

Phenomenology of vector-like leptons with Deep Learning at the Large Hadron Collider

Felipe F. Freitas,^a João Gonçalves,^a António P. Morais^a and Roman Pasechnik^b

^a*Departamento de Física da Universidade de Aveiro and Centre for Research and Development in Mathematics and Applications (CIDMA), Campus de Santiago, 3810-183 Aveiro, Portugal*

^b*Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, SE 223-62 Lund, Sweden*

E-mail: felipefreitas@ua.pt, jpedropino@ua.pt, aapmorais@ua.pt, Roman.Pasechnik@thep.lu.se

ABSTRACT: In this paper, a model inspired by Grand Unification principles featuring three generations of vector-like fermions, new Higgs doublets and a rich neutrino sector at the low scale is presented. Using the state-of-the-art Deep Learning techniques we perform the first phenomenological analysis of this model focusing on the study of new charged vector-like leptons (VLLs) and their possible signatures at CERN's Large Hadron Collider (LHC). In our numerical analysis we consider signal events for vector-boson fusion and VLL pair production topologies, both involving a final state containing a pair of charged leptons of different flavor and two sterile neutrinos that provide a missing energy. We also consider the case of VLL single production where, in addition to a pair of sterile neutrinos, the final state contains only one charged lepton. We propose a novel method to identify missing transverse energy vectors by comparing the detector response with Monte-Carlo simulated data. All calculated observables are provided as data sets for Deep Learning analysis, where a neural network is constructed, based on results obtained via an evolutive algorithm, whose objective is to maximise either the accuracy metric or the Asimov significance for different masses of the VLL. Taking into account the effect of the three analysed topologies, we have found that the combined significance for the observation of new VLLs at the high-luminosity LHC can range from 5.7σ , for a mass of 1.25 TeV, all the way up to 28σ if the VLL mass is 200 GeV. We have also shown that by the end of the LHC Run-III a 200 GeV VLL can be excluded with a confidence of 8.8 standard deviations. The results obtained show that our model can be probed well before the end of the LHC operations and, in particular, providing important phenomenological information to constrain the energy scale at which new gauge symmetries emergent from the considered Grand Unification picture can be manifest.

KEYWORDS: Beyond Standard Model, Quark Masses and SM Parameters, GUT, Gauge Symmetry

ARXIV EPRINT: [2010.01307](https://arxiv.org/abs/2010.01307)

Contents

1	Introduction	1
2	Theoretical background	3
2.1	A motivation	3
2.2	High-energy limit	4
2.3	Low-energy effective limit	6
2.3.1	Physically viable benchmark scenarios for masses	10
2.3.2	Physically viable benchmark scenarios for couplings	12
3	Searching for vector-like leptons at the LHC	13
3.1	Methodology: Deep Learning models and dataset	18
3.2	Results	22
4	Conclusions	40
A	Low-scale SHUT model Lagrangian and Feynman rules	42
B	Kinematic and angular variable for all topologies with detector effects	45
C	Neural Network models for different masses	49
D	Hierarchical clustering for the neural networks	55

1 Introduction

The ultimate goal of any scientific endeavour is to uncover the mysteries of the universe and the world around us and, so far, the best model that we devised to describe all the matter that surrounds us at the most fundamental level is modestly called the Standard Model (SM). The SM is a particle physics model based upon modern quantum field theory (QFT) framework whose predictions and results have matched the stringiest of tests and predictions [1–7]. However, there are clear indications that something is missing, from the fact that neutrinos have mass, as confirmed by the neutrino oscillation phenomena [8], and that it does not take into account the existence of dark matter (DM) [9]. Besides such experimental evidences, there are also theoretical motivations, as, e.g. the origin of the family replication found in nature, the fermion masses and mixing hierarchies and the origin of the SM gauge structure, where a consensual understanding is still lacking.

So, we can notice that there are certain deficiencies in our understanding of fundamental particle physics which leave us with the obvious question: what is missing and how to fix it? Well, so far, the most exotic theories have been put into the forefront, ranging

from models where extra spacetime dimensions exist [10] to models with a new symmetry between bosons and fermions known as supersymmetry (SUSY) [11]. While somewhat separate, these theories have a common underlying idea. The SM is an effective description of a more fundamental theory and is only valid up to a certain energy scale beyond which New Physics (NP) is needed. Therefore, the problems of the SM all result from our lack of understanding of what such theory really is and at which energy scale it becomes manifest.

High-scale theories like the string theory and SUSY, despite their mathematical complexity, provide a solid theoretical framework from which one can build upon in order to e.g. obtain NP models well motivated by the first principles. However, the amount of new states and model parameters emerging from such scenarios can be overwhelming. One possibility is to use a brute force method to analyse each combination of parameters and select the most promising ones or, alternatively, follow a smarter approach based upon Deep Learning (DL) techniques and optimization algorithms to find the best parameter space. Furthermore, one might have to overcome the typical challenges inherent to collider phenomenology, where the impact of background events can easily bury possible signal events preventing potential NP signatures from becoming observable. A better approach to handle such problem is the use of multivariate analysis to identify possible deviations from expected events which can be caused by NP. These deviations can be further amplified by combining multiple distributions into multidimensional distributions [12]. This state of affairs, the need to quickly identify subtle effects in multidimensional distributions of information, clearly calls for artificial intelligence methods. Particularly the use of Machine Learning and DL techniques [13].

With this being said, in this paper we revisit the key properties of a Grand Unified model recently introduced in [14–17] which attempts to unify all matter and fundamental interactions in a framework inspired by a flavour enhanced E_6 symmetry. Among the key features one can highlight a possible explanation for the fermion mass and mixing hierarchies observed in nature, as well as predicting NP states such as vector-like fermions, additional scalars doublets and a rich neutrino sector, well motivated by the model structure and the unification picture.

The goal of this article is to construct and study the low-energy limit of such a framework which offers interesting phenomenological implications for future explorations at particle colliders. In particular, we focus on the phenomenological study of vector-like leptons (VLLs) and propose potential smoking-gun signatures as direct search channels to probe our model both at the LHC Run-III as well as its high-luminosity upgrade. Furthermore, the techniques that we develop are rather generic and can be used well beyond the scope of the model under consideration. The numerical analysis will be performed using standard Monte Carlo tools, where the final step of our analysis consists in applying the DL techniques for statistical significance studies.

This paper is organized as follows. First, in section 2, we discuss the model structure. Here, we briefly review its basic properties both at the unification scale as well as its low energy (SM-like) limit, motivating the parameter choices used in the numerical analysis. The latter represents the main focus of this work which is performed in section 3 where a detailed description of the methods employed in our analysis and the results obtained is given. In section 4 we conclude and discuss future work and research directions.

2 Theoretical background

In this section, we introduce the model that is being explored in this article. We divide it into three main parts. In section 2.1 we motivate and introduce the concept of Grand Unified Theories (GUTs) our model is based upon. In particular, we consider an attractive low-scale picture where the presence of new vector-like fermions is well motivated alongside with three Higgs doublets. Then, in section 2.2 we make a short overview of the key properties of the unified framework. While the main purpose of this article is to study the phenomenological implications at the LHC and in particular the potential observability of VLLs, it is important to explain the origin of such NP and which parameter choices are relevant and well motivated. We then finalise in section 2.3 with an effective low-energy description by providing the full Lagrangian density, the particle masses as well as give a brief discussion of the benchmark scenarios that we use in our numerical analysis.

2.1 A motivation

One of the most attractive features of SUSY is an elegant solution to the well known hierarchy problem. Among the key predictions, every known particle in nature should have a SUSY partner with the same mass. However, none of the current or previous experiments have ever observed the existence of such particles. This means that SUSY cannot be an exact symmetry, at least, at phenomenologically relevant scales and should be broken in such a way to generate a larger mass contributions to the superpartners of the SM particles. The actual mass scale of such particles is not known, but the current lack of observation at the LHC [18–23] indicates that SUSY breaking should occur well above the electroweak (EW) scale. However, this by no means excludes SUSY as a well motivated formalism to describe realistic theories. This is the case of the model designed in [14–17] that we analyse in this article. While SUSY does not necessarily manifest at low scale and the effective theory can be treated as a standard non-SUSY model, its high scale limit is indeed supersymmetric with remarkable implications.

Such model can be thought of as a class of GUTs that can potentially emerge either from a single gauge group, as e.g. E_8 , or rather a semi-simple product group. We will denote this yet to be defined unifying force as G_U . One of the key properties of this framework is that flavour is promoted to a gauge symmetry that is part of G_U and treated in the same footing as conventional gauge interactions. The model aims at addressing various issues of the SM delving into fundamental questions such as the origin of gauge interactions and the origin of mass hierarchies for the different matter particles, which is typically known as the flavour problem. As a byproduct of this unification picture NP in the form of vector-like fermions (VLFs) may be manifest at the TeV scale. The emergence of light VLFs from other GUT models had previously been proposed in [24–27] where one of the key advantages of the presence of their leptonic counterparts is e.g. the possibility for explaining the muon anomalies [25, 28, 29]. In this article we will pay special attention to this sector since a potential discovery of VLLs at the LHC can offer relevant phenomenological probes of the high-energy theory and hints of the unification picture.

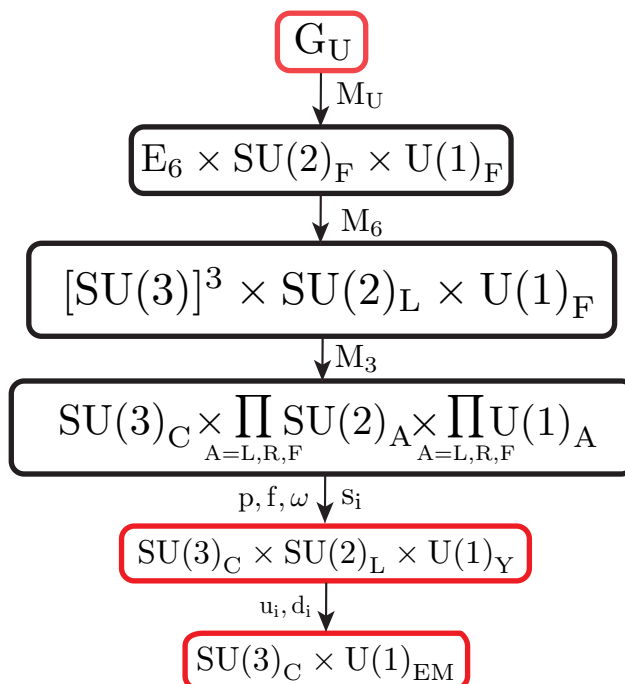


Figure 1. Symmetry breaking scheme from the original class of G_U gauge symmetries down to the strong and electromagnetic gauge groups ($SU(3)_C \times U(1)_{EM}$). The various terms between the different boxes (M_6 , M_3 , etc) represent the distinct mass scales involved in this scheme according to the discussion in [17]. M_U encodes unknown details G_U symmetry breaking.

2.2 High-energy limit

Here, we present the high-energy scale formulation of the model under consideration with focus on the main properties needed for a basic understanding relevant for our numerical analysis. A more detailed description can be found in [14–17, 30], where we highlight [17] as the most recent and complete reference.

The main idea of a GUT model is to embed all SM-gauge interactions, i.e. $SU(3)_C \times SU(2)_L \times U(1)_Y$, into a larger group. As already stated, an interesting possibility resides on the $E_6 \times SU(2)_F \times U(1)_F$ symmetry. It has been presented as a GUT candidate in [16, 17] where the exact connection and nature of G_U has not yet been fully established. However, the main low-scale properties have been thoroughly described and serve as a motivation to the current work.

Our model was engineered to address some of the main concerns one encounters in the SM. It proposes a first principles explanation for a common origin of the strong and EW interactions as well as the flavour structure observed in nature. The Higgs and matter fields are unified into a single superfield equipping both the scalar and the fermion sectors with the same flavour structures. This results in a rather reduced freedom in the Yukawa interactions allowing only for two free parameters, \mathcal{Y}_1 and \mathcal{Y}_2 , which will provide the dominant contributions to three generations of exotic vector-like quarks (VLQs) as well as the third and second generation SM-like quark masses. All remaining fermions, including the first-generation quarks and charged leptons, have their masses radiatively generated

making them naturally lighter. The CKM mixing is also emergent in this framework provided that there are, at least, three Higgs doublets developing vacuum expectation values (VEVs). Given its rather unique properties, this model has been named as SUSY Higgs-matter Unified Trinification or SHUT for short. Note that the trinification group emerges as a subgroup of E_6 which we consider below the M_6 scale in figure 1.

As stated before, the starting point is the E_6 gauge symmetry, where the strong and electroweak interactions get unified, extended with a family $SU(2)_F \times U(1)_F$ symmetry. From this point on, the sequence of steps by which we obtain the SM gauge group is schematically illustrated in figure 1. The SM particle content and all NP emergent at low-energy scales correspond to the states that remain light after the various breaking stages. Only second- and third-generation SM-like quarks and all three VLQ masses are tree-level generated, with their relative sizes controlled by the only two Yukawa couplings in the theory, which are of SUSY origin. To see this, let us consider the theory after the breaking step denoted by M_3 in figure 1 whose superpotential reads [17]

$$\begin{aligned}
 W = & \mathcal{Y}_1 \varepsilon_{ij} \left(\chi^i \mathbf{q}_L^3 \mathbf{q}_R^j + \ell_R^i \mathbf{D}_L^3 \mathbf{q}_R^j + \ell_L^i \mathbf{q}_L^3 \mathbf{D}_R^j + \phi^i \mathbf{D}_L^3 \mathbf{D}_R^j \right) \\
 & - \mathcal{Y}_2 \varepsilon_{ij} \left(\chi^i \mathbf{q}_L^j \mathbf{q}_R^3 + \ell_R^i \mathbf{D}_L^j \mathbf{q}_R^3 + \ell_L^i \mathbf{q}_L^j \mathbf{D}_R^3 + \phi^i \mathbf{D}_L^j \mathbf{D}_R^3 \right) \\
 & + \mathcal{Y}_2 \varepsilon_{ij} \left(\chi^3 \mathbf{q}_L^i \mathbf{q}_R^j + \ell_R^3 \mathbf{D}_L^i \mathbf{q}_R^j + \ell_L^3 \mathbf{q}_L^i \mathbf{D}_R^j + \phi^3 \mathbf{D}_L^i \mathbf{D}_R^j \right),
 \end{aligned}
 \tag{2.1}$$

where ε_{ij} is the two-dimensional Levi-Civita symbol in the generations' (or flavour, in what follows) space, \mathcal{Y}_1 and \mathcal{Y}_2 are the Yukawa couplings, and L/R denotes $SU(2)_{L/R}$ doublet superfields. Note that while χ is a $SU(2)_L \times SU(2)_R$ bi-doublet where the light Higgs sector resides, ϕ is a singlet carrying only family symmetry charges, typically dubbed as flavon. Despite some allowed mixing after symmetries are sequentially broken, the left- and right-handed leptons are essentially embedded in ℓ_L and ℓ_R respectively, whereas the SM-like quark sector belongs to both \mathbf{q}_L and \mathbf{q}_R . Note also that this model addresses neutrino masses due to the existence of six right-handed states, three in ℓ_R and three in ϕ . Last but not least, new down-type $SU(2)_L$ singlet VLQs and $SU(2)_L$ doublet VLLs are predicted in the SHUT model emerging from the fermionic components of $\mathbf{D}_{L,R}$ and χ , respectively. In this work, we will study the collider phenomenology of the latter and discuss possible implications for the high scale picture. All exotic scalars are assumed to be decoupled at the soft SUSY breaking scale beyond the reach of the LHC.

As we have mentioned above, one of the features of the SHUT model is that only second and third generation chiral quarks as well as the three VLQ generations are allowed to obtain masses at tree-level. For a better understanding of this statement let us inspect the superpotential in eq. (2.1). First, after the second last breaking stage in figure 1 all six neutral scalars in $\tilde{\phi}^i$ and $\tilde{\ell}_R^i$ develop VEVs which are denoted by p , f , ω and s_i (see [17] for details). We immediately see that mass terms for VLQs are generated from $\langle \tilde{\phi} \rangle \mathbf{D}_L \mathbf{D}_R$

and $\langle \tilde{\ell}_R \rangle \mathbf{D}_L \mathbf{q}_R$ type of terms resulting in [17]

$$\begin{aligned}
 m_{D/S}^2 &\simeq \frac{1}{2}(f^2 + p^2)\mathcal{Y}_2^2, & m_{S/D}^2 &\simeq \frac{\omega^2(f^2 + p^2 + \omega^2)}{2(f^2 + \omega^2)}\mathcal{Y}_2^2, \\
 m_B^2 &\simeq \frac{1}{2}(f^2 + \omega^2)\mathcal{Y}_1^2 + \frac{f^2 p^2}{2(f^2 + \omega^2)}\mathcal{Y}_2^2,
 \end{aligned}
 \tag{2.2}$$

where, for simplicity, we have ignored the subdominant effect of the s_i VEVs and where we adopt a notation such that the lightest VLQ is the D-quark. Along the same lines, SM-like quark masses are generated from $\langle \tilde{\chi} \rangle \mathbf{q}_L \mathbf{q}_R$ type of terms where, even for a generic setting with all six EW doublets in $\tilde{\chi}$ developing nonzero VEVs, the up and down-quark masses are always zero. Furthermore, it was shown in [16, 17] that a *proto-realistic* description of the CKM matrix requires a minimum of three light Higgs doublet VEVs where the quark masses become

$$m_u = 0, \quad m_c^2 = \frac{1}{2}\mathcal{Y}_2^2 (u_1^2 + u_2^2), \quad m_t^2 = \frac{1}{2}\mathcal{Y}_1^2 (u_1^2 + u_2^2)
 \tag{2.3}$$

and

$$m_d = 0, \quad m_s^2 = \frac{1}{2}\mathcal{Y}_2^2 \frac{d_2^2 p^2}{(f^2 + p^2 + \omega^2)}, \quad m_b^2 = \frac{1}{2}\mathcal{Y}_1^2 d_2^2,
 \tag{2.4}$$

with u_i and d_i being i-th family EW-symmetry breaking VEVs from Higgs doublets coupling to up-type and down-type quarks, respectively. If we consider, for simplicity, that $p \approx f \approx \omega$, we obtain the following ratios

$$\frac{\mathcal{Y}_1}{\mathcal{Y}_2} = \frac{m_t}{m_c} \approx \frac{m_b}{m_s} \approx \frac{m_B}{m_{D,S}} \sim \mathcal{O}(100),
 \tag{2.5}$$

implying also the presence of up to two generations of VLQs at the reach of the LHC if ω and f are around 100 TeV. This relation fixes the size of \mathcal{Y}_1 and \mathcal{Y}_2 and implies that mass ratios in the VLQ sector are the same as the ones found among their chiral counterparts. Note that in this work we will only study the VLL sector and leave a detailed numerical study of VLQs for a future work.

