binary classification. Table III lists the attribute content of each node type we construct. For simplicity, we do not specify edge-level attributes and instead let the GNN module learn these from the node features.

Sequential graph construction: We sequentially construct the event graph—Fig. 6 illustrates the construction.

- (1) *GNN layer 1–2*: In the first step, the SA nodes are connected with each other (i.e., a clique graph) to enable global interactions among the kinematics of the reconstructed objects and to ensure that each subsequent connection has some information about other nodes present at the global scale. We then perform a round of message passing.
- (2) *GNN layer 2–3*: Each SA node is connected to its corresponding auxiliary jet or fatjet node in the second step. This allows for object-specific embedding learning, where the kinematic information gathered from the previous layer and the node-level input from the current one are used to construct

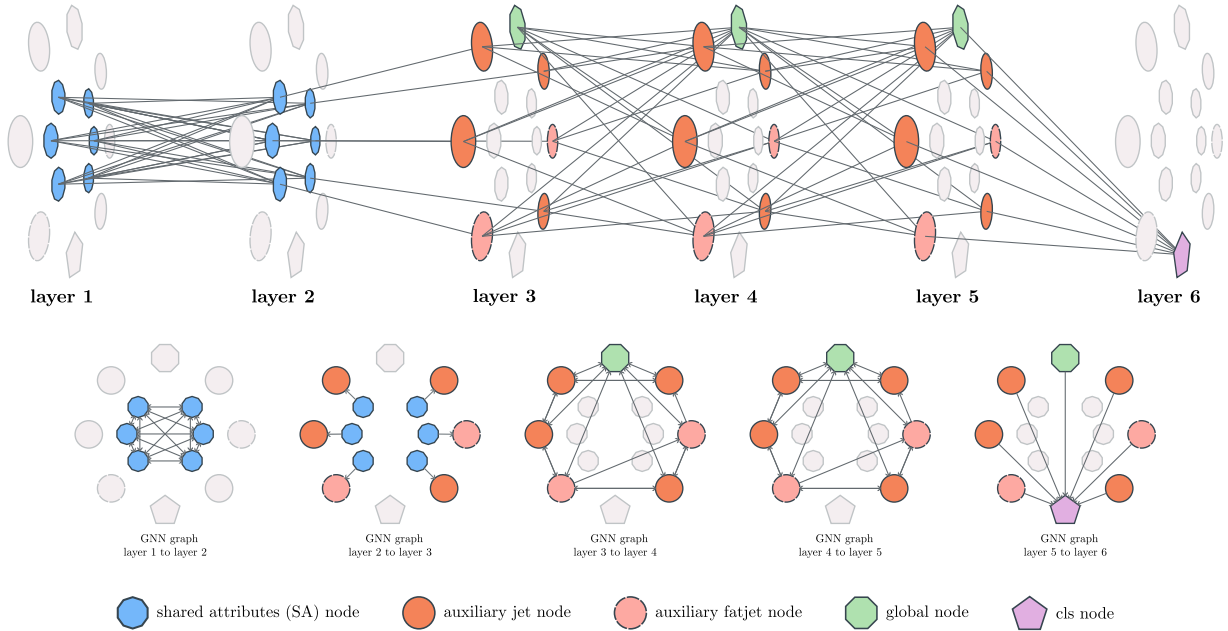


FIG. 6. *Constructing the GNN*: We show sequential construction of the graph along with flow of information (message passing) in the GNN model. We use GATv2Conv layer to implement the message passing operation in layer.

meaningful embeddings for each reconstructed object. A round of message passing is performed after the connection is made.

- (3) *GNN layer 3–5*: Jets and fatjets located nearby are likely related. For instance, a fatjet will be close to its constituent jets. Furthermore, independent jets with a common parent are also likely to be close. To incorporate this, we connect nodes to their nearest objects in the $\eta - \phi$ plane in the next step using the k nearest neighbors (k NN) algorithm. We also connect the `global` node of the jet and fatjet nodes using a bidirectional link. We do this for two reasons—we want to provide event-level information to each (fat) jet, and we also want to enable long-range interactions between distant jets, as this information transfer can occur through the `global` node. This choice is critical for large graphs, and similar strategies have been employed previously [52,53]. Once the connections are in place, we perform two rounds of message passing.
- (4) *GNN layer 5–6*: Once all message passings are performed, and each node has an appropriate embedding, we connect all the nodes to the `CLS` token/node to extract the graph-level class information and perform a round of message passing. This idea is commonly used in transformers in language-related tasks [54].