For the case of both SM-like leptons as well as VLLs, there are no allowed terms of the form $\langle \tilde{\chi} \rangle \ell_L \ell_R$ and $\langle \tilde{\phi} \rangle \chi \chi$, respectively, which means that, at tree-level, their masses are zero just like the first-generation quarks. However, below the second to last symmetry breaking stage in figure 1, such type of operators become allowed which means that they can be radiatively generated via loops with internal heavy scalar and fermion propagators.

2.3 Low-energy effective limit

While a direct probe for the high energy limit of the SHUT model at, or above, the $\omega - f - p$ scales is far beyond the reach of the LHC, exploring the corresponding NP signatures at the TeV-scale can offer us solid indications about the structure of the model at higher scales. Furthermore, such an analysis will provide an important piece of information about the low-scale properties of the model, which, although not explored in this work, can become relevant for matching of the low and the high scale regimes of the theory.

Field	SU(3) _C	SU(2) _L	U(1) _Y	# of generations
Q_L	3	2	1/3	3
L	1	2	-1	3
d_R	3	1	-2/3	3
u_R	3	1	4/3	3
e_R	1	1	-2	3

Table 1. SM-like sector for the fermions and quarks.

Field	SU(3) _C	SU(2) _L	U(1) _Y	# of generations
$E_{L,R}$	1	2	-1	3
$D_{L,R}$	3	1	-2/3	2
ν_R	1	1	0	6

Table 2. Beyond-the-SM sector for the fermions and quarks.

Field	SU(3) _C	SU(2) _L	U(1) _Y	# of generations
ϕ	1	2	1	3

Table 3. Scalar sector.

We consider in this section a possible low-energy scale limit of the SHUT model whose gauge symmetry is given in the second to last box of figure 1. All the quantum numbers for the gauge groups are shown in tables 1, 2 and 3 where the SU(2)_L doublets are defined as follows,

$$Q_L^i = \begin{bmatrix} u_L \\ d_L \end{bmatrix}^i \quad L^i = \begin{bmatrix} \nu_{eL} \\ e_L \end{bmatrix}^i \quad E_{L,R}^i = \begin{bmatrix} \nu'_{eL,R} \\ e'_{L,R} \end{bmatrix}^i, \quad (2.6)$$

with Q_L^i denoting the fermionic components of the $\mathbf{q}_L^{i=1,2,3}$ superfields, whereas L^i are the lepton doublet components of ℓ_L^i , and $E_{L,R}^i$ belong to the χ^i bi-doublet superfields.

Let us now describe the low-scale version of the SHUT model, step by step. The gauge boson's quantum numbers are not shown¹ since they are identical to the SM. On the other hand, the matter sector can be subdivided into two sub-sectors. The first, shown in table 1, represents the SM-like fermions from where ordinary matter emerges. The second sector, shown in table 2, is where NP appears including three new VLL generations, $E_{L,R}$, and two light VLQ generations which we denote as $D_{L,R}$. The Beyond-the-SM (BSM) sector also offers a rich neutrino content including six left-handed states originating from the E_L and E_R SU(2)_L doublets and six right-handed SM-singlet Majorana neutrinos which we denote as ν_R . Recall that the latter are embedded in three ℓ_R^i SU(2)_R-doublets and three

¹In fact, the model does allow for extra vector bosons, however those only become relevant at higher energy scales that are not particularly important for our discussion here.

ϕ^i flavons as stated above. Note that the lightest of the right-handed neutrinos, which we cast as ν_{BSM} in the remainder of this article, can be sterile enough to provide a good DM candidate [31]. While we do not perform DM studies in the current work, we will consider this scenario in our numerical analysis by setting its mass in the keV-MeV range and the mixing to the SM-like neutrinos to zero. In such a scenario ν_{BSM} escapes the detector and is treated as missing energy (MET). While the scalar sector also offers NP we will not further study it in this paper leaving any further details for a future work.

We can now introduce the relevant interaction terms for our analysis. We start with the low-scale Yukawa Lagrangian that reads as

$$\begin{aligned} \mathcal{L}_y = & (Y^a)_{ij} (\bar{Q}_L)^i (D_R)^j \phi_a + (\Gamma^a)_{ij} (\bar{Q}_L)^i (d_R)^j \phi_a + (\Delta^a)_{ij} (\bar{Q}_L)^i (u_R)^j \tilde{\phi}_a + \\ & + (\Theta^a)_{ij} (\bar{E}_L)^i (e_R)^j \phi_a + (\Upsilon^a)_{ij} (\bar{E}_L)^i (\nu_R)^j \tilde{\phi}_a + (\Sigma^a)_{ij} (\bar{L})^i (\nu_R)^j \tilde{\phi}_a + \\ & + (\Pi^a)_{ij} (\bar{L})^i (e_R)^j \phi_a + (\Omega^a)_{ij} (\bar{E}_R)^i (\nu_R)^j \tilde{\phi}_a + \text{h.c.}, \end{aligned} \quad (2.7)$$

where Γ , Δ , Θ and Π are the 3×3 Yukawa matrices, Υ , Σ , and Ω are 3×6 matrices whereas Y is a 3×2 one. Note that only Y , Γ and Δ contain entries whose leading contributions are proportional to \mathcal{Y}_1 and \mathcal{Y}_2 . The remaining ones are purely of a radiative origin. Unlike what we have in the SM, here, the gauge symmetries allow for explicit construction of invariant bilinear and mass terms

$$\begin{aligned} \mathcal{L}_{\text{bil}} = & (M_D)_{ij} (\bar{D}_L)^i (D_R)^j + (M_E)_{ij} (\bar{E}_L)^i (E_R)^j + \frac{1}{2} (M_{\nu_R})_{ij} (\bar{\nu}_R)^i (\nu_R)^j + \\ & + (M_{LE})_{ij} (\bar{L})^i (E_R)^j + (\Xi)_{ij} (\bar{D}_L)^i (d_R)^j. \end{aligned} \quad (2.8)$$

These arise from the vector-like nature of the involved fields where $\text{SU}(2)_L$ transformations do not distinguish between left and right chiralities. All such mass terms in (2.8) are generated at the ω - f - p scales, thus larger than the EW scale. Note that the neutrino mass matrix M_{ν_R} is generated once the p , f , ω and s_i VEVs are developed. However, contrary to all remaining bilinear and Yukawa terms in the leptonic sector, its entries are generated by tree-level diagrams once the corresponding operators become allowed (see [17] for details). Therefore, small loop factors will not suppress the size of M_{ν_R} , whose entries can be up to an order of p , f and ω scales. As a byproduct, the neutrino sector automatically contains a seesaw mechanism and hence an explanation for the smallness of SM neutrino masses as we further discuss below. For completeness, we show the remaining Lagrangian terms in appendix A.

With the model fully defined, we finalise this section by showing the fermion mass matrices in the gauge eigenbasis that are implemented in our numerical analysis. First, for the quarks, and considering the components of the Q_L $\text{SU}(2)_L$ doublets as in (2.6), the new Lagrangian is written as

$$\begin{aligned} \mathcal{L}_{q,\text{SB}} = & \frac{v_a}{\sqrt{2}} (Y^a)_{ij} (\bar{d}_L)^i (\bar{D}_R)^j + \frac{v_a}{\sqrt{2}} (\Gamma^a)_{ij} (\bar{d}_L)^i (\bar{d}_R)^j + \frac{v_a}{\sqrt{2}} (\Delta^a)_{ij} (\bar{u}_L)^i (\bar{u}_R)^j + \\ & + (M_D)_{ij} (\bar{D}_L)^i (\bar{D}_R)^j + (\Xi)_{ij} (\bar{D}_L)^i (\bar{d}_R)^j, \end{aligned} \quad (2.9)$$

with v_a being the VEV of the respective Higgs doublet ϕ_a . The up-type quark mass matrix written in the basis $\{u_L^1, u_L^2, u_L^3\} \otimes \{u_R^1, u_R^2, u_R^3\}$ takes the form

$$[M_u] = \frac{v_a}{\sqrt{2}} \begin{bmatrix} \Delta_{11}^a & \Delta_{12}^a & \Delta_{13}^a \\ \Delta_{21}^a & \Delta_{22}^a & \Delta_{23}^a \\ \Delta_{31}^a & \Delta_{32}^a & \Delta_{33}^a \end{bmatrix}. \quad (2.10)$$

The eigenvalues of $[M_u]$ give the masses of the up-type quarks whose leading contributions are proportional to (2.3). A similar strategy can be now employed for the down quark sector where, in the basis $\{d_L^1, d_L^2, d_L^3, D_L^1, D_L^2\} \otimes \{d_R^1, d_R^2, d_R^3, D_R^1, D_R^2\}$, we have

$$[M_d] = \begin{bmatrix} \frac{v_a}{\sqrt{2}} \Gamma_{11}^a & \frac{v_a}{\sqrt{2}} \Gamma_{12}^a & \frac{v_a}{\sqrt{2}} \Gamma_{13}^a & \frac{v_a}{\sqrt{2}} Y_{11}^a & \frac{v_a}{\sqrt{2}} Y_{12}^a \\ \frac{v_a}{\sqrt{2}} \Gamma_{21}^a & \frac{v_a}{\sqrt{2}} \Gamma_{22}^a & \frac{v_a}{\sqrt{2}} \Gamma_{23}^a & \frac{v_a}{\sqrt{2}} Y_{21}^a & \frac{v_a}{\sqrt{2}} Y_{22}^a \\ \frac{v_a}{\sqrt{2}} \Gamma_{31}^a & \frac{v_a}{\sqrt{2}} \Gamma_{32}^a & \frac{v_a}{\sqrt{2}} \Gamma_{33}^a & \frac{v_a}{\sqrt{2}} Y_{31}^a & \frac{v_a}{\sqrt{2}} Y_{32}^a \\ \Xi_{11} & \Xi_{21} & \Xi_{31} & (M_D)_{11} & (M_D)_{12} \\ \Xi_{12} & \Xi_{22} & \Xi_{32} & (M_D)_{21} & (M_D)_{22} \end{bmatrix}. \quad (2.11)$$

Unlike the up sector, here we have NP contributions. Besides the down, strange and bottom quarks, we also have two new VLQs which we name as d_4 and d_5 ² defined in such a way that $m_{d_5} > m_{d_4}$. The leading contributions to the down-type quark masses are proportional to eqs. (2.2) and (2.4).

We can now extend this analysis to the lepton sector and write down the mass matrices for the charged leptons and neutrinos. Starting with the charged leptons, in the basis $\{e_L^i, e_L^i\} \otimes \{e_R^j, e_R^j\}$ one gets

$$[M_L] = \begin{bmatrix} [(M_E)_{ij}]_{3 \times 3} & \left[\frac{v_a}{\sqrt{2}} (\Theta^a)_{ij} \right]_{3 \times 3} \\ [(M_{LE})_{ij}]_{3 \times 3} & \left[\frac{v_a}{\sqrt{2}} (\Pi^a)_{ij} \right]_{3 \times 3} \end{bmatrix}, \quad (2.12)$$

and for the neutrinos, in the basis $\{\nu_{e_L}^i, \nu_{e_L}^i, \nu_{e_R}^i, \nu_R^j\} \otimes \{\nu_{e_L}^i, \nu_{e_L}^i, \nu_{e_R}^i, \nu_R^j\}$ we arrive at

$$[M_\nu] = \begin{bmatrix} [0]_{3 \times 3} & [0]_{3 \times 3} & [M_{LE}]_{3 \times 3} & \left[\frac{v_a \Sigma^a}{\sqrt{2}} \right]_{3 \times 6} \\ [0]_{3 \times 3} & [0]_{3 \times 3} & [M_E]_{3 \times 3} & \left[\frac{v_a \Upsilon^a}{\sqrt{2}} \right]_{3 \times 6} \\ [M_{LE}]_{3 \times 3}^\dagger & [M_E]_{3 \times 3}^\dagger & [0]_{3 \times 3} & \left[\frac{v_a \Omega^a}{\sqrt{2}} \right]_{3 \times 6} \\ \left[\frac{v_a \Sigma^a}{\sqrt{2}} \right]_{6 \times 3}^\dagger & \left[\frac{v_a \Upsilon^a}{\sqrt{2}} \right]_{6 \times 3}^\dagger & \left[\frac{v_a \Omega^a}{\sqrt{2}} \right]_{6 \times 3}^\dagger & [M_{\nu_R}]_{6 \times 6} \end{bmatrix}, \quad (2.13)$$

where $i = 1, 2, 3$ as usual and $j = 1, \dots, 6$. For charged leptons, besides the SM-like states we also have exotic VLLs which we name as e_4 , e_5 and e_6 ,³ defined in such a way that

²This rather simplistic nomenclature is used to facilitate the designation when doing numerical analysis, as this is the name of the particle as defined in the UFO files. The designation in [16, 17] and above in eq. (2.2) corresponds to $d_4 \equiv D$, $d_5 \equiv S$.

³Again, in accordance with [16, 17], we have $e_4 \equiv E$, $e_5 \equiv M$, $e_6 \equiv T$.

$m_{e_6} > m_{e_5} > m_{e_4}$. The neutrino sector is quite rich in new particles, besides the three SM-like ones, we have a total of twelve new states. The numerical analysis will only consider the three lightest, keV-MeV scale BSM neutrinos which are denoted as $\nu_4 \equiv \nu_{\text{BSM}}$, ν_5 and ν_6 .

2.3.1 Physically viable benchmark scenarios for masses

Before moving to the numerical analysis we present possible benchmark scenarios for couplings and masses in such a way to preserve the key properties emergent from the unification picture as well as complying with the measured phenomenological quantities.

The main focus of this work is the construction of an analysis framework dedicated to the study of VLLs and how important the DL techniques can be. This will enable us to propose robust signal events to be tested via direct searches at the LHC as well as understanding whether the model under consideration can be probed in such a sector. As it was shown in [17], under certain approximations and before EW symmetry breaking (EWSB), the lepton mass matrix is reduced to⁴

$$[M_L] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa_7 \omega & \kappa_5 \omega \\ 0 & 0 & 0 & 0 & \kappa_6 \omega & \kappa_8 \omega \\ 0 & 0 & 0 & 0 & \kappa_1 p & \kappa_3 f \\ 0 & 0 & 0 & \kappa_2 p & 0 & 0 \\ 0 & 0 & 0 & \kappa_4 f & 0 & 0 \end{bmatrix}, \quad (2.14)$$

where the various κ_i terms are radiatively generated Yukawa couplings, thus expected to be smaller than unity. The VLL masses are then

$$\begin{aligned} m_{e_6}^2 &= p^2 \kappa_2^2 + f^2 \kappa_4^2, \\ m_{e_5, e_4}^2 &= \frac{1}{2} \left(\omega^2 \Lambda_1 + p^2 \kappa_1^2 + f^2 \kappa_3^2 \pm \left[\left(\omega^2 \Lambda_1 + p^2 \kappa_1^2 + f^2 \kappa_3^2 \right)^2 \right. \right. \\ &\quad \left. \left. - 4 \omega^2 \left(\omega^2 \Lambda_2 - 2 f p \Lambda_3 + p^2 \Lambda_4 + f^2 \Lambda_5 \right) \right]^{1/2} \right), \end{aligned} \quad (2.15)$$

where we defined $\Lambda_1 = \kappa_5^2 + \kappa_6^2 + \kappa_7^2 + \kappa_8^2$, $\Lambda_2 = (\kappa_5 \kappa_6 - \kappa_7 \kappa_8)^2$, $\Lambda_3 = (\kappa_5 \kappa_7 + \kappa_6 \kappa_8) \kappa_1 \kappa_3$, $\Lambda_4 = (\kappa_5^2 + \kappa_8^2) \kappa_1^2$ and $\Lambda_5 = (\kappa_6^2 + \kappa_7^2) \kappa_3^2$. The plus sign in (2.15) corresponds to e_5 and the minus sign to e_4 .

Considering a scenario where $\omega \sim f \ll p$, Taylor expansion of (2.15) leads to the simplified expressions

$$\begin{aligned} m_{e_6} &\approx p \kappa_2, \\ m_{e_5} &\approx p \kappa_1, \\ m_{e_4} &\approx \omega \sqrt{\kappa_5^2 + \kappa_8^2}. \end{aligned} \quad (2.16)$$

⁴It is important to note that *this does not* represent a one-to-one correspondence between (2.14) and (2.12). One should interpret (2.14) as a matrix one would get by following all symmetry breaking steps as seen in figure 1, while (2.12) corresponds to the stage immediately after the ω , f and p VEVs.

Along the lines of what was discussed in [17], let us consider a set of possible solutions with

- $\kappa_2 \sim \mathcal{O}(10^{-2})$, $\kappa_1 \sim \mathcal{O}(10^{-3.5} - 10^{-2})$ and $\kappa_{5,8} \sim \mathcal{O}(10^{-3} - 10^{-2})$,
- $p \sim \mathcal{O}(500 - 1000 \text{ TeV})$ and $\omega \sim f \sim \mathcal{O}(100 \text{ TeV})$.

This benchmark scenario leads to the following mass ranges

- $m_{e_6} \sim \mathcal{O}(5 - 10 \text{ TeV})$,
- $m_{e_5} \sim \mathcal{O}(0.15 - 10 \text{ TeV})$,
- $m_{e_4} \sim \mathcal{O}(0.1 - 1 \text{ TeV})$,

which we will use as a guiding principle for our numerical analysis. In particular, we see that for the model under consideration e_4 can be light enough to be probed at the LHC. On another hand, e_6 will always be rather heavy and a potential observation at the LHC would likely be very challenging. Regarding e_5 , we see that it can either be as heavy as e_6 or as light as e_4 depending on yet unexplored model details. Based on this estimation, we will consider both possibilities in the numerical studies.

To finalise this subsection, let us consider the neutrino sector. Before EWSB, the mass matrix is block diagonal,

$$M_\nu = \begin{bmatrix} \bar{M}_{9 \times 9} & 0 \\ 0 & M_{6 \times 6} \end{bmatrix}, \quad (2.17)$$

where \bar{M} represents neutral components belonging to $SU(2)_L$ doublets while M denotes SM singlets corresponding to ν_R in table 2. Starting with the $M_{6 \times 6}$ block, which corresponds to M_{ν_R} in (2.8), its components offer the larger contributions to the neutrino mass matrix. In this sector, hierarchies among gauge eigenstates result from the relative sizes of the EW-preserving VEVs. On the other hand, the \bar{M} components are radiatively generated and share the same properties as the VLLs. Thus, after the p , f and ω VEVs one can write

$$\bar{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \kappa_7 \omega & \kappa_5 \omega \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \kappa_6 \omega & \kappa_8 \omega \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \kappa_1 p & \kappa_3 f \\ 0 & 0 & 0 & 0 & 0 & 0 & \kappa_2 p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \kappa_4 f & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa_2 p & \kappa_4 f & 0 & 0 & 0 \\ 0 & \kappa_7 \omega & \kappa_6 \omega & \kappa_1 p & 0 & 0 & 0 & 0 & 0 \\ 0 & \kappa_5 \omega & \kappa_8 \omega & \kappa_3 f & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2.18)$$

with eigenvalues,

$$m_{\nu_{1,2,3}}^2 = 0, \quad m_{\nu_{4,5}}^2 = m_{e_6}^2, \quad m_{\nu_{6,7}}^2 = m_{e_5}^2, \quad m_{\nu_{8,9}}^2 = m_{e_4}^2, \quad (2.19)$$

such that, the left-handed neutrino components, at this stage, share the same masses as their charged lepton partners. In total, and before EWSB, we have three massless, and twelve massive neutrinos (six from the doublets and six from singlets). In the corresponding mass basis, if we identify the massive states as μ_i ($i = 1, \dots, 12$), we can recast the neutrino mass matrix in a condensed notation as

$$m_\nu = \begin{pmatrix} \mathbf{0}_{3 \times 3} & \frac{v_{\text{EW}}}{\sqrt{2}} (\mathbf{y}_\nu)_{3 \times 12} \\ \frac{v_{\text{EW}}}{\sqrt{2}} (\mathbf{y}_\nu^\top)_{12 \times 3} & (\boldsymbol{\mu}_N)_{12 \times 12} \end{pmatrix}, \quad (2.20)$$

where the contribution of EWSB VEVs was already included. Note that \mathbf{y}_ν are the 3×12 Yukawa matrices whose entries are all radiatively generated. While a more dedicated analysis is beyond the scope of this work, this structure can potentially offer three sub-eV states as well as light keV-MeV order sterile neutrinos as we will assume in our numerical studies.

2.3.2 Physically viable benchmark scenarios for couplings

In our numerical analysis we will be using `MadGraph5` [32] which is a tool that requires a theory written in the mass basis. Therefore, not only masses but also couplings need to be rotated to such a basis. To this end we use `SARAH` [33], which also offers a complete set of Feynman rules with physical fields. All relevant diagrams for our studies are shown in appendix A. Note that all signal and background processes that we will consider involve only triple gauge self-interactions as well as fermion-fermion-gauge vertices. While the gauge sector is purely SM-like with well-known parameters, the Feynman rules involving fermion vertices will be sensitive to elements of the mixing matrices in the charged lepton (including VLLs) and neutrino sectors, defined by the bi-unitary transformations

$$\begin{aligned} U_L^e \cdot M_L \cdot U_R^{e\dagger} &= m_e^{\text{diag}}, \\ U_\nu \cdot m_\nu \cdot U_\nu^\dagger &= m_\nu^{\text{diag}}. \end{aligned} \quad (2.21)$$

Let us now discuss which phenomenological constraints are applied to these matrices. First, for the charged leptons, we consider the limit where the SM-like sector is flavour-diagonal. Therefore, in U_L^e and U_R^e , we add a 3×3 identity block and consider a limiting scenario where there is no mixing with VLLs. While this may not be the case in general, a realistic scenario cannot strongly deviate from the flavour alignment limit that we impose. A complete study with flavour mixing is beyond the scope of this work. For the VLL block, we consider a generic mixing with the only restriction being that both U_L^e and U_R^e are unitary. To summarize, the lepton mixing matrices used in the numerical analysis are given by

$$U_{L,R}^e = \begin{bmatrix} \mathbb{1}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & (U_{L,R}^{\text{VLL}})_{3 \times 3} \end{bmatrix}, \quad (2.22)$$

where $U_{L,R}^{\text{VLL}} \cdot U_{L,R}^{\text{VLL}\dagger} = \mathbb{1}_{3 \times 3}$.

For the neutrino sector, we also consider a limiting scenario where, for simplicity, the mixing between the three light active neutrinos and the remaining twelve BSM states is

zero. Once again, a more generic case with flavour mixing is beyond the scope of our analysis and does not significantly affect our main conclusions. Note, however, that mixing among light neutrinos is allowed and fixed by the PMNS matrix. For the remaining BSM 12×12 block we recall that the mixing among right-handed and left-handed components is radiatively generated and is thus likely small. Here, we consider that those elements are always smaller than 10^{-3} . Having said this, the full neutrino mixing matrix reads

$$U_\nu = \begin{bmatrix} (U_{\text{PMNS}})_{3 \times 3} & 0_{3 \times 12} \\ 0_{12 \times 3} & (U_\nu^{\text{BSM}})_{12 \times 12} \end{bmatrix}, \quad (2.23)$$

with,

$$U_\nu^{\text{BSM}} = \begin{bmatrix} (U_1)_{6 \times 6} & (D_1)_{6 \times 6} \\ (D_1)^\dagger_{6 \times 6} & (U_2)_{6 \times 6} \end{bmatrix}. \quad (2.24)$$

We set the matrix elements in D_1 to be of order $\mathcal{O}(10^{-3} - 10^{-8})$ while in $U_{1,2}$ they are randomly generated in a way that preserves unitarity and guarantees that e_4 couples democratically to the sterile neutrinos. With the above ingredients we have defined a possible benchmark scenario to start our collider phenomenology studies while preserving the essential features of the model under consideration.

3 Searching for vector-like leptons at the LHC

As a first step, we create the necessary UFO [34] files using SARAH. These are later used by MadGraph5 (MG5) to generate the signal and background Monte-Carlo events. Hard-scattering events are generated with Pythia8 [35], and then Delphes [36] is used to include hadronization and detector effects.

All hard-scattering events are generated using pp collisions at 14 TeV center-of-mass energy, with the parton distribution function *nn23lo1* and with the strong coupling constant α_s fixed automatically by MG5. We have generated a total of 250000 events. The background channels up to two extra jets are generated with the MLM matching scheme [37]. While substantial theoretical work has already been done over the last decades [24, 25, 28, 38–51], only recently, the searches for exotic charged leptons have started. The most recent analysis was done by the CMS collaboration at the LHC in 2019 [52], where a search for VLLs coupling to taus was performed. In fact, one of the three topologies that we propose in our analysis is very similar to the one in figure 1 of [52], and more in line to what we see, for example, in figure 30 of [53]. In the context of our model, such topology can be seen in figure 2 which, in what follows, will be referred to as “ZA”.

Recall that we are treating the lightest, sterile, BSM neutrinos as MET as long as their mass is in the keV-MeV mass range. This means that possible decays are kinematically forbidden and therefore, such neutrinos escape the detector. Of course, it would be interesting to study a scenario where we have SM neutrinos instead of the BSM ones provided that the final state would be purely the SM one. However, due to the structure of the mixing matrix used in this study (2.22), such a mixing is non-existing. We are leaving for

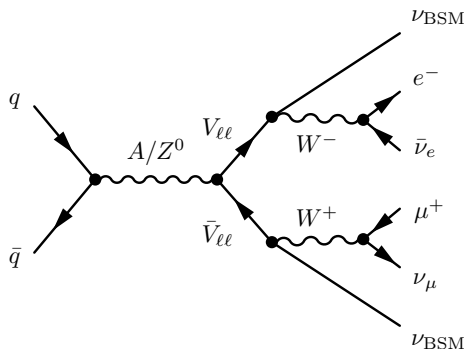


Figure 2. Leading-order Feynman diagram for the ZA topologies. Here, q and \bar{q} correspond to quarks originating from the colliding protons, $V_{\ell\ell}$ represents VLLs and ν_{BSM} denotes the lightest BSM neutrino. There are two purely SM leptonic final channels identified with ℓ and ν_ℓ .

a future work the scenario where inclusion of a non-zero mixing between SM leptons and VLLs, in consistency with both the flavour observables constraints and predictions from the high-scale theory, is implemented.

We also consider vector-boson fusion events whose topology is shown in figure 3 and that we shall refer to as “VBF”. While the latter is expected to have a smaller cross section, the presence of two well-defined forward jets enables us to tag such events using the high transverse mass of the forward jets. The signal channels, we propose here, provide a good starting point for our analysis. However, due to the expected low cross section for the signal events in comparison with the overwhelming cross-section of the irreducible background, searching for such particles at the LHC solely considering these two processes can become rather challenging. A third channel, denoted as “VLBSM”, with only four internal vertices (VBF diagrams contain eight vertices while ZA ones — six) is then considered and can be seen in figure 4. Furthermore, we use DL techniques inspired from previous works [13, 54, 55] and tailored for our analysis, in order to efficiently discriminate signal from background.

Another possibility would be to consider $\bar{e}_4 e_4$ pair production in ZA events leaving charged tracks in the detector. This would provide us a sizeable cross section and a clean signal potentially offering a good smoking gun for our model. However, preliminary numerical calculations have shown that e_4 decays too quickly before reaching the detector track chamber. To see this let us estimate the probability of travelling a distance equal or larger than x_0 given by [56]

$$P(x_0) = e^{-m_{e_4} x_0 \Gamma_{e_4} / |p|} . \tag{3.1}$$

Considering for example $m_{e_4} = 1$ TeV we have determined with MadGraph that $\Gamma_{e_4} = 314$ GeV. If we now require that the VLL needs to travel through a few layers of the ATLAS Semiconductor Tracker then $x_0 \approx 370$ mm. If the momentum of the particle is of the order $|p| \sim \mathcal{O}(100 \text{ GeV})$ the probability to leave a track in the Inner Detector is zero. In fact, already for a distance of $x_0 = 1$ mm, well before the pixel detector, we have $P(1 \text{ mm}) = 0.043$. The probability of e_4 , or any other VLL, surviving and leaving any imprints in the detectors is almost non-existent implying that one has to reconstruct e_4 from its main product decays. We will only consider the ZA, VBF and VLBSM topologies in our analysis and leave other possibilities for a future work.

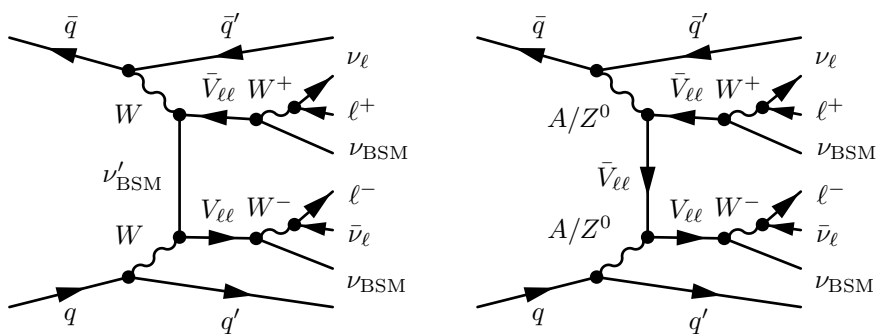


Figure 3. Leading-order Feynman diagrams for the VBF topologies. The same nomenclature as seen in figure 2 applies here. ν'_{BSM} correspond to any BSM neutrino.

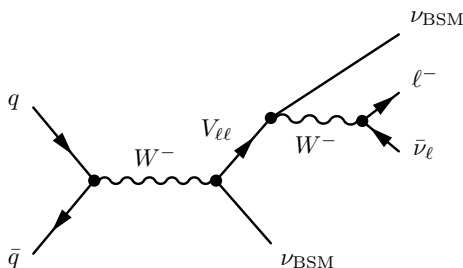


Figure 4. Leading-order Feynman diagrams for the VLBSM topologies. The same nomenclature as seen in figure 2 applies here.

The main irreducible background for each signal channel is chosen as follows:

- For ZA topologies, we consider $t\bar{t}$ and W^+W^- both with fully leptonic final states, $t\bar{t} + Z^0$, with Z^0 decaying into lepton/anti-lepton pair and $t\bar{t}$ fully leptonic decay and finally $t\bar{t} + Z^0$ with Z^0 decaying into neutrinos.
- For VBF topologies, we consider W^+W^- with fully leptonic final states, $t\bar{t} + (j, jj)$, where tops decay into leptons accompanied with either one or two light jets.
- For VLBSM topologies, we consider the single lepton production $pp \rightarrow \ell\nu_\ell$ with zero, one and two light jets.

Both the background and signal leptonic final states are chosen to be identical. The Feynman diagrams for W^+W^- and $t\bar{t}$ with fully leptonic final states are displayed in figures 5 and 6.

To facilitate the signal detection and reduce the background contamination we consider final-state leptons of different flavour. In particular, we choose W^+ decaying to μ^+ and ν_μ , while for W^- we consider the $e^- + \bar{\nu}_e$ channel. We also choose the following simple kinematic cuts as event selection criteria:

1. Charged leptons (e^- and μ^+) are required to have a transverse momentum $p_T > 25 \text{ GeV}$ and $|\eta| \leq 5$ and

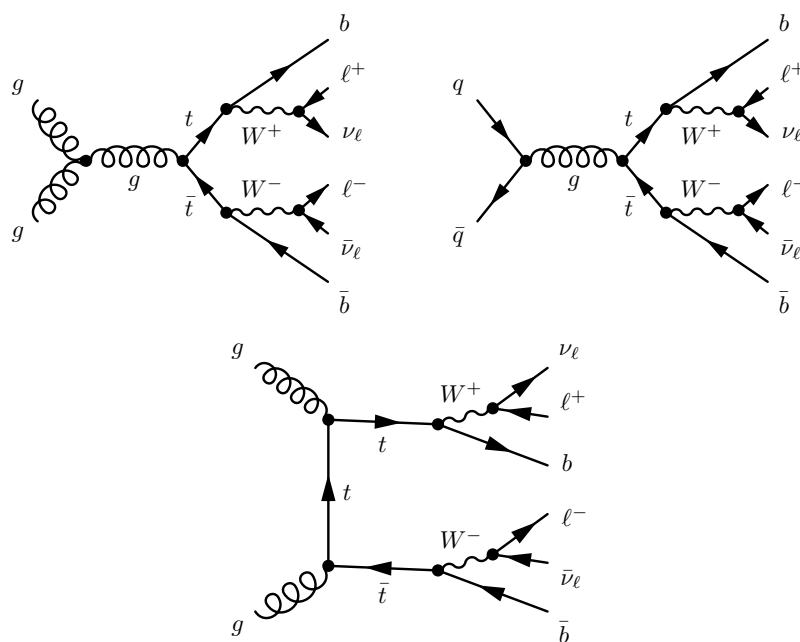


Figure 5. $t\bar{t}$ background topologies with fully leptonic final states. All particles shown here are purely SM ones. The final leptonic channels in both the signal events and the $t\bar{t}$ background processes shown here are chosen to be identical.

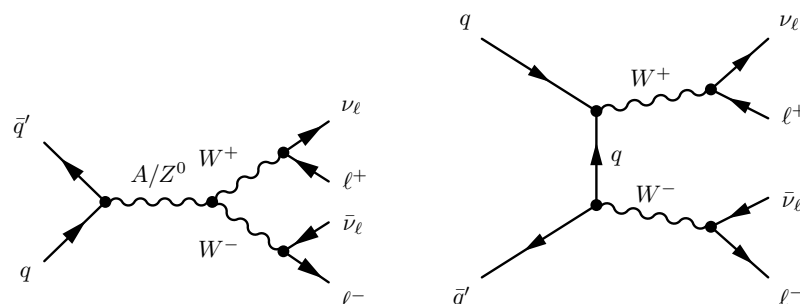


Figure 6. W^+W^- background topologies. All particles shown here are purely SM ones. The final leptonic channels in both the signal events and the W^+W^- background processes shown here are chosen to be identical.

2. Missing transverse energy $\cancel{E}_T > 15$ GeV.
3. For events with jets, we use the Cambridge/Aachen jet clustering algorithm with $\Delta R = 1.0$, transverse momentum $p_T > 35$ GeV and pseudo-rapidity $|\eta| \leq 5$.

At this point we are able to reconstruct all particles up to the VLLs with relative precision using the information from the final state visible particles, tracks and calorimetric towers provided by the `Delphes` output. It is important to note that in `Delphes` we only have at our disposal the total MET information, \vec{E}_T^{miss} , which is solely based on the calorimetric cells and can be represented as $\vec{E}_T^{\text{miss}} = -\sum_i^{\text{cell}} \vec{E}_T(i)$. However, one can make

use of this information and reprocess it in a more suitable manner. Our approach consists in picking up the generator-level information for the η , ϕ and E_T variables of the particles that will later be treated as invisible. We then use the $\eta \times \phi$ coordinates of the invisible particles in order to screen down our attention into nearby calorimeter tower cells. Then, based on the fact that invisible particles should not deposit any energy in the calorimeter, if such neighbouring cells show some activity or have any energy deposition the event is discarded, otherwise, the event is recorded, in particular, its $\eta \times \phi$ position. The following step of our procedure consists in building a list with the $\eta \times \phi$ inactive cell coordinates, dividing the total MET of the event by the number of such inactive cells and using the generator-level E_T information from the invisible particles as weights for each MET-bin. This allows one to identify which cells correspond to the ν_ℓ coordinates and then to build the missing transverse energy vectors which will later be used to reconstruct the W^\pm bosons and also to estimate the angular distributions $\Delta R(\ell, \nu_\ell)$ and $\cos(\theta_{\ell, \nu_\ell})$. In essence, our method, engineered to be used with real data, relies on a simple criterion: leptons need to be accompanied with MET in neighbouring cells for W^\pm events. Then, with the guidance of Monte-Carlo generator information, the strategy is to look into the calorimeter response and identify candidate events by assessing whether such a criterion is fulfilled or not.

Meanwhile, for the background we cannot reconstruct such objects, which means that we cannot use observables such as e.g. $(p_T(V_{\ell\ell}), M(V_{\ell\ell}), \dots)$ to train our deep learning models. In order to train our network models we build an object to mimic the reconstructed $V_{\ell\ell}$ for each background channel. First, for the backgrounds $pp \rightarrow t\bar{t} + j(jj)$ and $pp \rightarrow t\bar{t} + Z(Z \rightarrow \ell^+\ell^-)$, the object that plays the role of $V_{\ell\ell}$ is composed by the reconstructed W^\pm . On the other hand, the b -jets associated to the top quark decay products, that is, next to the W -boson, play the role of the would be ν_{BSM} . For the background $pp \rightarrow t\bar{t} + Z(Z \rightarrow \nu_\ell\bar{\nu}_\ell)$ we use the reconstructed W while neutrinos coming from the Z -boson decay play the role of ν_{BSM} . Finally, for the $pp \rightarrow W^+W^-$, we first reconstruct the W^+ using the associated charged lepton l^+ and the missing vector associated to the ν_l and to play the role of ν_{BSM} we use the missing vector associated to the $\bar{\nu}_l$ from the W^- decay. This way we have constructed an object to mimic $V_{\ell\ell}$ for this latter background. The object that mimics $\bar{V}_{\ell\ell}$ is reconstruct in the same way but considering the opposite charges.

The key feature of this method is that it allows one to clean the signal by completely removing background events. Note that, as a consequence of sizeable VLL decay widths, one can see in figures 20, 21 and 22 of appendix B that signal events can be characterized by rather large tails in, e.g. $p_T(e, \mu, W)$ distributions. For sufficiently large transverse momentum, say 400 GeV to be conservative, SM background processes are absent and any event is promptly identified by our deep learning tool as signal.