We primarily use the GATv2Conv [55] message-passing module to construct our GNN as it shows superior performance over other methods such as GCNConv [56] and SAGEConv [57]. This performance improvement could be attributed to the attention mechanism employed by the GNN. The GATv2Conv module incorporates the attention

mechanism to learn an embedding for a node by selectively attending to information obtained by its neighbors,

$$\mathbf{x}'_i = \sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{i,j} \Theta_t \mathbf{x}_j, \quad (12)$$

where the attention coefficients $\alpha_{i,j}$ are computed as

$$\alpha_{i,j} = \frac{\exp(\mathbf{a}^\top \text{LeakyReLU}(\Theta_s \mathbf{x}_i + \Theta_t \mathbf{x}_j))}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp(\mathbf{a}^\top \text{LeakyReLU}(\Theta_s \mathbf{x}_i + \Theta_t \mathbf{x}_k))}. \quad (13)$$

Here, subscripts s and t indicate the source and the target in the message passing framework, and $\Theta_{s,t}$ are their corresponding parameters. Using the constructed event embeddings from `CLS` node, we use a simple, fully connected deep neural network (DNN) to perform the final event classification.

B. Classifier description and hyperparameters tuning

Our classifier has two components: a GNN model (with six graph-convolution layers) to learn the event embedding and a DNN model (with five fully connected layers) for event classification using the embeddings. Figure 7 shows a representative diagram of the classification pipeline. We use BatchNorm [58] and Dropout [59] layers to regularize our model with dropout probability $p = 0.5$. We use the ADAMW optimizer [60] with a weight decay value of $\lambda = 10^{-4}$. We run a comprehensive hyperparameter search on the depth of the model (both GNN and DNN), the sizes of the various nodes of the different layers, the regularization parameters, the learning rate and weight decay, the

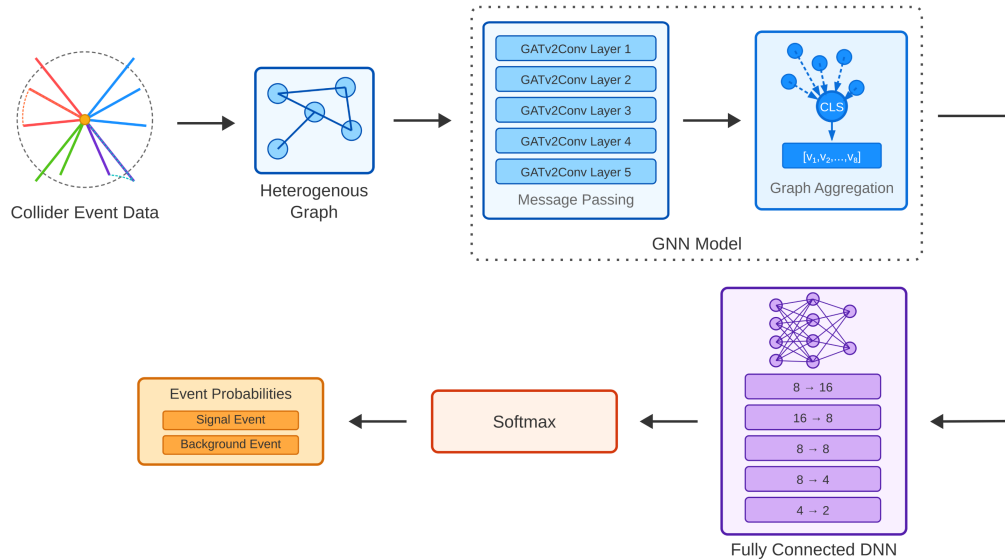


FIG. 7. Schematic of the event classification pipeline. We perform an event-level signal vs background classification on the learned event embeddings using a DNN model. We use a GNN model (Fig. 6) to learn the embeddings of the event.

batch size, the number of epochs of training, and the value of k for k NN for generating the graph. Our hyperparameters are optimized using the Bayesian search strategy on WEIGHTS AND BIASES [61].

C. Training strategy

We train the classifier with a two-step strategy:

- (1) *Pretraining*: We pretrain our model on all signal parameter points we consider.
- (2) *Fine-tuning*: After pretraining the classifier, we finetune our model on each mass point.

Our strategy is mainly motivated by the mounting evidence that large models trained on vast amounts of data generalize better than specific models [62–65]. Models trained on larger datasets are better able to capture the underlying properties.

Our signal is characterized by M_B and M_Φ . We treat the branching ratio of B to the new mode, i.e., $\text{BR}(B \rightarrow b\Phi)$, as a free parameter that controls the yield of the signal. As mentioned earlier, we pick eight benchmark values for M_Φ (from 300 GeV to 1 TeV, sampled at an interval of 100 GeV), and eleven values for M_B (1 TeV to 3 TeV, sampled at an interval of 200 GeV). Considering all combinations leaves us with 87 distinct parameter points (the parameter point $M_B = M_\Phi = 1$ TeV is kinematically forbidden) to pretrain the model on. The model is trained to perform binary classification at these parameter points.

Loss calculation: We train an unbiased signal vs background classifier in our pretraining stage; i.e., we set the total background weight equal to the total signal weight. For this, we weigh the background processes such that the total background weight is the proportionate sum of the rates (cross sections). Therefore, the weight of a particular background process is

$$w_{b_i} \propto \frac{\sigma_{b_i}}{N_{b_i}}, \quad (14)$$

where σ_{b_i} is the cross section of the background process i and N_{b_i} is the number of samples of background process i in the dataset. Similarly, we weigh the signal events. However, since the signal cross section varies across the parameter range we scan, we ensure that all points contribute equally to the final loss by setting the weights as

$$w_s \propto \frac{1}{N_s}, \quad (15)$$

where N_s is the number of signal samples for a parameter point. As we showed earlier [27], a bias-adjusted loss function performs better than the conventional cross-entropy loss. We find that such a loss function effectively balances the performance across all the signal parameter points while cutting out heavy background processes in the current case as well.

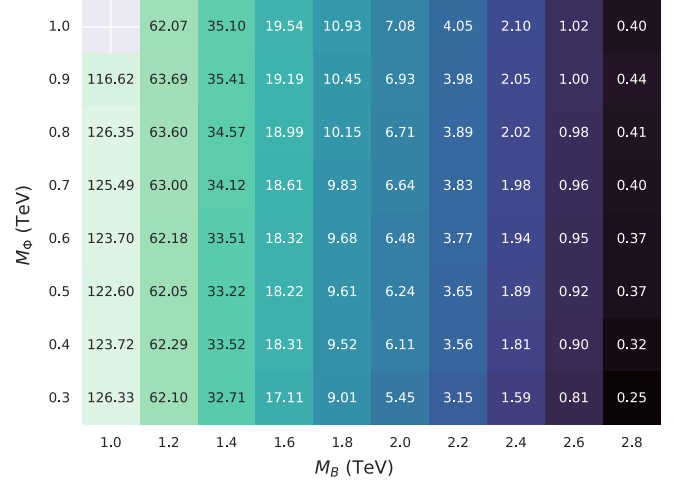


FIG. 8. Discovery sensitivity [defined in Eq. (16)] obtained from the GNN + DNN model for each parameter point on the $M_B - M_\Phi$ grid considered in Fig. 5.

Fine-tuning: While fine-tuning to a particular parameter point, we keep the GNN model frozen and retrain only the fully connected DNN. Each parameter point is fine-tuned independently, giving us 87 models in total. We scan the threshold values of the classifier response curve and select the best value that maximizes the discovery sensitivity [66], given as

$$\mathcal{Z}_D = \sqrt{2(N_S + N_B) \ln \left(\frac{N_S + N_B}{N_B} \right)} - 2N_S, \quad (16)$$

where N_S and N_B are the number of surviving signal and background events at the HL-LHC, respectively. We take the branching ratio in the new mode $\beta_{b\Phi} = 1$. The pretrained generic model performs well on most parameter points, as seen in Fig. 8. Fine-tuning shows a modest improvement in the final $\mathcal{Z}_D$ score, with noticeable improvements at the extremums of the parameter points.

V. PROSPECTS

Figure 9 shows the LHC reach obtained using the fine-tuned GNN model for two different projected luminosities, $\mathcal{L} = 3000 \text{ fb}^{-1}$ [Figs. (a) and (b) in Fig. 9] and 300 fb^{-1} [Figs. (c) and (d) in Fig. 9]. We show the 5σ (discovery) and 2σ ($\sim$ exclusion) contours on the $M_B - M_\Phi$ plane with different contours corresponding to different values of $\beta_{b\Phi}$. Since the number of signal events scales with the branching ratio as $\beta_{b\Phi}^2$, these contours can be easily interpreted for other scenarios (e.g., when the standard decay modes dominate or the B quark has additional decay modes) as well by scaling down the $\beta_{b\Phi} = 1.0$ line appropriately. We estimate the discovery contours with the $\mathcal{Z}_D$ score [Eq. (16)], and the exclusion limits with [66]