All chosen observables in our studies are detailed in section 3.2. We compute the observables both in the lab frame, as well as the W frame. We also emphasize that all Monte Carlo simulations and posterior data processing (PYTHIA8, MadGraph, Delphes) are performed in the blafis⁵ and ARGUS computer clusters as part of the overall computing infrastructure at the University of Aveiro.

⁵Technical details can be found at the Gr@v's website [57].

3.1 Methodology: Deep Learning models and dataset

In this section, we describe the construction of our neural network (NN) models and what are the best architectures we found to accurately separate and identify signal events from their respective backgrounds. For the unfamiliar reader, NNs, and by extension DL algorithms, are rooted in the *universal approximation theorem*, which essentially states that a given NN with a given number of hidden layers and a finite number of hidden units, a.k.a neurons, can approximate any arbitrary continuous function on compact subsets of \mathbb{R}^n ; this function can describe a hyperdimensional plane which separate samples from distinct classes (classification problems) or predict new samples based on previous one (regression problems). However, the universal approximation theorem does not tell us how deep the NN, neither the number of hidden units needed to better approximate the desired arbitrary function. This is a problem that one faces when identifying clever solutions and finding better designs for NNs.

The main goal of our NN model is to classify the signal channels defined in the previous section over each respective background. This procedure is often denoted as classification. For a better performance one must determine what kind of architecture should be employed, the number layers, how many neurons each layer needs, etc. An appropriate choice of such parameters can lead to models which are capable of giving very accurate predictions, which in the context of high energy physics can potentially lead to discoveries using the available data.

The problem of selecting the correct parameters is referred to as hyperparameter optimization or tuning. There are many methods to search for the best combinations of parameters including a brute force approach by testing each possible combination until the optimal NN model is recovered. However, as one might expect such a method is very time consuming. A more efficient procedure consists in using an evolutionary algorithm search [13, 55]. We define for our analysis the following set of hyperparameters:

- number of hidden layers: 1 to 5
- number of neurons in each layer: 256, 512, 1024 or 2048
- kernel initializer: ‘normal’, ‘he normal’, ‘he uniform’
- L2 regularization penalty: 1e-3, 1e-5, 1e-7
- activation function: ‘relu’, ‘elu’, ‘tanh’, ‘sigmoid’
- optimizer: ‘adam’, ‘sgd’, ‘adamax’, ‘nadam’

Our evolutionary algorithm is initialized by building a set of ten NNs using *Keras* [58] with *TensorFlow* [59] as back-end. The hyperparameters are then randomly chosen from the ones presented in the list above. Each NN is trained for 200 epochs and once the training phase is complete we select the top five NNs that have shown better performances in order to “breed” new NNs for the next iteration. Such NNs are then initialized with the hyperparameters of the selected ones and treated as “parent traits” while randomly

including new ones as mutations. We have set a 20% probability of a random mutation to occur. We then construct a new population set and repeat the training/evaluation process for five times, five generations until finally retrieving the best NN architecture.

Another important aspect of the evolutionary algorithm is the fitness function. This helps the algorithm to select the best architectures based on a pre-defined metric. In our case, we set two fitness functions, one where the best models are ranked according to their accuracy on the test set, and a second one that ranks the models according to their Asimov significance defined as:

$$\mathcal{Z}_A = \left[2 \left((s + b) \ln \left(\frac{(s + b)(b + \sigma_b^2)}{b^2 + (s + b)\sigma_b^2} \right) - \frac{b^2}{\sigma_b^2} \ln \left(1 + \frac{\sigma_b^2 s}{b(b + \sigma_b^2)} \right) \right) \right]^{1/2}, \quad (3.2)$$

with s and b being the number of signal and background events, respectively, and σ_b^2 is the variance of background events. We include a similar procedure to that described in [60] and the necessary modifications for the training methodology into our evolutionary algorithm.

Although, the main “body” of our NN is built using the principles of natural selection [61], some characteristics of the model construction are universal to all NN models that we summarize as follows:

- As input data, the NN receives a standard normalised vector, i.e. such a vector has mean 0 and standard deviation 1, and data sets from all observables are extracted from the ROOT detector output. This data set is then reshuffled and divided into a training set (80 % of data) and a test set (20 % of data). To avoid overfitting, we use the cross-validation with a five-fold scheme during the training of the NN.
- We employ a cyclic learning rate during the training phase with 0.01 initial value and maximal value of 0.1.
- In the output layer, the data is transported to a vector, with entries between 0 and 1 (which correspond to probabilities), in the format (S, B_j) , where S is the signal and B_j correspond to different backgrounds. The index j runs over the number of backgrounds chosen for a given signal. As an example, in ZA we consider 4 distinct backgrounds ($j = 4$). So, such a vector would correspond to (S, B_1, B_2, B_3, B_4) .
- Batch size of 32768 entries.
- We impose a total limit of 200 epochs with a patience of 5 epochs and a validation loss monitor, i.e. if the loss value on the test/validation set did not change for 5 epochs the training is resumed and all metrics are computed and stored to be passed to the evolutionary algorithm.
- To select the NN models with better accuracy we use the binary cross entropy (BCE) loss, while for the selection of those models that maximize the Asimov significance we use eq. (3.2) as a loss function during the evolutionary scan of the best hyperparameters.

	Dimension-full	Dimensionless	
Lab. frame	$p_T(e^-), p_T(\mu^+), p_T(e_4)$ $p_T(\bar{e}_4), m_{e_4}, m_{\bar{e}_4}$ $M_T(W^-), M_T(W^+), \text{MET}$	$\cos(\theta_{\bar{\nu}_e e}), \cos(\theta_{\bar{\nu}_\mu \mu^+}),$ $\cos(\theta_{W^- W^+}),$ $\cos(\Delta\phi), \cos(\Delta\theta),$ $\eta_e, \eta_{\mu^+}, \eta_{e_4}, \eta_{\bar{e}_4}$	$\Delta R(e, \bar{\nu}_e), \Delta R(\mu^+, \nu_{\mu^+})$
W^- frame	$p_T(e^-), p_T(e_4)$	$\cos(\theta_{\bar{\nu}_e e}),$ η_e, η_{e_4}	
W^+ frame	$p_T(\mu^+), p_T(\bar{e}_4)$	$\cos(\theta_{\nu_\mu \mu^+}),$ $\eta_{\mu^+}, \eta_{\bar{e}_4}$	
$\ell' \bar{\ell}'$ frame		$\cos(\Delta\phi), \cos(\Delta\theta)$	

Table 4. Kinematic (dimension-full) and angular (dimensionless) observables selected to study the ZA and VBF channels. We include observables in four different frames of reference: laboratory frame (top row), W^- rest frame (second row), W^+ rest frame (third row) and $\ell' \bar{\ell}'$ frame. $\theta_{i,j}$ denotes the angle between the respective particles from either the final state or reconstructed objects. Note that ν is the reconstructed missing energy vector, following the discussion in section 3.

	Dimension-full	Dimensionless
Lab. frame	$p_T(\mu^+), M_T(W),$ $p_T(W), \text{MET}$	$\cos(\theta_{\mu^+}), \cos(\theta_{\bar{\nu}_\mu \mu^+}),$ $\cos(\theta_W), \eta_{\mu^+}, \eta_W, \phi_{\mu^+}$

Table 5. Kinematic (dimension-full) and angular (dimensionless) observables selected to study the VLBSM channel. We compute observables in the laboratory frame where $\theta_{i,j}$ is the angle between the respective particles from either the final state or reconstructed objects. Note that ν is the reconstructed missing energy vector, following the discussion in section 3.

Another important aspect to mention is the fact that our data is unbalanced, i.e., we have more data points for some of the classes we are analyzing (in some cases, with a ratio of $s/b \approx 1.83$). This is a result of the event selection criteria that we impose for the signal and background. Unbalanced data can lead to models with lower predictive power for substantially outnumbered classes, which is a serious issue if one wants to search for NP phenomena. To avoid this, we use the Synthetic Minority Oversampling Technique (SMOTE) [62] to create synthetic entries for the minority classes in our training dataset. Note that we do not employ any re-sampling technique on the test dataset. This method is faster and more efficient than generating additional Monte-Carlo events and passing them through subsequent hadronization and detector effects.

Our dataset is stored in a table format where each row corresponds to an event entry that has successfully passed the selection cuts described in 3, and each column corresponds to the observables described in tables 4 and 5. The dimensions of each training and test datasets are displayed in table 6.

Original dataset:Training (SMOTE) Test	ZA		VBF		VLBSM	
	Signal	Backgrounds	Signal	Backgrounds	Signal	Backgrounds
$m_{e_4} = 200 \text{ GeV}$	(77405, 36) (65983, 36) (15481, 36)	$t\bar{t}$: (44725, 36) (65983, 36) (8945, 36) (81825, 36) $t\bar{t}, Z^0(l^+l^-)$: (65983, 36) (16365, 36) (46705, 36) $t\bar{t}, Z^0(\nu_l\bar{\nu}_l)$: (65983, 36) (9341, 36) (48475, 36) W^+W^- : (65983, 36) (9695, 36)	(115330,36) (91444, 36) (23066, 36)	$t\bar{t}+j(jj)$: (85870, 36) (91444, 36) (17174, 36) (48170, 36) W^+W^- : (91444, 36) (9634, 36)	(147825,10) (128870, 10) (29565, 10)	$l\nu_l$: (84185, 10) (128870, 10) (16837, 10) (160630,10) $l\nu_l+j(jj)$: (128870,10) (32126, 10)
$m_{e_4} = 486 \text{ GeV}$	(125455,36) (100656, 36) (25091, 36)	$t\bar{t}$: (45625, 36) (100656,36) (9125, 36) (81845, 36) $t\bar{t}, Z^0(l^+l^-)$: (100656, 36) (16369, 36) (46885, 36) $t\bar{t}, Z^0(\nu_l\bar{\nu}_l)$: (100656, 36) (9377, 36) (48630, 36) W^+W^- : (100656, 36) (9726, 36)	(143455,36) (114892, 36) (28691, 36)	$t\bar{t}+j(jj)$: (85840, 36) (114892, 36) (17168, 36) (49145, 36) W^+W^- : (114892, 36) (9829, 36)	(187530,10) (149901, 10) (37506, 10)	$l\nu_l$: (83065, 10) (149901, 10) (16613, 10) (161460,10) $l\nu_l+j(jj)$: (149901,10) (32292, 10)
$m_{e_4} = 677 \text{ GeV}$	(137310,36) (111257, 36) (27462, 36)	$t\bar{t}$: (45060, 36) (111257, 36) (9012, 36) (83905, 36) $t\bar{t}, Z^0(l^+l^-)$: (111257, 36) (16781, 36) (46790, 36) $t\bar{t}, Z^0(\nu_l\bar{\nu}_l)$: (111257, 36) (9358, 36) (48345, 36) W^+W^- : (111257, 36) (9669, 36)	(143455,36) (122589, 36) (28691, 36)	$t\bar{t}+j(jj)$: (85840, 36) (122589, 36) (17168, 36) (49145, 36) W^+W^- : (122589, 36) (9829, 36)	(195230,10) (156047, 10) (39046, 10)	$l\nu_l$: (83640, 10) (156047, 10) (16728, 10) (160870, 10) $l\nu_l+j(jj)$: (156047, 10) (32174, 10)
$m_{e_4} = 868 \text{ GeV}$	(146085,36) (116302, 36) (29217, 36)	$t\bar{t}$: (44460, 36) (116302, 36) (8892, 36) (82045, 36) $t\bar{t}, Z^0(l^+l^-)$: (116302, 36) (16409, 36) (47240, 36) $t\bar{t}, Z^0(\nu_l\bar{\nu}_l)$: (116302, 36) (9448, 36) (48380, 36) W^+W^- : (116302, 36) (9676, 36)	(157250,36) (125058, 36) (31450, 36)	$t\bar{t}+j(jj)$: (85245, 36) (125058, 36) (17049, 36) (48870, 36) W^+W^- : (125058, 36) (9774, 36)	(198290,10) (158405, 10) (39658, 10)	$l\nu_l$: (83740, 10) (158405, 10) (16748, 10) (160680, 10) $l\nu_l+j(jj)$: (158405, 10) (32136, 10)
$m_{e_4} = 1250 \text{ GeV}$	(151020,36) (120489, 36) (30204, 36)	$t\bar{t}$: (44370, 36) (120489, 36) (8874, 36) (81935, 36) $t\bar{t}, Z^0(l^+l^-)$: (120489, 36) (16387, 36) (47425, 36) $t\bar{t}, Z^0(\nu_l\bar{\nu}_l)$: (120489, 36) (9485, 36) (48635, 36) W^+W^- : (120489, 36) (9727, 36)	(159445,36) (126781, 36) (31889, 36)	$t\bar{t}+j(jj)$: (86005, 36) (126781, 36) (17201, 36) (48080, 36) W^+W^- : (126781, 36) (9616, 36)	(196480,10) (156931, 10) (39296, 10)	$l\nu_l$: (84140, 10) (156931, 10) (16828, 10) (160255, 10) $l\nu_l+j(jj)$: (156931, 10) (32051, 10)

Table 6. Dataset dimensions for each value of the lightest VLL mass, m_{e_4} . In the pairings (X, Y) , X denotes the number of events (rows) while Y is the number of features in each dataset. The (X, Y) pairs are organized in groups of three. The top ones correspond to the original dataset before splitting (80/20% ratio) and re-sampling using the SMOTE technique, the middle ones are the training set already balanced while the bottom ones correspond to the remaining test set (20% of the original one).

3.2 Results

We start our discussion by presenting a specific benchmark point whose parametric choice was guided by our discussion in sections 2.3.1 and 2.3.2. Note that the analysis methodology is independent of such a parametric choice. We will then study events for a light VLL (e_4) accompanied by the lightest BSM neutrino (ν_4), which is treated as MET, and whose masses read as

$$m_{e_4} = 677 \text{ GeV}, \quad m_{\nu_4} = 216 \text{ keV}. \quad (3.3)$$

The decay width is automatically calculated in **MadGraph** in the narrow width approximation. While for the ZA and VLBSM events heavier neutrinos are not important, in the VBF case they should be taken into account since they appear as intermediate states. We then fix their masses as

$$\begin{aligned} m_{\nu_5} &= 0.138 \text{ GeV}, & m_{\nu_6} &= 36.7 \text{ GeV}, & m_{\nu_7} &= 2140 \text{ GeV}, & m_{\nu_8} &= 2537 \text{ GeV}, \\ m_{\nu_9} &= 3035 \text{ GeV}, & m_{\nu_{10,11}} &= m_{e_4}, & m_{\nu_{12,13}} &= m_{e_5}, & m_{\nu_{14,15}} &= m_{e_6}, \end{aligned} \quad (3.4)$$

with

$$m_{e_5} = 3258 \text{ GeV}, \quad m_{e_6} = 4240 \text{ GeV}. \quad (3.5)$$

The BSM couplings are essentially the mixing matrices seen in section 2.3.2. Here we adopt,

$$\begin{aligned} U_L^{\text{VLL}} &= \begin{bmatrix} -0.162 - 0.381i & -0.683 - 0.321i & 0.318 + i0.379 \\ -0.315 - 0.089i & -0.341 + 0.225i & -0.746 - 0.411i \\ 0.844 - 0.0970i & -0.035 + 0.498i & -0.105 - 0.134i \end{bmatrix}, \\ U_R^{\text{VLL}} &= \begin{bmatrix} -0.186 - 0.490i & -0.266 - 0.462i & -0.654 + i0.113 \\ 0.389 - 0.153i & -0.750 - 0.366i & 0.306 - 0.186i \\ 0.640 - 0.375i & 0.035 + 0.133i & -0.497 - 0.429i \end{bmatrix}, \end{aligned} \quad (3.6)$$

for the VLLs, while for the neutrinos we have⁶

$$\begin{aligned}
 U_1 &= \begin{bmatrix} -0.490 & -0.408 & -0.081 & -0.192 & 0.568 & 0.476 \\ 0.491 & -0.745 & 0.436 & -0.063 & 0.014 & -0.10 \\ 0.098 & -0.109 & -0.511 & -0.337 & 0.385 & -0.675 \\ -0.445 & 0.003 & 0.455 & 0.511 & 0.278 & -0.506 \\ -0.553 & -0.267 & 0.073 & -0.434 & -0.614 & -0.227 \\ 0.068 & 0.443 & 0.574 & -0.630 & 0.270 & -0.029 \end{bmatrix}, \\
 U_2 &= \begin{bmatrix} -0.413 & 0.166 & 0.360 & -0.313 & 0.270 & 0.708 \\ 0.304 & -0.082 & 0.292 & -0.630 & -0.647 & 0.017 \\ -0.053 & 0.842 & 0.374 & 0.062 & -0.033 & -0.379 \\ -0.080 & 0.145 & -0.040 & 0.597 & -0.643 & 0.449 \\ -0.837 & -0.259 & 0.064 & -0.062 & -0.282 & -0.380 \\ 0.163 & -0.411 & 0.800 & 0.376 & 0.121 & -0.095 \end{bmatrix}.
 \end{aligned} \tag{3.7}$$

Let us stress here that the numerical values above were randomly generated but in consistency with the theory requirements discussed in sections 2.3.1 and 2.3.2.

The overall cross sections for both signal and background events are estimated by MadGraph and for our benchmark point read:

$$\begin{aligned}
 \text{ZA: } \sigma &= 4.40 \times 10^{-7} \pm 2.62 \times 10^{-10} \text{ pb}, \\
 \text{VBF: } \sigma &= 8.96 \times 10^{-7} \pm 5.88 \times 10^{-10} \text{ pb}, \\
 \text{VLBSM: } \sigma &= 7.70 \times 10^{-5} \pm 4.33 \times 10^{-8} \text{ pb}, \\
 t\bar{t} &: \sigma = 6.72 \pm 3.01 \times 10^{-3} \text{ pb}, \\
 t\bar{t} + j &: \sigma = 7.85 \pm 5.06 \times 10^{-3} \text{ pb}, \\
 t\bar{t} + jj &: \sigma = 5.99 \pm 3.70 \times 10^{-3} \text{ pb}, \\
 t\bar{t} + Z^0(\ell^+\ell^-) &: \sigma = 5.36 \times 10^{-4} \pm 3.81 \times 10^{-7} \text{ pb}, \\
 t\bar{t} + Z^0(\bar{\nu}_\ell\nu_\ell) &: \sigma = 1.06 \times 10^{-3} \pm 6.95 \times 10^{-7} \text{ pb}, \\
 W^+W^- &: \sigma = 0.839 \pm 5.45 \times 10^{-4} \text{ pb}, \\
 pp \rightarrow \ell\nu_\ell &: \sigma = 10309.1 \pm 5.4 \text{ pb}, \\
 pp \rightarrow \ell\nu_\ell + j &: \sigma = 2943.6 \pm 2.1 \text{ pb}, \\
 pp \rightarrow \ell\nu_\ell + jj &: \sigma = 1233.2 \pm 0.7 \text{ pb}.
 \end{aligned}$$

As one can see the main problem we face is the overwhelming background resulting from $t\bar{t}$ events whose cross-section largely overtakes that of ZA and VBF channels, as well as $pp \rightarrow \ell\nu_\ell + (0, 1 \text{ and } 2)$ jets whose cross-section exceeds that of the VLBSM channel by at least eight orders of magnitude. While each diagram in figure 3 has a larger suppression

⁶The numerical values for D_1 are small, $\mathcal{O}(10^{-3} - 10^{-8})$, so not to occupy too much space, they are omitted. The dominant contributions all come from U_1 and U_2 .