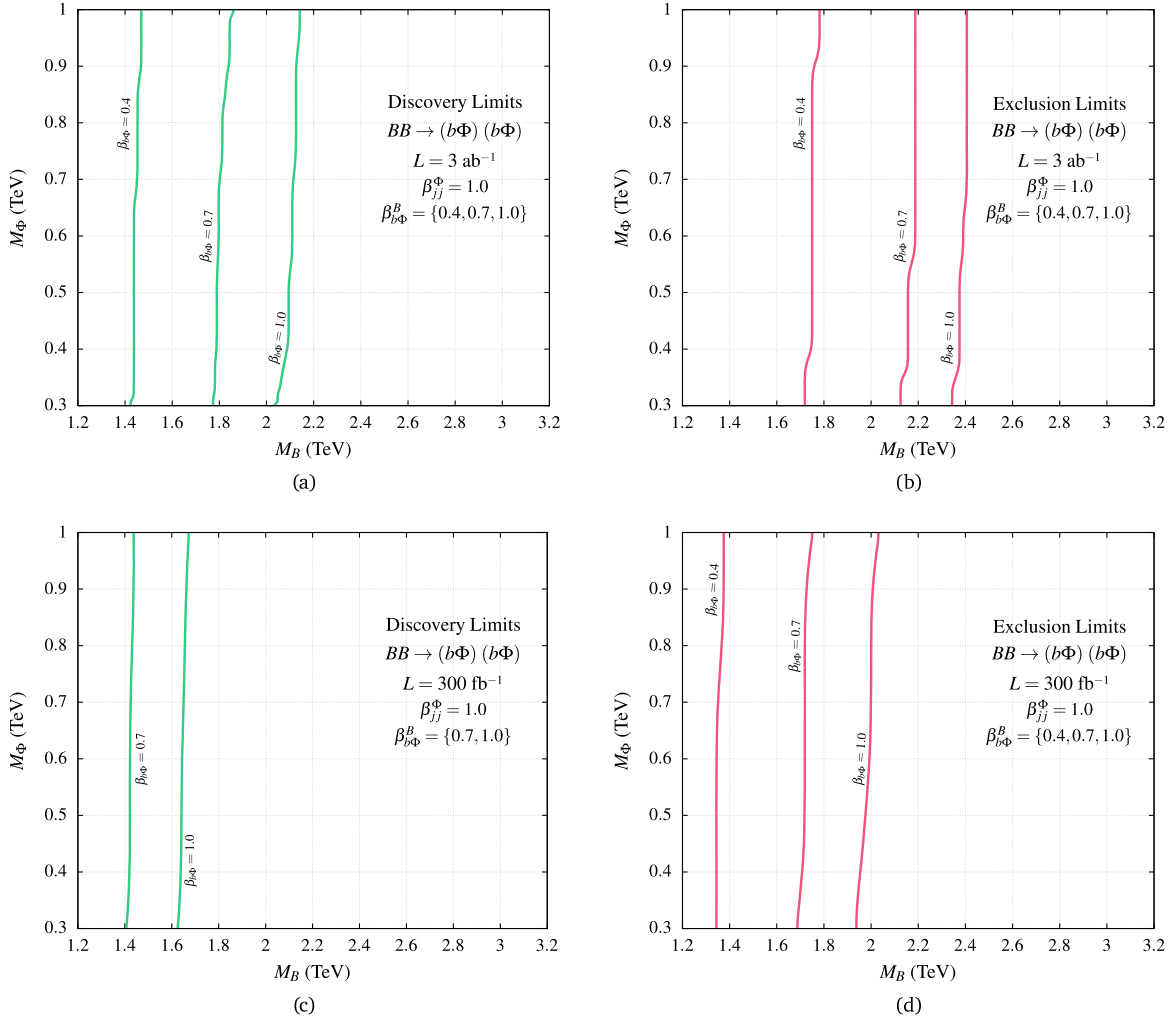


FIG. 9. LHC reach plots for $BB \rightarrow (b\Phi)(b\Phi)$ signal at the HL-LHC, i.e., for luminosity, $L = 3000 \text{ fb}^{-1}$ [(a) and (b)]. The left panel shows the $\mathcal{Z}_D = 5$ [discovery, Eq. (16)] contours and the right panel shows the $\mathcal{Z}_E = 2$ [exclusion, Eq. (17)] contours for three different values of branching in the $B \rightarrow b\Phi$ mode: $\beta_{b\Phi} = 40\%, 70\%, 100\%$. The boson, Φ , has a two-prong decay to either a pair of b -jets or gluon jets. We also show these plots for $L = 300 \text{ fb}^{-1}$ in (c) and (d).

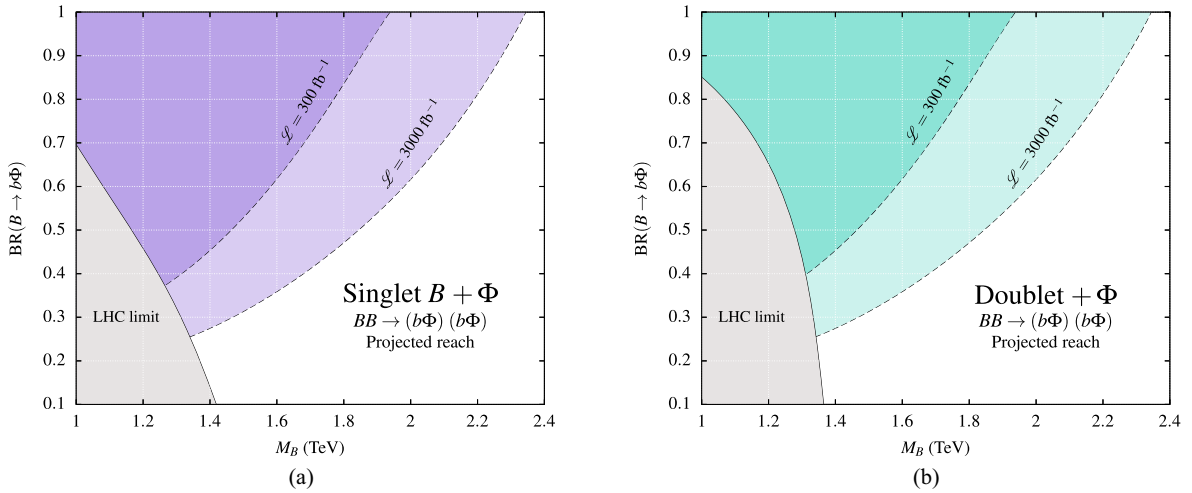


FIG. 10. Projected LHC reach for (a) the Singlet $B + \Phi$ model and (b) the Doublet $+ \Phi$ in the proposed channel along with the regions ruled out from the current LHC data (in gray) for $L = 300$ and 3000 fb^{-1} .