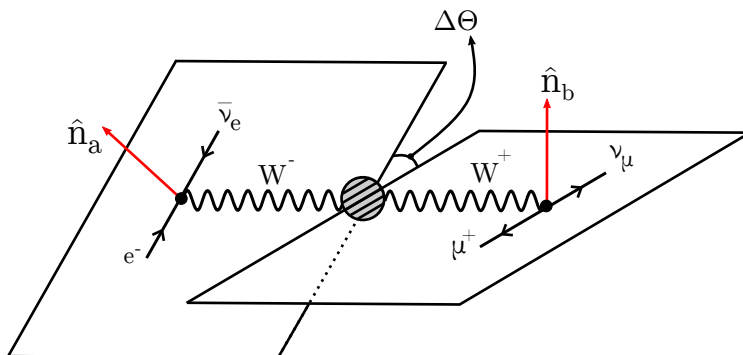


Figure 7. Schematic representation of the new angles for the variables $\cos(\Delta\theta)$ and $\cos(\Delta\phi)$. The vectors \hat{n}_a and \hat{n}_b are normal to the plane formed by the decay products of W^- and W^+ . We can use these vectors to determine the angle $\Delta\theta$. Note that $\nu_{e,\mu}$ should be solely regarded as two distinct MET vectors, say $\vec{E}_{T(1)}^{\text{miss}}$ and $\vec{E}_{T(2)}^{\text{miss}}$, corresponding to inactive cells in the vicinity of a charged lepton as explained in the three paragraphs before of section 3.1.

factor associated with the presence of more interaction vertices and internal propagators, it ends up generating more contributions to the overall cross section. Indeed, while in figure 2 we only have e_4 as an intermediate state, so less combinations are concerned, in figure 3 we have all BSM neutrinos contributing to the ν'_{BSM} propagator thus implying a larger number of possible VBF processes. Furthermore, two of the BSM neutrinos in ν'_{BSM} are rather light with masses of the order of 100 keV (for ν_4) and 100 MeV (for ν_5), which, on its own, offers an enhancement factor of at least 6 and 3 orders of magnitude, respectively, in comparison to massive EW-scale (or above) propagators.

The relevant observables for VLBSM signals are detailed in table 5, where both angular and kinematic variables are determined in the laboratory frame. The most obvious distinction between this dataset and the previous two is in the number of features. While the VLBSM signal with less internal propagators and only one lepton final state yields cross-sections larger than those of VBF and ZA events, as a drawback, we do not have a wealth of distinct variables to choose from. Not only that, the VLBSM topology and its corresponding backgrounds do not allow for a direct one-to-one correspondence between variables such as VLL invariant mass distributions, as well as the azimuthal and polar angles $\cos(\Delta\phi)$ and $\cos(\Delta\theta)$ are absent in VLBSM events. A schematic representation of these new angular variables can be seen in figure 7, where $\Delta\theta$ is the angle between the W^- and W^+ planes, and $\Delta\phi$ is the azimuthal angle between those two planes. The only relevant distributions, as specified in table 5, can be seen in figure 20 of appendix B.

Starting with the ZA/VBF signal and background topologies one can reconstruct the top, W and e_4 masses within the expected range provided that there is a noticeable difference between them. We also observe a sizeable separation between signal events and background, especially for ΔR distributions, where, for the former, $\Delta R_{e-\bar{\nu}_e}$ and $\Delta R_{\mu^+\nu_\mu}$ showcase a peak near zero, well separated from background events. For pseudo-rapidity distributions the majority of signal events have a peak at around zero indicating particle trajectories perpendicular to the beam axis, which helps to separate events from some of

the background channels ($t\bar{t}, W^+W^-$). While there are indeed sizeable differences in some variables, others are clearly dominated by background, especially for angular variables $\cos(\theta)$. Similar conclusions arise when observing the distributions for the VLBSM channel (see figure 20), with pseudo-rapidity distributions providing a good signal-to-background separation, whereas angular distributions suffer from the considered backgrounds.

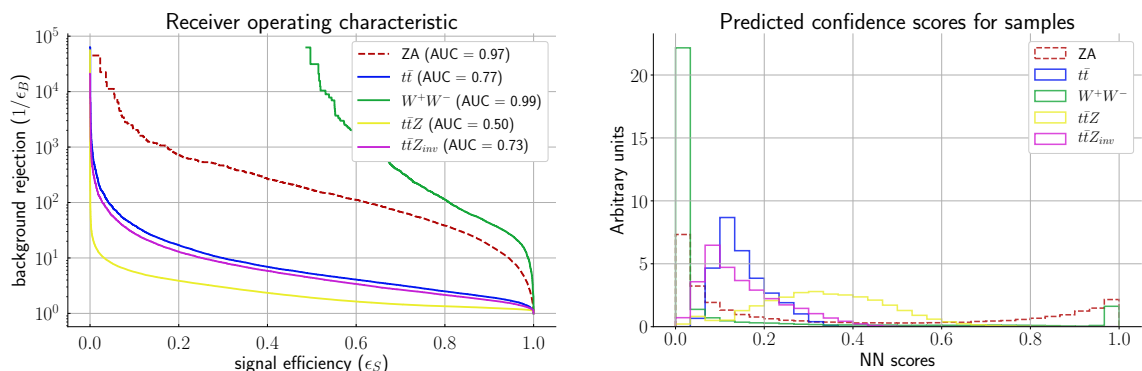
Generally speaking, for all studied channels, kinematic distributions such as transverse momentum and missing transverse energy can offer some degree of distinction in order to separate signal and background events. In fact, for backgrounds, these distributions tend to accumulate at lower energies when compared to signals. In particular, for transverse momentum distributions we have such an accumulation of events at $p_T < 200$ GeV and for missing energy the preferred region is $\text{MET} < 200$ GeV. On another hand, for signal events, we have a significant accumulation in the high energy region where $p_T/\text{MET} > 300$ GeV. In fact, due to a rich neutrino sector, MET distributions are of particular interest as the signals we are considering here which contain both BSM and SM missing energy.

However, it is important to note that the information available at experiments is limited, typically referred to as low level observable. This operates mostly on counting the number of hits (or events) that were “observed” by a given detector. A high level approach would combine the different information from these detectors with further complex and sophisticated observables, such as the ones we explore in this work. The use of such a multitude of observables, including the variables in the W frame, will then serve as an important step in the subsequent analysis. It allows us to build a vast dataset for DL studies, which in turn, allows for a quicker training and a greater overall accuracy despite lower cross-sections.

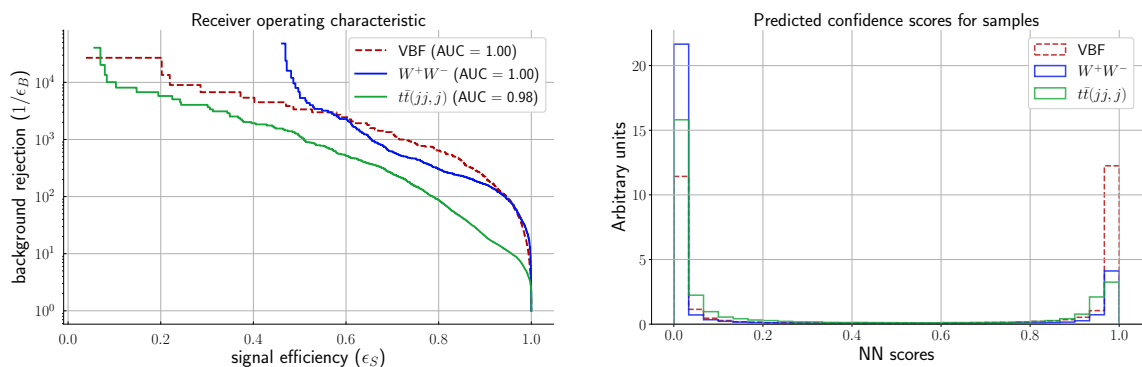
For the benchmark scenario considered above (a VLL with $m_e = 667$ GeV and a BSM neutrino in the keV range), the architecture that maximizes the accuracy can be seen in tables 10, 11 and 12. On the other hand, the architecture that maximizes the Asimov significance is shown in tables 13, 14 and 15. A quantitative approach to evaluate our NN models can be done with the help of ROC (Receiver operating characteristic) curves, which represent a measure of how well the NN classification has performed. In figure 8 we show our results for the best accuracy whereas the best Asimov significance can be seen in figure 9. As one can observe, for the models which perform with a better accuracy, the selected architectures are capable of separating signal events from background with almost 100 % efficiency. In particular, we see that signal events are above background events with signal efficiencies of about $\varepsilon_S = 1$, with 97% accuracy for ZA, 100% for VBF and 98% for VLBSM channels.

However, we note that a large significance does not necessarily imply a good accuracy. In fact, for the NN architectures that maximize the Asimov significance, the accuracy is substantially reduced. A particularly relevant example is that of the VLBSM channel exhibiting the lowest accuracy with a value $\text{AUC} = 0.32$. The predicted confidence scores can be found on the right panels of figure 8. Note that the NN assigns a different score to each prediction. For example, taking figure 8(a), the NN score of 1.0 labels an event that is either a signal or a W^+W^- background event.

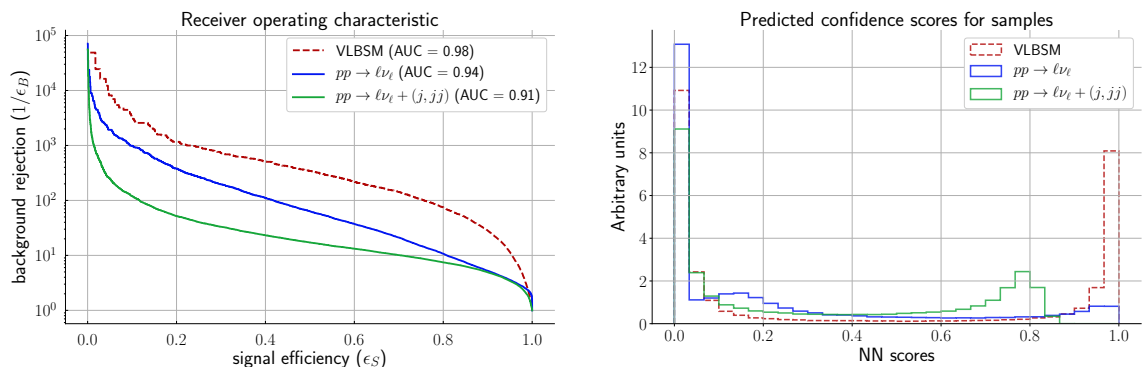
The significance of signal events is typically regarded by experimental physicists as the most meaningful measure to either claim a discovery or that a given NP candidate is



(a) ZA topologies.

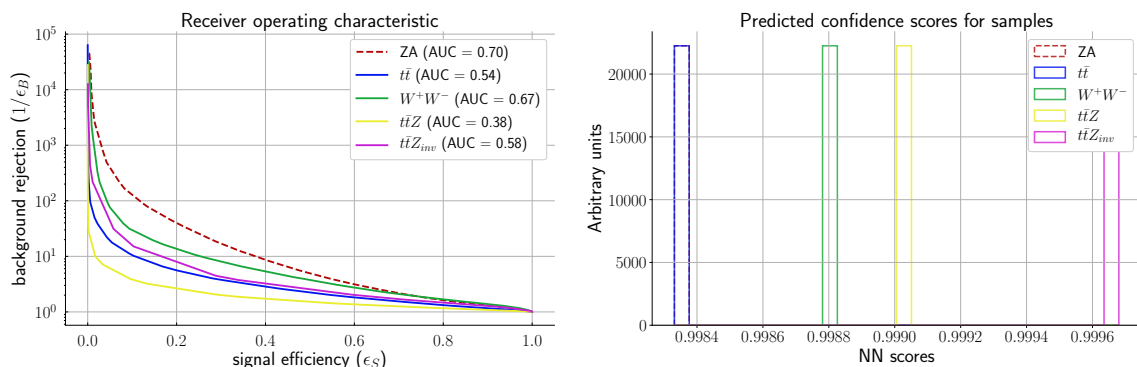


(b) VBF topologies.

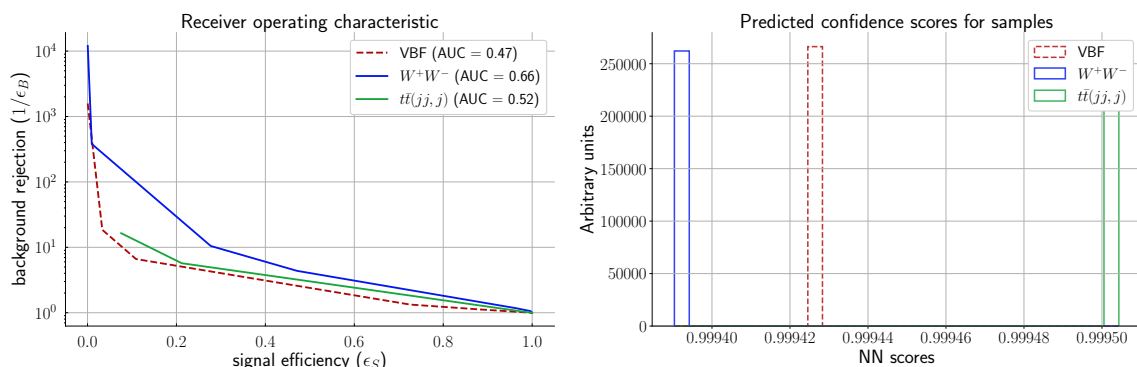


(c) VLBSM topologies.

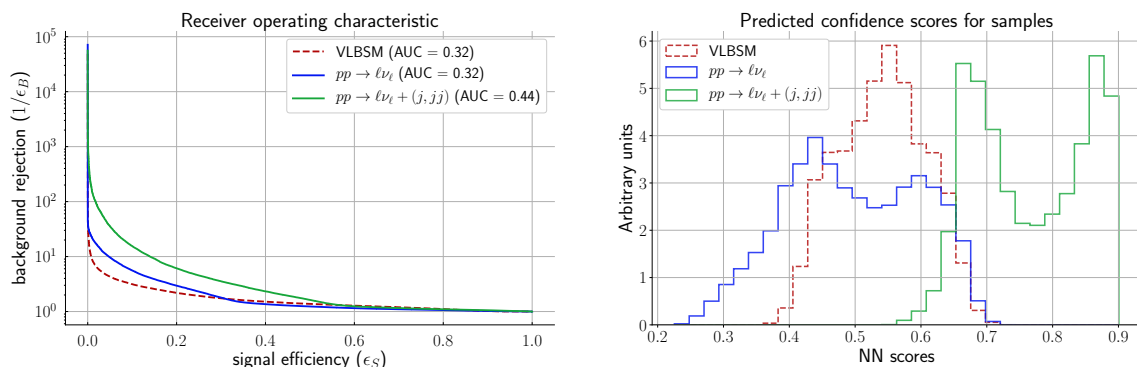
Figure 8. ROC and predicted confidence scores for each signal for a light VLL with mass $m_{e_4} = 677$ GeV and integrated luminosity $\mathcal{L} = 3000$ fb $^{-1}$. Signal events are represented by dashed curves in red. AUC denotes accuracy. The distributions are computed following an implementation of an evolutive algorithm that maximizes the accuracy metric.



(a) ZA topologies.



(b) VBF topologies.



(c) VLBSM topologies.

Figure 9. The same as in figure 8 but now computed for an evolutive algorithm that maximizes the Asimov significance.

excluded. Therefore, we compute the significance for all channels proposed so far showing our results in figures 10 and 11. While in the form the accuracy metric is maximized, in the latter it is the Asimov significance that is maximized. For the sake of rigour and completeness of our analysis we compute the significance for three distinct statistics:

1. First, we consider what we denote as *naive significance*. This is calculated solely by counting the number of background b and signal s events according to the well known formula $s/\sqrt{s+b}$.

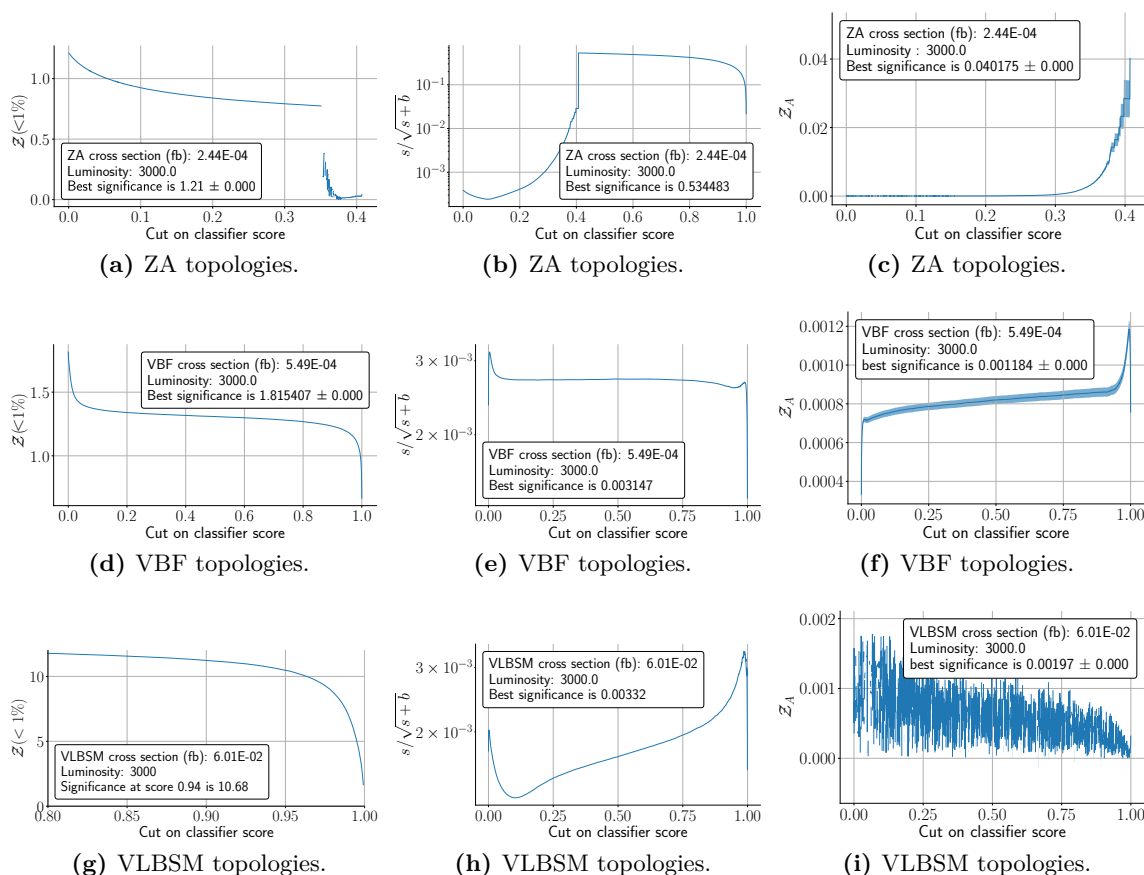


Figure 10. Significance as a function of the NN scores for each signal topology for a light VLL with mass $m_{e_4} = 677$ GeV and an integrated luminosity of $\mathcal{L} = 3000$ fb $^{-1}$. For plots (a), (d) and (g) we showcase the adapted Asimov significance where backgrounds are known with an error of up to 1%, for (b), (e) and (h) the naive significance $s/\sqrt{s+b}$ and for (c), (f) and (i) the Asimov significance. Plots are computed following an implementation of an evolutive algorithm that maximizes the accuracy metric.