$$\mathcal{Z}_E = \sqrt{2N_S - 2N_B \ln\left(\frac{N_S + N_B}{N_B}\right)}, \quad (17)$$

where N_S , N_B are the number of surviving signal and background events at the LHC, respectively. In general, we see that with the complex GNN model, it is possible to attain a discovery significance score of 5σ at values $M_B > 2$ TeV. The reach is only slightly better towards the higher end of M_Φ considered due to the lower boost of Φ jet; overall, the GNN + DNN model efficiency follows the same trend as Fig. 5.

As expected, the LHC can exclude regions beyond its discovery reach. The projected exclusion limits apply to both the singlet $B + \Phi$ and doublet $+ \Phi$ models. Since the contours in Fig. 9 are insensitive to M_Φ , we reinterpret the exclusion contours and plot them on the $\text{BR}(B \rightarrow b\Phi)$ vs M_B plane in Fig. 10 by picking the maximum value M_B allowed at the 2σ level for each $\text{BR}(B \rightarrow b\Phi)$ value $\in [0.1, 1.0]$. The purple (turquoise) regions are excludable in the singlet $B + \Phi$ (doublet $+ \Phi$) model. The current limits (recast from the LHC data for $L = 139 \text{ fb}^{-1}$) are colored in gray. The gain is considerable, particularly at the HL-LHC. It also extends beyond the HL-LHC limits for the monoleptonic final states, as obtained in our previous paper [27]. There, the singlet $B + \Phi$ exclusion limit stood at $M_B = 1.8$ TeV, for $\text{BR}(B \rightarrow b\Phi) = 50\%$, corresponding to the maximal signal yield.

VI. CONCLUSIONS

In this paper, we studied the pair production of vectorlike bottom-type quarks (B), with each decaying to a b quark and a new gauge-singlet scalar or pseudoscalar boson Φ that dominantly decays to gg or $b\bar{b}$ pairs. Such final states ($2b + 4j$ or $6b$) are difficult to isolate due to the absence of any leptonic handles and an overwhelmingly large hadronic background at the LHC. We designed a sophisticated GNN model-based event classification pipeline to tag the resulting fully hadronic final states. By representing collision events as heterogeneous graphs containing jets, fatjets, and global event features, our model utilized the permutation invariance and message-passing strengths of GNNs. We also employed a novel two-step training strategy to train our event classifier: a pretraining over the entire set of benchmark parameter points before fine-tuning the model for classification at a particular parameter point. This strategy closely follows the one used to train large foundation models in various domains (see, e.g., Refs. [62–65]). Our estimation showed that, at the HL-LHC ($L = 3000 \text{ fb}^{-1}$), M_B up to 2.4 TeV could be excluded by a dedicated search in the joint $2b + 4j/6b$ channel for $\text{BR}(B \rightarrow b\Phi) \approx 100\%$. Even for a

moderate BR in the new mode, $\beta_{b\Phi} = 0.4$, the exclusion limit in M_B goes up to about 1.8 TeV. Our projections indicate that, in the future, the LHC sensitivity to the singlet and doublet VLQ scenarios (see Fig. 10) could be significantly enhanced, particularly in regions of parameter space currently unconstrained. Thus, a dedicated search for pair-produced VLQs decaying through the $4j + 2b/6b$ channel could uncover or constrain a wide range of new physics scenarios.

Though our analysis is thorough, a more rigorous analysis with real data, higher-order simulations, or better simulation of detector effects (e.g., with Geant4 [67]) could improve our results. Our estimations can be taken as conservative since, in the signal, we only included events where each pair-produced B quark decayed to a b and a Φ pair. Including hadronic final states from all pair-production signatures arising from standard decays of the B quark (i.e., $B \rightarrow \{tW, bZ, bh\}$) that satisfy our selection criteria could enhance the signal strength, as we showed in Ref. [27] while studying the prospects of an inclusive $tW + X$ signal arising from the decays of pair-produced B quarks. Moreover, as Refs. [68,69] showed, accounting for contaminating contributions from other production processes (e.g., single productions) could further enhance sensitivities. This is interesting because, beyond 2 TeV, the pair production process may not remain the leading production process of VLQs. Moreover, nonconventional single productions ($pp \rightarrow B\Phi$) also come into play in that regime. On the analysis front, transformer-based models have been shown to be effective at HEP tasks [70–75]. Though their performance on event-level features or heterogeneous reconstructed objects remains to be studied rigorously, using them in the analysis pipeline could improve the prospects further.