2. The second metric to consider is the *plain* Asimov significance \mathcal{Z}_A as given in eq. (3.2). This is typically the most conservative measure in our analysis and it assumes that the background is known with 1% uncertainty.
3. Finally, we consider the Asimov significance in the case of backgrounds known with an uncertainty much smaller than 1% referring to it as $\mathcal{Z}(<1\%)$. In particular, we choose background uncertainty of $10^{-3}\%$ in our studies. This measure typically gives the best results but requires that all physics backgrounds are under control by the experiment. We expect this to be realizable by the ATLAS and CMS communities upon accumulated knowledge and experience over time. Note that in the limit of vanishing background uncertainty we recover the naive significance formula.

We then study these three significance measures in terms of the NN scores specializing to the case of the High-Luminosity (HL) LHC runs, i.e. $\mathcal{L} = 3000$ fb $^{-1}$. Considering

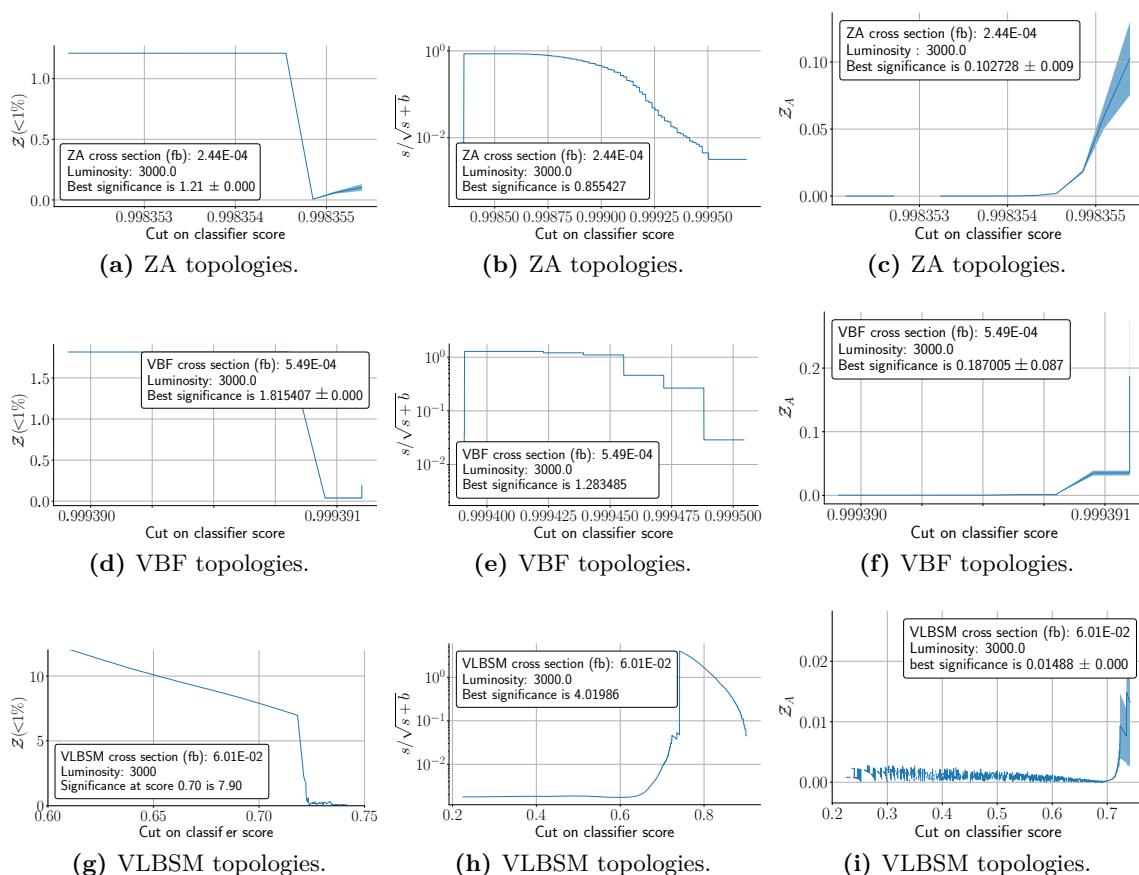


Figure 11. The same as in figure 10 but for an evolutive algorithm that maximizes the Asimov significance.

the results from both scenarios, we note that, as a general rule for all channels we get $\mathcal{Z}(< 1\%) > s/\sqrt{s+b} > \mathcal{Z}_A$. For the case of the VLBSM channel, note that we have not selected the highest significance that our algorithm has found. The reason for this is that we have regions of the NN parameter space where only signal is present. For example, in figure 10(g), we observe that for scores of ~ 0.80 we obtain a significance greater of 10σ . However, for a realistic evaluation, we have asked the NN to guarantee the existence of both signal and background events. This is not the case of the ZA and VBF channels where the largest significance is found in a region where both signal and background is always present.

Under the assumption that all signal topologies represent independent events, we can define the combined significance as the sum of all three contributions, i.e.

$$\sigma_C = \sigma_{ZA} + \sigma_{VBF} + \sigma_{VLBSM}. \quad (3.8)$$

For the results in figure 10 we see that if we privilege an evolutionary algorithm that looks for a better accuracy we get

- $\mathcal{Z}(< 1\%)$: $\sigma_C = 13.71\sigma$,
- $s/\sqrt{s+b}$: $\sigma_C = 0.55\sigma$,
- \mathcal{Z}_A : $\sigma_C = 0.04\sigma$.

while if we choose to maximize the Asimov significance, the results depicted in figure 11 are translated into

- $\mathcal{Z}(< 1\%)$: $\sigma_C = 10.93\sigma$,
- $s/\sqrt{s+b}$: $\sigma_C = 6.16\sigma$,
- \mathcal{Z}_A : $\sigma_C = 0.33\sigma$.

Note that for both metrics we surpass the 5σ threshold if the $\mathcal{Z}(< 1\%)$ measure is considered. However, note that this statistics works under the assumption that all backgrounds are very well under control. Nicely, for the case of the Asimov metric, we obtain for the naive significance $s/\sqrt{s+b} = 6.16\sigma$ providing us a stronger argument towards the possibility of probing VLLs with masses around 700 GeV perhaps even before the end of the HL-LHC runs. We also observe that an evolutive algorithm engineered to maximize the Asimov significance offers overall better results. We should mention here that the combined (or even individual) significance grows with luminosity. Therefore, and based on the results so far discussed, it provides a compelling argument in favour of high-luminosity machines in the longer term.

The results presented so far for a single point are already rather interesting not only in the context of the model formulation we discuss here but also for any other model with VLLs and sterile neutrinos. For completeness and better scrutiny we will study the impact of varying the mass of both the two lightest VLLs as well as of the lightest BSM neutrino. In particular, we are interested in understanding under which circumstances one can reach or surpass a signal significance of 5 standard deviations in order to motivate direct VLL searches for this class of models at the LHC. To do this we repeat the numerical procedure explained above considering the following cases:

1. varying the lightest VLL mass, m_{e_4} , between 200 GeV and 1.25 TeV while keeping m_{ν_4} the same as in eq. (3.3);
2. keeping the m_{e_4} fixed as in eq. (3.3) and varying m_{ν_4} between 100 keV and 100 MeV;
3. Varying the masses of the two lightest VLLs such that $m_{e_4} < m_{e_5}$;
4. the same as in 1 but with varying luminosity.

For a fixed luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$ we study the signal significance calculated for the VLL masses $m_{e_4} = 200, 486, 868$ and 1250 GeV. First, we consider an evolutive algorithm that maximizes the accuracy metric showing our results in table 7. We have repeated the same procedure for $m_{e_4} = 200$ and 486 GeV considering an evolutive algorithm that maximizes the Asimov significance. Our results can be found in table 8. From the aforementioned tables we notice that for heavy VLL scenarios with $m_{e_4} = 1.25$ TeV the combined significance for all event signals drops to values near zero indicating that such channels can be rather challenging for direct searches at the LHC. However, if the background is well under control, the $\mathcal{Z}(< 1\%)$ statistics offers a combined significance of 5.66σ . Note that the larger component of the combined significance results from the VLBSM channel with 4.9σ , which means that a potential observation of heavy VLLs with

Mass of e_4	$s/\sqrt{s+b}$			$\mathcal{Z}(<1\%)$			\mathcal{Z}_A		
	ZA	VBF	VLBSM	ZA	VBF	VLBSM	ZA	VBF	VLBSM
200 GeV	4.01	9.4×10^{-3}	0.31	12.18	2.83	12.95	2.05	2.47×10^{-3}	1.41
486 GeV	0.95	1.51	6.66×10^{-3}	2.59	2.13	7.83	0.12	4.6×10^{-4}	2.15×10^{-4}
677 GeV	0.53	3.15×10^{-3}	3.32×10^{-3}	1.21	1.82	10.68	0.040	1.18×10^{-3}	1.97×10^{-3}
868 GeV	0.26	0.93	6.18×10^{-4}	0.52	1.32	6.60	0.01	0.30	2.47×10^{-4}
1250 GeV	0.05	4.37×10^{-4}	1.20×10^{-4}	0.17	0.59	4.90	4.28×10^{-4}	2.05×10^{-4}	2.65×10^{-3}

Table 7. Signal significance for an evolutive algorithm that maximizes accuracy metric. All significances are computed for $\mathcal{L} = 3000 \text{ fb}^{-1}$.

Mass of e_4	$s/\sqrt{s+b}$			$\mathcal{Z}(<1\%)$			\mathcal{Z}_A		
	ZA	VBF	VLBSM	ZA	VBF	VLBSM	ZA	VBF	VLBSM
200 GeV	6.10	2.00	12.65	12.18	2.83	12.70	4.44	0.145	4.28
486 GeV	1.77	1.50	11.26	2.60	2.13	8.62	0.30	0.53	0.20
677 GeV	0.86	1.28	4.02	1.21	1.82	7.90	0.11	0.187	0.015

Table 8. Signal significance for an evolutive algorithm that maximizes Asimov significance metric. All significances are computed for $\mathcal{L} = 3000 \text{ fb}^{-1}$.

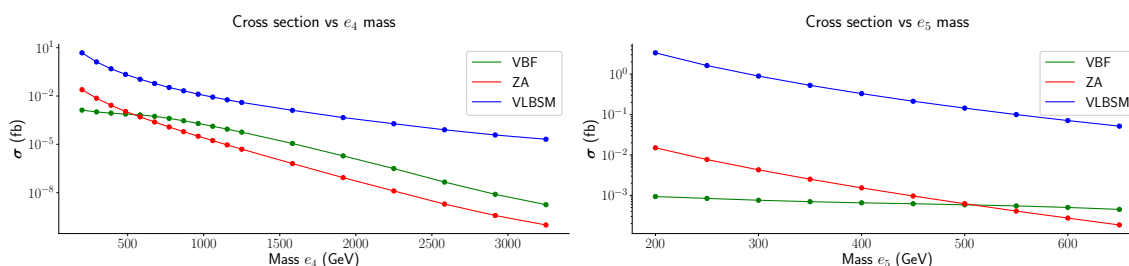


Figure 12. Total production and decay cross section (in femtobarn) as a function of e_4 (left) and e_5 masses (right). While on the left panel $m_{e_5} = 3.2 \text{ TeV}$, on the right one we have $m_{e_4} = 200 \text{ GeV}$.

masses in the TeV range can only be possible if the $pp \rightarrow \ell\nu_\ell + (0j, j, jj)$ backgrounds are known and with a high precision. The fast decrease in significance for larger masses is a consequence of the, also fast, decrease in cross-section with increasing mass, as shown in figure 12, left panel. A small significance is also obtained for the heavier VLLs, e_5 and e_6 , whose masses lie beyond $\mathcal{O}(3 \text{ TeV})$ and likely out of the reach of the HL-LHC.

However, recall that all three signal events represent independent variables, which means that we can evaluate a global significance as the sum of the individual ones from each process. This implies that we can consider additional event signals that would boost this global significance. In particular, this entails that including channels with jets from W decays to quarks can be relevant due to a larger expected number of events. In fact, the W decay width is larger for light jets with a branching ratio (BR) of approximately 67.4%, rather than for leptons, whose BR is 10.86% [56]. For the present case, and when the

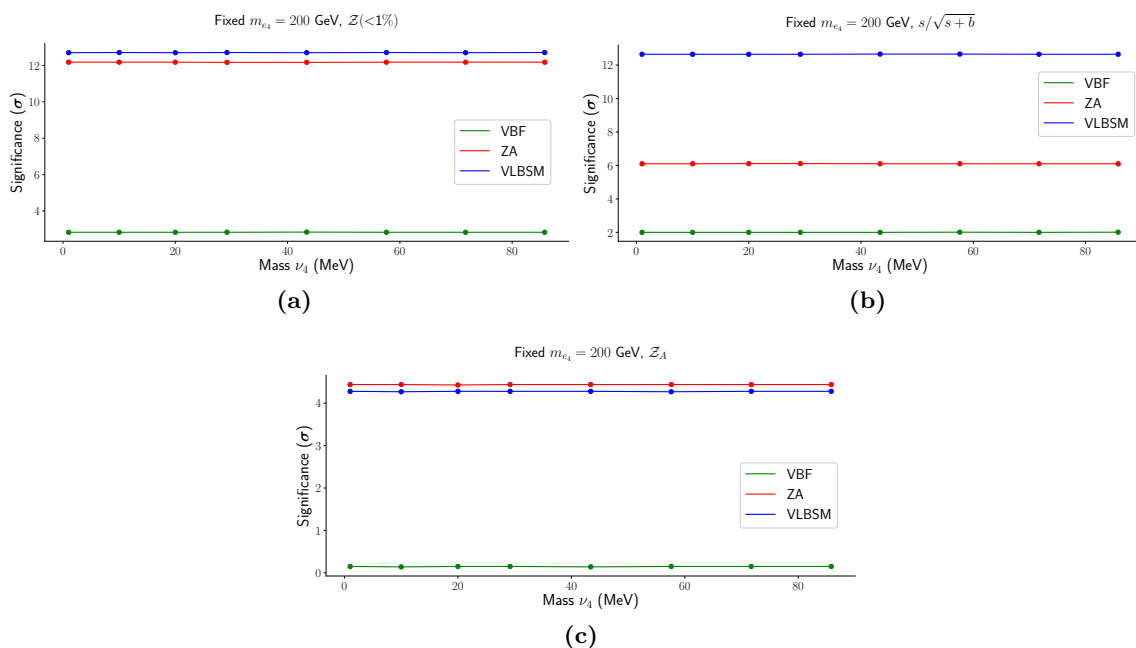


Figure 13. Significance as a function of the neutrino’s mass for a luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$. The significance is computed following an evolutive algorithm that maximises the Asimov metric.

accuracy metric maximisation is concerned, the combined significance at $m_{e_4} = 200$ GeV is 27.16σ for $\mathcal{Z}(< 1\%)$, 3.46σ for \mathcal{Z}_A and 4.33σ for the naive significance. If one instead maximizes the Asimov metric, the same benchmark point ($m_{e_4} = 200$ GeV) yields 27.71σ for $\mathcal{Z}(< 1\%)$, 8.57σ for \mathcal{Z}_A and 20.75σ for the naive significance. These results, and in particular those for the naive significance, indicate that a light VLL characteristic of our model is expected to have a strong presence for a high-luminosity run at the LHC and can be probed well before the end of the LHC programme.

In general, we note that for an ever increasing mass of e_4 , the overall significance reduces, as we have already discussed above. We also see that the rate of decrease is faster for the ZA channel rather than for the VBF and VLBSM ones. Note that VBF signals yield a maximum significance in $\mathcal{Z}(< 1\%)$ of 2.83σ for $m_{e_4} = 200$ GeV and for both the Asimov and accuracy metrics maximization. On the other hand, for the ZA and VLBSM channels with maximized accuracy, the same VLL mass yields a significance of 12.18σ for ZA events and 12.95σ for the VLBSM channel (see table 7). However, for results obtained upon maximization of the Asimov metric the naive significance can be as large as 6.10σ for ZA events and 12.65 for VLBSM ones (see table 8).

For completeness of information, we show in figure 13 that fixing $m_{e_4} = 200$ GeV and for $m_{\nu_4} < m_{\nu_5}$, the effect of varying the sterile neutrino mass, m_{ν_4} , is negligible for any value of the lightest BSM neutrino mass in the range 100 keV to 100 MeV. Note that, while our analysis is generic enough, due to a combination between the seesaw nature of neutrino masses and the radiative origin of Yukawa interactions in the low-scale SHUT model, these mass scales for ν_4 are realistic, compatible with the model’s structure and should be seriously considered.