The Universal FeynRules Output [29] files used in our analysis are available at [76] under the name SingBplusPhi.

ACKNOWLEDGMENTS

T.M. acknowledges partial support from the SERB/ANRF, Government of India, through the Core Research Grant (CRG) No. CRG/2023/007031.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

- [1] Shrihari Gopalakrishna, Tanumoy Mandal, Subhadip Mitra, and Rakesh Tibrewala, LHC signatures of a vector-like b' , *Phys. Rev. D* **84**, 055001 (2011).
- [2] Shrihari Gopalakrishna, Tanumoy Mandal, Subhadip Mitra, and Grégory Moreau, LHC signatures of warped-space vectorlike quarks, *J. High Energy Phys.* **08** (2014) 079.
- [3] João M. Alves, G. C. Branco, A. L. Cherchiglia, C. C. Nishi, J. T. Penedo, Pedro M. F. Pereira, M. N. Rebelo, and J. I. Silva-Marcos, Vector-like singlet quarks: A roadmap, *Phys. Rep.* **1057**, 1 (2024).
- [4] Aram Hayrapetyan *et al.* (CMS Collaboration), Review of searches for vector-like quarks, vector-like leptons, and heavy neutral leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV at the CMS experiment, *Phys. Rep.* **1115**, 570 (2025).
- [5] Javi Serra, Beyond the minimal top partner decay, *J. High Energy Phys.* **09** (2015) 176.
- [6] A. Arhrib, R. Benbrik, S. J. D. King, B. Manaut, S. Moretti, and C. S. Un, Phenomenology of 2HDM with vectorlike quarks, *Phys. Rev. D* **97**, 095015 (2018).
- [7] Huayong Han, Li Huang, Teng Ma, Jing Shu, Tim M. P. Tait, and Yongcheng Wu, Six top messages of new physics at the LHC, *J. High Energy Phys.* **10** (2019) 008.
- [8] Nicolas Bizot, Giacomo Cacciapaglia, and Thomas Flacke, Common exotic decays of top partners, *J. High Energy Phys.* **06** (2018) 065.
- [9] Rachid Benbrik *et al.*, Signatures of vector-like top partners decaying into new neutral scalar or pseudoscalar bosons, *J. High Energy Phys.* **05** (2020) 028.
- [10] Giacomo Cacciapaglia, Gabriele Ferretti, Thomas Flacke, and Hugo Serôdio, Light scalars in composite Higgs models, *Front. Phys.* **7**, 22 (2019).
- [11] J. A. Aguilar-Saavedra, J. Alonso-González, L. Merlo, and J. M. No, Exotic vectorlike quark phenomenology in the minimal linear σ model, *Phys. Rev. D* **101**, 035015 (2020).
- [12] Giacomo Cacciapaglia, Thomas Flacke, Myeonghun Park, and Mengchao Zhang, Exotic decays of top partners: Mind the search gap, *Phys. Lett. B* **798**, 135015 (2019).
- [13] Daohan Wang, Lei Wu, and Mengchao Zhang, Hunting for top partner with a new signature at the LHC, *Phys. Rev. D* **103**, 115017 (2021).
- [14] Avik Banerjee *et al.*, Phenomenological aspects of composite Higgs scenarios: exotic scalars and vector-like quarks, *arXiv:2203.07270*.
- [15] Avik Banerjee, Diogo Buarque Franzosi, and Gabriele Ferretti, Modelling vector-like quarks in partial compositeness framework, *J. High Energy Phys.* **03** (2022) 200.
- [16] Daniel Elander, Ali Fatemiabhari, and Maurizio Piai, Toward minimal composite Higgs models from regular geometries in bottom-up holography, *Phys. Rev. D* **107**, 115021 (2023).
- [17] Roberto Franceschini, Physics beyond the standard model associated with the top quark, *Annu. Rev. Nucl. Part. Sci.* **73**, 397 (2023).
- [18] Avik Banerjee, Venugopal Ellajosyula, and Luca Panizzi, Heavy vector-like quarks decaying to exotic scalars: A case study with triplets, *J. High Energy Phys.* **01** (2024) 187.
- [19] Ed Bennett, Ho Hsiao, Jong-Wan Lee, Biagio Lucini, Axel Maas, Maurizio Piai, and Fabian Zierler, Singlets in gauge theories with fundamental matter, *Phys. Rev. D* **109**, 034504 (2024).
- [20] Avik Banerjee, Elin Bergeaas Kuutmann, Venugopal Ellajosyula, Rikard Enberg, Gabriele Ferretti, and Luca Panizzi, Vector-like quarks: Status and new directions at the LHC, *SciPost Phys. Core* **7**, 079 (2024).
- [21] Umar Sohail Qureshi, Alfredo Gurrola, Andres Flórez, and Cristian Rodriguez, Probing light scalars and vector-like quarks at the high-luminosity LHC, *Eur. Phys. J. C* **85**, 379 (2025).
- [22] Jeong Han Kim, Samuel D. Lane, Hye-Sung Lee, Ian M. Lewis, and Matthew Sullivan, Searching for dark photons with maverick top partners, *Phys. Rev. D* **101**, 035041 (2020).
- [23] Shivam Verma, Sanjoy Biswas, Anirban Chatterjee, and Joy Ganguly, Exploring maverick top partner decays at the LHC, *Phys. Rev. D* **107**, 115024 (2023).
- [24] Shivam Verma, Sanjoy Biswas, Tanumoy Mandal, and Subhadip Mitra, Machine learning tagged boosted dark photon: A signature of fermionic portal matter at the LHC, *arXiv:2410.06925*.
- [25] Akanksha Bhardwaj, Tanumoy Mandal, Subhadip Mitra, and Cyrin Neeraj, Roadmap to explore vectorlike quarks decaying to a new scalar or pseudoscalar, *Phys. Rev. D* **106**, 095014 (2022).
- [26] Akanksha Bhardwaj, Kartik Bhide, Tanumoy Mandal, Subhadip Mitra, and Cyrin Neeraj, Discovery prospects of a vectorlike top partner decaying to a singlet boson, *Phys. Rev. D* **106**, 075024 (2022).
- [27] Jai Bardhan, Tanumoy Mandal, Subhadip Mitra, and Cyrin Neeraj, Machine learning-enhanced search for a vectorlike singlet B quark decaying to a singlet scalar or pseudoscalar, *Phys. Rev. D* **107**, 115001 (2023).
- [28] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks, FeynRules 2.0—A complete toolbox for tree-level phenomenology, *Comput. Phys. Commun.* **185**, 2250 (2014).
- [29] Céline Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter, UFO—The 2.0, *Comput. Phys. Commun.* **183**, 1201 (2012).
- [30] 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.
- [31] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, An introduction to PYTHIA8.2, *Comput. Phys. Commun.* **191**, 159 (2015).
- [32] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [33] A. M. Sirunyan *et al.* (CMS Collaboration), Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, *J. Instrum.* **13**, P05011 (2018).