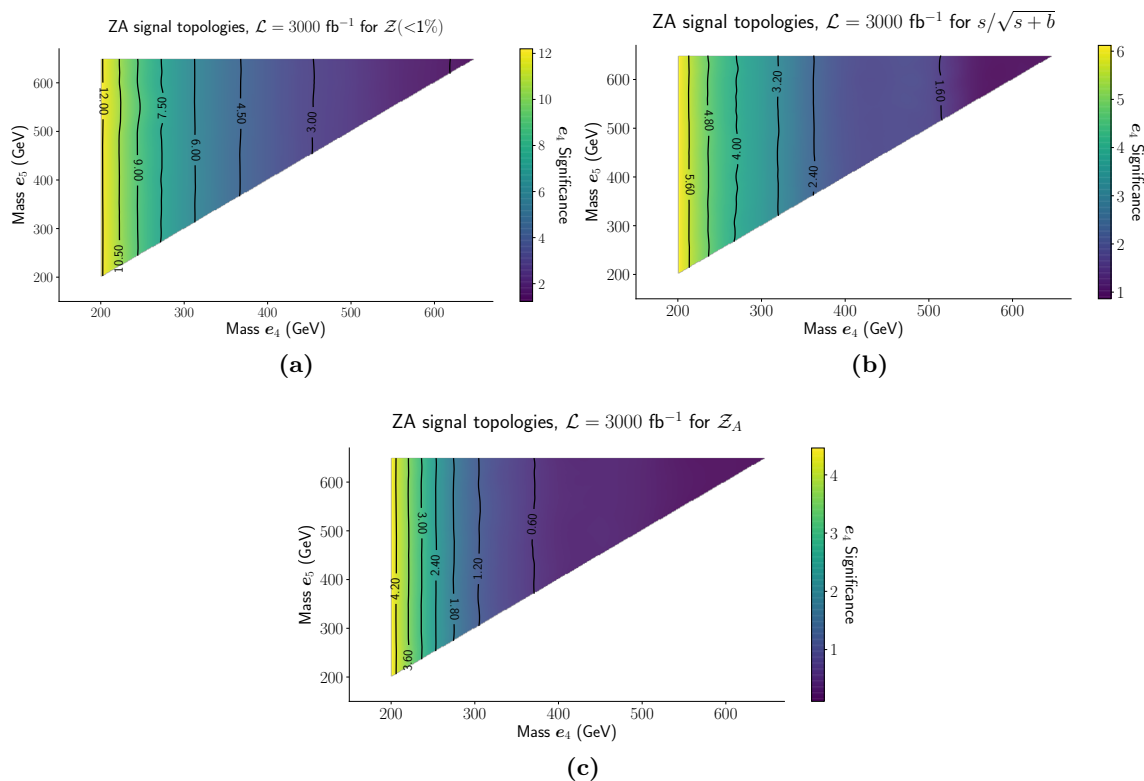


Figure 14. Significance contour plots for (a) $\mathcal{Z}(<1\%)$, (b) $s/\sqrt{s+b}$ and (c) \mathcal{Z}_A . The colour scale represents a ZA event signal for e_4 at a luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$. The significance is computed with an evolutive algorithm that maximises the Asimov metric.

Provided that lighter VLLs represent a more interesting case for forthcoming explorations at the LHC we will essentially focus our attention in the m mass range [200, 700] GeV for both e_4 and e_5 , in such a way that $m_{e_4} < m_{e_5}$. This choice is based on our discussion in section 2.3.1 just below eq. (2.16), where two light VLLs below 1 TeV order with a heavy one at around 5 TeV is a viable scenario. First, we fix the mass of the e_4 to be $m_{e_4} = 200 \text{ GeV}$, as this represents the case where we obtain the greatest significance, as shown in tables 7 and 8. We also consider the BSM neutrino to be in the hundreds of keV order. The results of these scans can be seen in figure 14 where we show e_4 significance contours in terms of m_{e_4} and m_{e_5} . One can immediately see that a varying e_5 mass has a very marginal impact on the e_4 significance, indicating that it is independent of m_{e_5} . This can be understood from the $e_5 \rightarrow W\nu_i$ decay branching fractions shown in table 9. In fact, we observe that the overall BRs do not suffer significant alterations with the varying e_5 mass, and as such, the computed e_4 production and decay cross-section remains very much the same. Therefore we do not expect visible changes in the significance and thus, the impact on the significance is understandably small.

So far all results have been computed for a luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$. However, the LHC is only scheduled to run at such luminosities around 2026-2030 [63]. Therefore it is equally relevant to study how the significance changes for lower luminosities, in particular,

	$m_{e_4} = 200 \text{ GeV}$ $m_{e_5} = 200 \text{ GeV}$	$m_{e_4} = 200 \text{ GeV}$ $m_{e_5} = 250 \text{ GeV}$	$m_{e_4} = 200 \text{ GeV}$ $m_{e_5} = 300 \text{ GeV}$
$e_5 \rightarrow W\nu_4$	0.5082513	0.5069336	0.5063337
$e_5 \rightarrow W\nu_5$	0.3111182	0.3119451	0.3123246
$e_5 \rightarrow W\nu_6$	0.1806305	0.1811213	0.1813417

Table 9. Branching fractions of e_5 decaying into W and BSM neutrinos for three different e_5 masses and fixed $m_{e_4} = 200 \text{ GeV}$.

for 300 fb^{-1} , which is planned to be delivered in the Run III, scheduled to start during 2021. With this in mind, we show in figures 15 and 16 the dependency of the significance over the projected luminosities for a maximized accuracy. Results for the case where the Asimov significance is maximized are shown in figures 17, 18 and 19.

Looking first at figures 15 and 16, for $\mathcal{L} = 300 \text{ fb}^{-1}$, we note that a signal significance at or beyond 5σ level is only achievable for a light VLL of 200 GeV. In particular, the combined significance is 8.84σ for $\mathcal{Z}(< 1\%)$. This means that, if all backgrounds introduced in section 3 are known with high precision, it will be possible to either discover or exclude a VLL with mass 200 GeV, possibly even before the end of Run III.

For the case of heavy VLLs, the $\mathcal{Z}(< 1\%)$ significance is no larger than 1.78σ for the case of $m_{e_4} = 1.25 \text{ TeV}$, where VLBSM signals provide the larger contribution. With these results in mind, we would only expect to observe such heavy states in a high luminosity run. In fact, since we are simulating proton-proton collisions at $\sqrt{s} = 14 \text{ TeV}$, at such high masses, pair production of VLLs is increasingly unlikely, meaning that we only expect such heavy states to become visible in either high-luminosity runs or at higher energy colliders.

These results confirm that light states are favored to be probed at Run III of the LHC. This is especially noticeable when we use an evolutive algorithm that maximizes the Asimov significance. In plots 17, 18 and 19, we obtain, for a VLL of 200 GeV, significances well above 5σ , at $\mathcal{L} = 300 \text{ fb}^{-1}$ in all three statistics. Therefore, a 200 GeV VLL characteristic of our model can already be probed by the LHC Run III. We also note that for a VLL of 486 GeV, we are already able to obtain a significance of 4.25σ for $\mathcal{Z}(< 1\%)$ and 4.64σ for $s/\sqrt{s+b}$. While the latter two do not pass a 5σ baseline, they already represent significant deviations from pure SM processes. We argue that the addition of new signals, such as the ones with jets as mentioned before, should offer the necessary boost to achieve 5σ . This type of argumentation can also apply, for example, for a VLL of 677 GeV where we read a combined significance of 3.45σ .

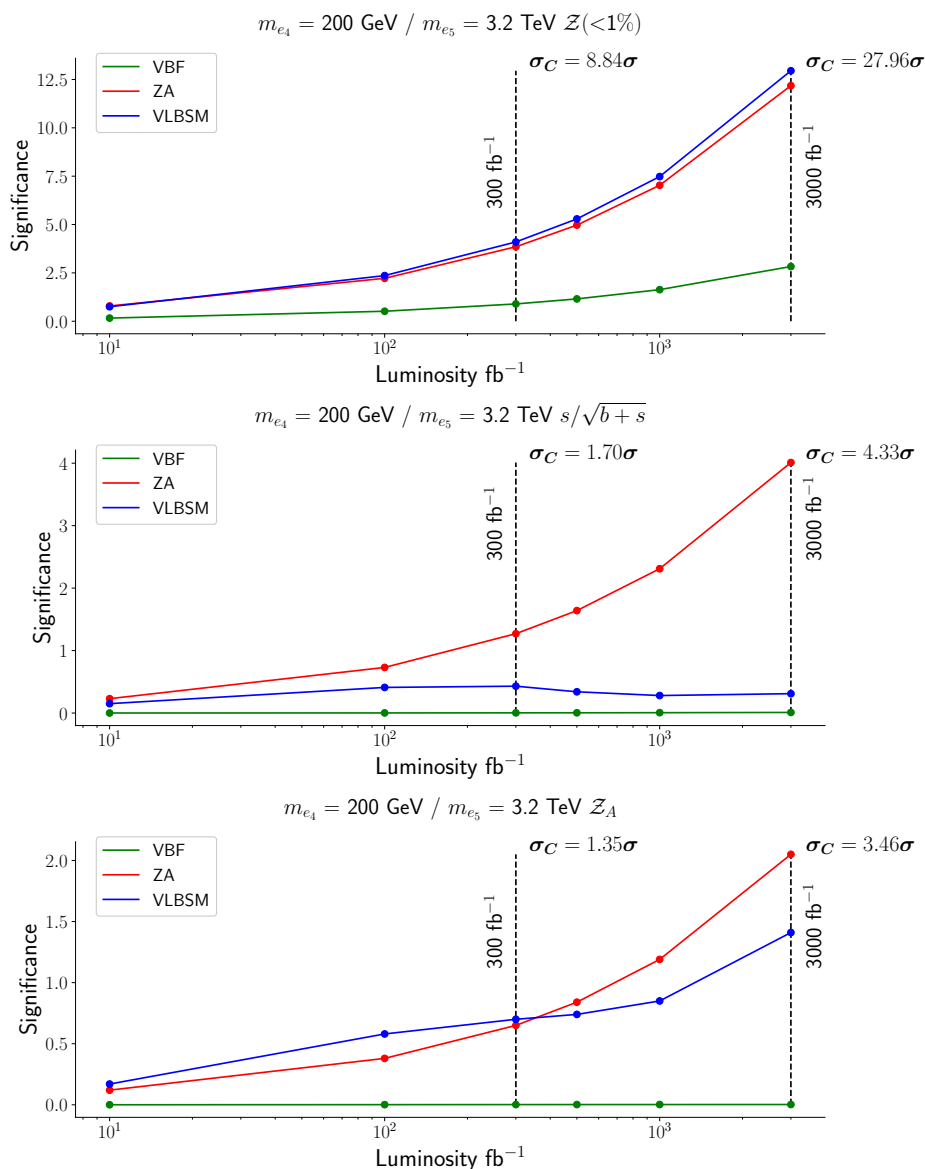


Figure 15. Significance as a function of the luminosity for different statistics and topologies that results from an evolutive algorithm that maximises accuracy. Here, x axis is in logarithmic scale for all plots and in the bottom plot the y axis is also in logarithmic scale. Green plots correspond to VBF signal events, red plots correspond to ZA signals and blue plots are representative of VLBSM topologies. The top plot corresponds to the Asimov significance where backgrounds are exactly known with a systematics of 1%, the middle plot is representative of the naive significance ($\sigma = s/\sqrt{s+b}$) and the bottom plot corresponds to the Asimov significance where backgrounds are not exactly known. The combined significance shown for $\mathcal{L} = 300 \text{ fb}^{-1}$ and $\mathcal{L} = 3000 \text{ fb}^{-1}$ is defined as $\sigma_C = \sigma_{\text{VBF}} + \sigma_{\text{ZA}} + \sigma_{\text{VLBSM}}$. The lightest VLL has mass of 200 GeV, while the second lightest has mass of 3.2 TeV.

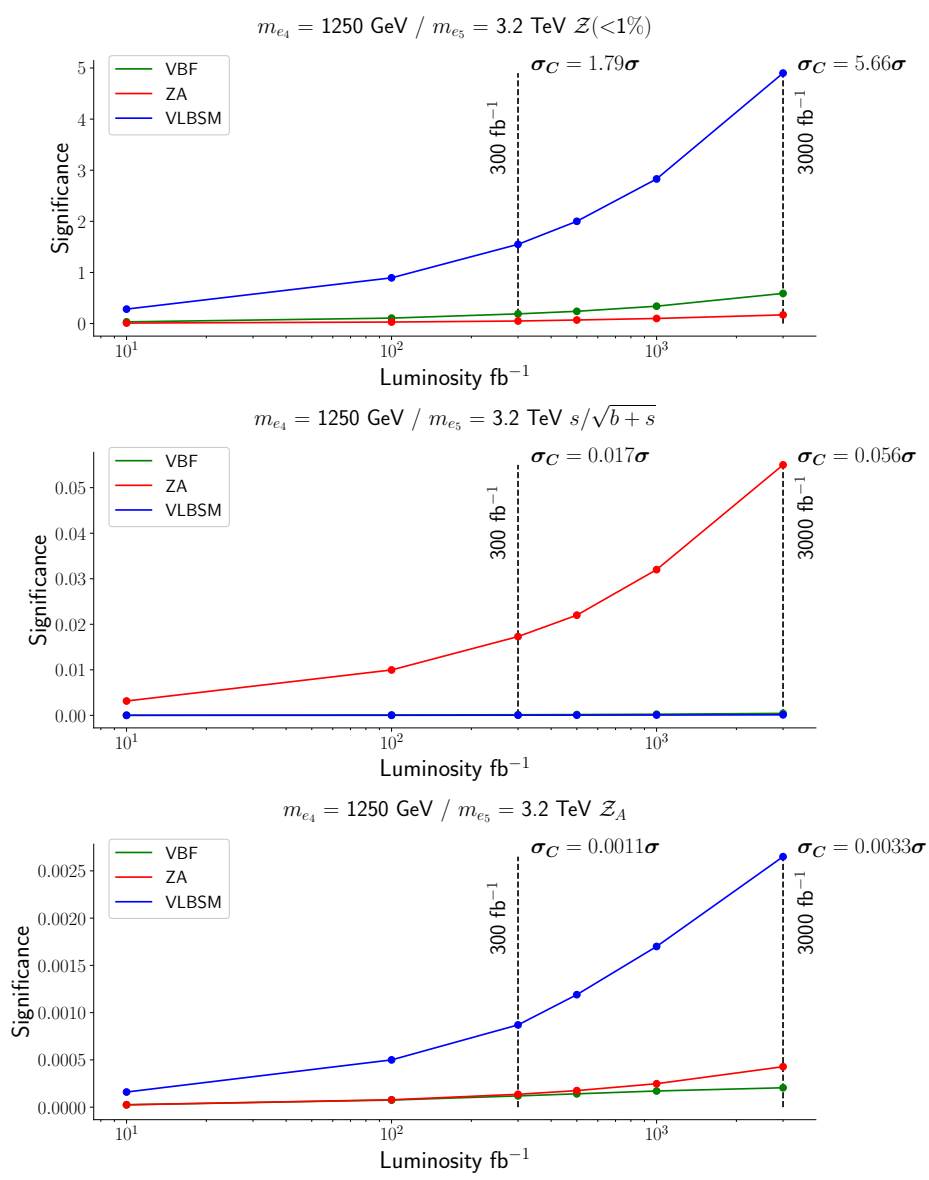


Figure 16. The same as in figure 15 but for a lightest VLL mass $m = 1250 \text{ GeV}$.

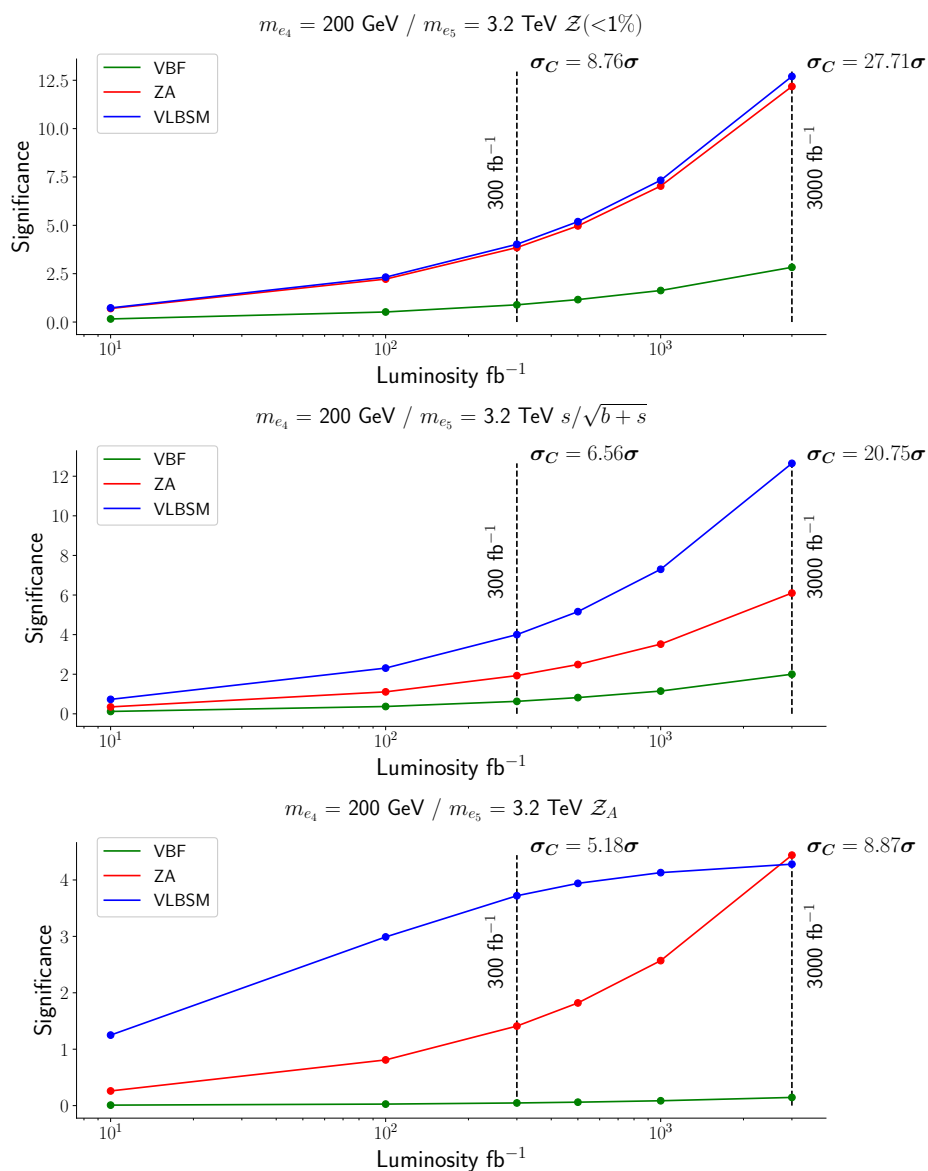


Figure 17. The same as in figure 15 but for an evolutive algorithm that maximizes the Asimov significance.