- [34] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [35] Albert M Sirunyan *et al.* (CMS Collaboration), Search for pair production of vectorlike quarks in the fully hadronic final state, *Phys. Rev. D* **100**, 072001 (2019).
- [36] Giovanni Balossini, Guido Montagna, Carlo Michel Carloni Calame, Mauro Moretti, Oreste Nicosini, Fulvio Piccinini, Michele Treccani, and Alessandro Vicini, Combination of electroweak and QCD corrections to single W production at the Fermilab Tevatron and the CERN LHC, *J. High Energy Phys.* **01** (2010) 013.
- [37] Stefano Catani, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini, Vector boson production at hadron colliders: A fully exclusive QCD calculation at next-to-next-to-leading order, *Phys. Rev. Lett.* **103**, 082001 (2009).
- [38] Claudio Muselli, Marco Bonvini, Stefano Forte, Simone Marzani, and Giovanni Ridolfi, Top quark pair production beyond NNLO, *J. High Energy Phys.* **08** (2015) 076.
- [39] Nikolaos Kidonakis, Theoretical results for electroweak-boson and single-top production, *Proc. Sci. DIS2015* (2015) 170 [arXiv:1506.04072].
- [40] John M. Campbell, R. Keith Ellis, and Ciaran Williams, Vector boson pair production at the LHC, *J. High Energy Phys.* **07** (2011) 018.
- [41] M. Cepeda *et al.*, Report from working group 2: Higgs physics at the HL-LHC and HE-LHC, *CERN Yellow Rep. Monogr.* **7**, 221 (2019).
- [42] Anna Kulesza, Leszek Motyka, Daniel Schwartländer, Tomasz Stebel, and Vincent Theeuwes, Associated production of a top quark pair with a heavy electroweak gauge boson at NLO + NNLL accuracy, *Eur. Phys. J. C* **79**, 249 (2019).
- [43] D. de Florian *et al.* (LHC Higgs Cross Section Working Group), Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, *10.23731/CYRM-2017-002* (2016).
- [44] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions, *J. High Energy Phys.* **01** (2007) 013.
- [45] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant, Graph neural networks in particle physics, *Mach. Learn. Sci. Technol.* **2**, 021001 (2021).
- [46] Jan Kieseler, Object condensation: One-stage grid-free multi-object reconstruction in physics detectors, graph and image data, *Eur. Phys. J. C* **80**, 886 (2020).
- [47] Huilin Qu and Loukas Gouskos, ParticleNet: Jet tagging via particle clouds, *Phys. Rev. D* **101**, 056019 (2020).
- [48] Vinicius Mikuni and Florencia Canelli, ABCNet: An attention-based method for particle tagging, *Eur. Phys. J. Plus* **135**, 463 (2020).
- [49] Murat Abdughani, Daohan Wang, Lei Wu, Jin Min Yang, and Jun Zhao, Probing the triple Higgs boson coupling with machine learning at the LHC, *Phys. Rev. D* **104**, 056003 (2021).
- [50] Nicholas Choma, Federico Monti, Lisa Gerhardt, Tomasz Palczewski, Zahra Ronaghi, Prabhat, Wahid Bhimji, Michael M. Bronstein, Spencer R. Klein, and Joan Bruna (IceCube Collaboration), Graph neural networks for IceCube signal classification, arXiv:1809.06166.
- [51] R. Abbasi *et al.*, Graph neural networks for low-energy event classification & reconstruction in IceCube, *J. Instrum.* **17**, P11003 (2022).
- [52] Md. Shamim Hussain, Mohammed J. Zaki, and Dharmashankar Subramanian, Edge-augmented graph transformers: Global self-attention is enough for graphs, arXiv:2108.03348.
- [53] EunJeong Hwang, Veronika Thost, and Tengfei Ma Shib Sankar Dasgupta, An analysis of virtual nodes in graph neural networks for link prediction, *Learning on Graphs* (2022).
- [54] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805.
- [55] Shaked Brody, Uri Alon, and Eran Yahav, How attentive are graph attention networks?, arXiv:2105.14491.
- [56] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio, Graph attention networks, arXiv:1710.10903.
- [57] Will Hamilton, Zhitao Ying, and Jure Leskovec, Inductive representation learning on large graphs, *Adv. Neural Inf. Process. Syst.* **30** (2017).
- [58] Sergey Ioffe and Christian Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv:1502.03167.
- [59] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* **15**, 1929 (2014).
- [60] Ilya Loshchilov and Frank Hutter, Decoupled weight decay regularization, arXiv:1711.05101.
- [61] Lukas Biewald, Experiment tracking with weights and biases (2020), software available from wandb.com.
- [62] Congqiao Li, Antonios Agapitos, Jovin Drews, Javier Duarte, Dawei Fu, Leyun Gao, Raghav Kansal, Gregor Kasieczka, Louis Moureaux, Huilin Qu, Cristina Mantilla Suarez, and Qiang Li, Accelerating resonance searches via signature-oriented pre-training, arXiv:2405.12972.
- [63] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample, Llama: Open and efficient foundation language models, arXiv:2302.13971.
- [64] Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lillicrap, Jean-baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser *et al.*, Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, arXiv:2403.05530.
- [65] Tom B. Brown *et al.*, Language models are few-shot learners, *Adv. Neural Inf. Process. Syst.* **33**, 1877 (2020).
- [66] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71**, 1554 (2011); **73**, 2501(E) (2013).
- [67] S. Agostinelli *et al.* (GEANT4 Collaboration), GEANT4—A simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect. A* **506**, 250 (2003).

- [68] Tanumoy Mandal and Subhadip Mitra, Probing color octet electrons at the LHC, *Phys. Rev. D* **87**, 095008 (2013).
- [69] Tanumoy Mandal, Subhadip Mitra, and Satyajit Seth, Single productions of colored particles at the LHC: An example with scalar leptoquarks, *J. High Energy Phys.* **07** (2015) 028.
- [70] Huilin Qu, Congqiao Li, and Sitian Qian, Particle transformer for jet tagging, [arXiv:2202.03772](https://arxiv.org/abs/2202.03772).
- [71] Alexander Bogatskiy, Timothy Hoffman, David W. Miller, Jan T. Offermann, and Xiaoyang Liu, Explainable equivariant neural networks for particle physics: PELICAN, *J. High Energy Phys.* **03** (2024) 113.
- [72] Thorben Finke, Michael Krämer, Alexander Mück, and Jan Tönshoff, Learning the language of QCD jets with transformers, *J. High Energy Phys.* **06** (2023) 184.
- [73] Tobias Golling, Lukas Heinrich, Michael Kagan, Samuel Klein, Matthew Leigh, Margarita Osadchy, and John Andrew Raine, Masked particle modeling on sets: towards self-supervised high energy physics foundation models, *Mach. Learn. Sci. Tech.* **5**, 035074 (2024).
- [74] Yifan Wu, Kun Wang, Congqiao Li, Huilin Qu, and Jingya Zhu, Jet tagging with more-interaction particle transformer, *Chin. Phys. C* **49**, 013110 (2025).
- [75] Jai Bardhan, Radhikesh Agrawal, Abhiram Tilak, Cyryn Neeraj, and Subhadip Mitra, HEP-JEPA: A foundation model for collider physics using joint embedding predictive architecture, [arXiv:2502.03933](https://arxiv.org/abs/2502.03933).
- [76] <https://github.com/rsrchtsm/vectorlikequarks>