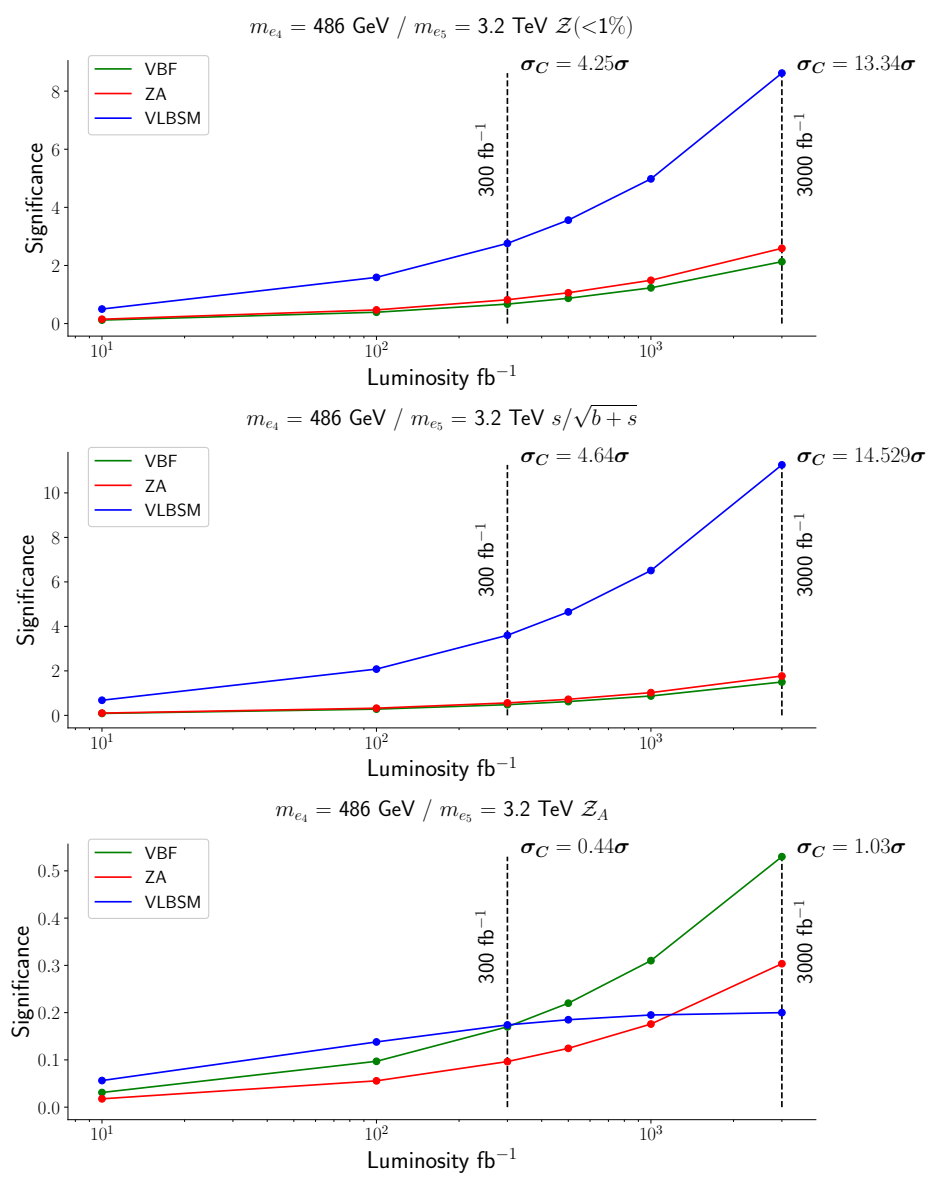


Figure 18. The same as in figure 17 but for a lightest VLL mass $m = 486 \text{ GeV}$.

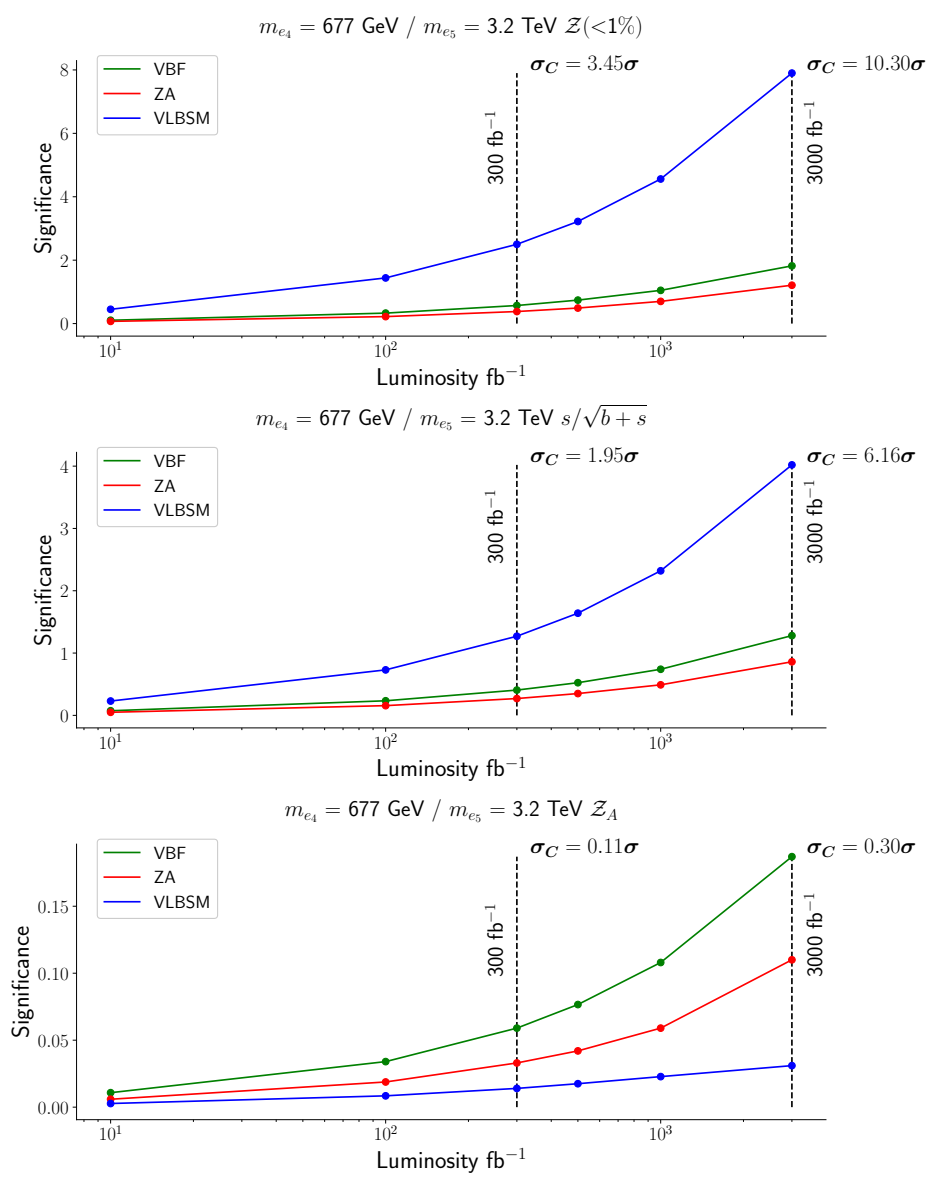


Figure 19. The same as in figure 17 but for a lightest VLL mass $m = 677 \text{ GeV}$.

4 Conclusions

In this work we have studied the collider phenomenology inherent to $SU(2)_L$ -doublet VLLs at the LHC, relevant for Run III and beyond. The properties of such exotic leptons were based on a framework built upon unification principles where the strong and EW interactions are ultimately unified with a local family symmetry. One of the major goals of the model under consideration is to offer a potential solution to the flavour problem, including a description for neutrino masses. It features a low-scale theory where new TeV-scale VLLs and VLQs are a natural consequence of the unification of Higgs and matter in common representations, sharing the same gauge and flavour quantum numbers.

We have performed Monte Carlo simulations relying on DL techniques with the aim of determining the statistical significance of an hypothetical VLL discovery at future LHC runs. We have also proposed a method to identify distinct MET cells from W^\pm leptonic decays by directly assessing the calorimeter's response and comparing it to Monte-Carlo simulated data, enabling us to clean the signal by completely removing background events. In this work, simple neural networks were considered, following the implementation of an evolutive algorithm that maximizes either the accuracy metric or the Asimov significance. For the first scenario, we are able to distinguish background events from signal events with an accuracy between 98% to 100%, depending in the signal topology in question, while the second scenario provides an increase in the Asimov significance at the cost of lower overall accuracy (32% to 70%).

We have proposed three distinct signatures for VLLs in our model which can be searched for at the LHC in the ZA, VBF and VLBSM channels with purely leptonic final states. Three distinct statistical significances were subject of our analysis, namely, the Asimov significance \mathcal{Z}_A , an adapted version of the Asimov significance $\mathcal{Z}(< 1\%)$ and the well known *naive* significance $s/\sqrt{s+b}$. A combined result of 27.71σ for $\mathcal{Z}(< 1\%)$, 20.75σ for $s/\sqrt{s+b}$ and finally 8.87σ for \mathcal{Z}_A was obtained for a luminosity of 3000 fb^{-1} , a center of mass beam energy $\sqrt{s} = 14 \text{ TeV}$, a VLL mass $m_{e_4} = 200 \text{ GeV}$ and a maximized Asimov metric significance. Under the same conditions, but for an accuracy metric search we have obtained a combined significance of 27.96σ for $\mathcal{Z}(< 1\%)$, 4.33σ for $s/\sqrt{s+b}$ and 3.46σ for \mathcal{Z}_A . In this mass range, the ZA and VLBSM channels provide the dominant contributions. Note that we have also considered relatively light BSM neutrinos with a mass of the order of $\mathcal{O}(100 \text{ keV})$. Furthermore, we have shown that varying the mass of the lightest BSM neutrino up to $\mathcal{O}(100 \text{ MeV})$ has a residual effect on the significance. As expected, the luminosity also has a noticeable impact on the significance. In particular, a value at or above 5σ can be achieved for $\mathcal{L} = 300 \text{ fb}^{-1}$, meaning that a 200 GeV VLL as predicted in our model can already be probed at the LHC Run-III. However, for larger VLL masses, and in particular for $m_{e_4} = 1.25 \text{ TeV}$, we have observed that a combined 5σ significance in the fully leptonic channels can only be achieved for $\mathcal{L} = 3000 \text{ fb}^{-1}$, that is, at the high-luminosity LHC, and with a combined significance of 5.66σ if the backgrounds are known with a high precision.

A scenario with two light VLLs, both below 650 GeV, was also studied, where an identical significance for an e_4 discovery was achieved. This follows from a negligible effect played by different m_{e_5} masses on the decay branching fractions and thus on the signal cross-section.

For all studied cases one has observed that the significance quickly drops if the VLL masses lie beyond 1 TeV. One of the first steps beyond the work presented here is to add jets to the final states from W decays into light quarks. Since such channels offer larger branching ratios it is likely possible that the significance for higher masses can be increased. Adding these channels would possibly increase the significance of the region around 1 TeV, but also provide a decisive boost for the $m_{e_4} = 486\text{GeV}$ scenario, where significances for Run III luminosities over 4σ are already obtained. As such, we conclude that VLLs in the mass range $\sim [200, 500]$ GeV are under the capabilities of being either discovered or excluded before the end of Run III.

Based on the model under consideration, the observation of VLLs at the reach of forthcoming LHC runs can offer a crucial probe to falsify our model and obtain hints about the high scale dynamics. For example, if the New Physics scale above the EW one, which was defined by the p , f and ω VEVs, is of the order 100 TeV, the e_4 mass as given in (2.16) would imply that the radiatively generated Yukawa couplings $\kappa_{5,8}$ need to be approximately $\mathcal{O}(10^{-2.7})$. However, one should comment here that an even stronger link to the high-scale can be obtained through the study of VLQs due to the tree-level nature of their masses. In fact, while for VLLs there is still some degree of uncertainty steaming from a non-trivial functional dependency of the κ_i parameters with masses and couplings, leading contributions to the two lightest VLQ masses are well understood and are proportional to the Yukawa coupling $\mathcal{Y}_2 \sim \mathcal{O}(10^{-2})$. In turn, this would allow us to fix the $\omega \sim f$ and p scales establishing a direct link to the scale where larger symmetries are broken. Having said this, performing collider phenomenology studies for VLQs using similar methods to those employed in this work is one of our key priorities for the near future. A combined study of both VLQ and VLL sectors can offer us a rather complete information about the ω , f and p scales, the sizes of the radiatively generated Yukawa couplings and the physics involved at such high energies.

Other important studies to perform concern Higgs and flavour physics which, due to the presence of three $SU(2)_L$ scalar doublets, is highly relevant. In fact, we have chosen a basis where the SM lepton sector has zero mixing with other BSM fermions. However, a more complete approach should consider small deviations from this limit up to flavour physics constraints. Last but not least, the presence of keV-MeV scale neutrinos can potentially offer a DM candidate if, in a basis with a more generic neutrino mixing, it is stable enough. In the longer term, with all such phenomenological studies, we can determine more precisely what are the viable regions of the parameter space that will help us in performing a direct matching between the low-scale and the high-scale regimes of the theory.

Acknowledgments

The authors want to thank António Onofre for rather thorough and insightful discussions about the subjects addressed in this manuscript. APM and FFF are supported by the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT — Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. APM, FFF and JG

are supported by the project PTDC/FIS-PAR/31000/2017. APM is also supported by the projects CERN/FIS-PAR/0027/2019, CERN/FISPAR/0002/2017 and by national funds (OE), through FCT, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19. R.P. is supported in part by the Swedish Research Council grants, contract numbers 621-2013-4287 and 2016-05996, as well as by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 668679).

A Low-scale SHUT model Lagrangian and Feynman rules

In this appendix the tree-level Lagrangian and Feynman rules between the EW gauge bosons and leptons are presented. To simplify some notation, projection operators are defined as

$$P_R = \frac{1 + \gamma^5}{2}, \quad P_L = \frac{1 - \gamma^5}{2} \quad (\text{A.1})$$

The Yukawa and fermion bilinear interactions are presented in section 2.3 in eqs. (2.7) and (2.8). Here we write the remaining Lagrangian terms of the low-scale SHUT model by considering all renormalizable, Lorentz and gauge invariant operators. We start by writing out all kinetic terms for the fermions,

$$\begin{aligned} \mathcal{L}_{\text{kin,f}} = & i(\bar{Q}_L)^i \not{D}(Q_L)_i + i(\bar{L})^i \not{D}(L)_i + i(\bar{d}_R)^i \not{D}(d_R)_i + i(\bar{u}_R)^i \not{D}(u_R)_i + i(\bar{e}_R)^i \not{D}(e_R)_i + \\ & + i(\bar{E}_L)^i \not{D}(E_L)_i + i(\bar{E}_R)^i \not{D}(E_R)_i + i(\bar{D}_L)^i \not{D}(D_L)_i + i(\bar{D}_R)^i \not{D}(D_R)_i \\ & + i(\bar{\nu}_R)^i \not{D}(\nu_R)_i, \end{aligned} \quad (\text{A.2})$$

where repeated index i represents summation over the different generations. The covariant derivative is defined as⁷

$$D_\mu = \partial_\mu - ig \frac{Y}{2} B_\mu - ig_w \frac{\sigma_a}{2} A_\mu^a - ig_s \frac{\lambda_a}{2} G_\mu^a. \quad (\text{A.3})$$

The kinetic terms for the bosonic sector reads

$$\mathcal{L}_{\text{bos,f}} = -\frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} A_b^{\mu\nu} A_{\mu\nu}^b - \frac{1}{4} G_c^{\mu\nu} G_{\mu\nu}^c + \frac{1}{2} (D_\mu \phi_a)(D^\mu \phi^a)^\dagger, \quad (\text{A.4})$$

where b and c represent $SU(2)_L$ and $SU(3)_C$ adjoint indices, while a denotes scalar generations. The scale potential is that of a generic 3HDM model and reads as:

$$V(\phi, \phi^\dagger) = (m_i)^2 |\phi^i|^2 + \left(m_{ij}^2 \phi^i (\phi^j)^\dagger + \text{H.c.} \right) + \lambda_{ijkl} \left(\phi^i (\phi^j)^\dagger \phi^k (\phi^l)^\dagger + \text{H.c.} \right). \quad (\text{A.5})$$

⁷We have a quite strong abuse of language here. In fact, the covariant derivate for the term $i(\bar{Q}_L)^i \not{D}(Q_L)_i$ is different than, for example, $i(\bar{E}_L)^i \not{D}(E_L)_i$. That is because quarks couple to gluons, while the leptons do not, so the covariant derivative for leptons *does not* have the last term of (A.3). One should interpret the definition in this fashion, that is, if it interacts, it exists, if it does not, it does not exist.

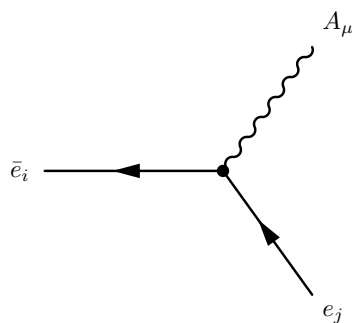
The full Lagrangian density for the effective low-energy 3HDM is the sum of all previous sectors and reads as

$$\mathcal{L}_{3\text{HDM}} = \mathcal{L}_{\text{kin},f} + \mathcal{L}_{\text{bos},f} + \mathcal{L}_y + \mathcal{L}_{\text{bil}} - V(\phi, \phi^\dagger). \quad (\text{A.6})$$

Expanding and rotating (A.6) to the mass basis one arrives to the following Feynman rules:

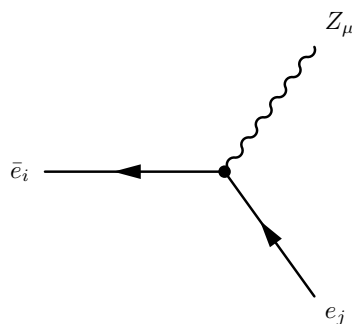
- **Lepton and Gauge bosons interactions**

- Charged Leptons — Photon vertex



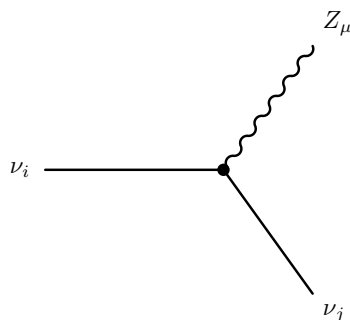
$$\begin{aligned} \frac{i}{2} \delta_{ij} (g_1 \cos \theta_W + g_2 \sin \theta_W) (\gamma^\mu \cdot P_L) + \frac{i}{2} \left(2g_1 \cos \theta_W \sum_{a=1}^3 U_{R,ia}^{e,*} U_{R,ja}^e \right. & (\text{A.7}) \\ \left. + (g_1 \cos \theta_W + g_2 \sin \theta_W) \sum_{a=1}^3 U_{R,i3+a}^{e,*} U_{R,j3+a}^e \right) (\gamma^\mu \cdot P_R) \end{aligned}$$

- Charged Leptons — Z boson vertex



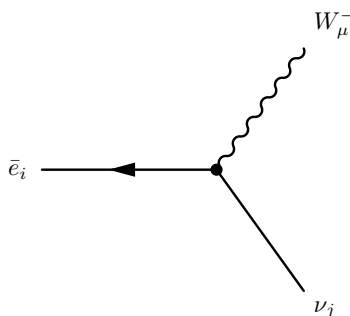
$$\begin{aligned} \frac{i}{2} \delta_{ij} (-g_1 \sin \theta_W + g_2 \cos \theta_W) (\gamma^\mu \cdot P_L) + \frac{i}{2} \left(2g_1 \sin \theta_W \sum_{a=1}^3 U_{R,ia}^{e,*} U_{R,ja}^e \right. & (\text{A.8}) \\ \left. + (g_1 \sin \theta_W - g_2 \cos \theta_W) \sum_{a=1}^3 U_{R,i3+a}^{e,*} U_{R,j3+a}^e \right) (\gamma^\mu \cdot P_R) \end{aligned}$$

– Neutrinos — Z boson vertex



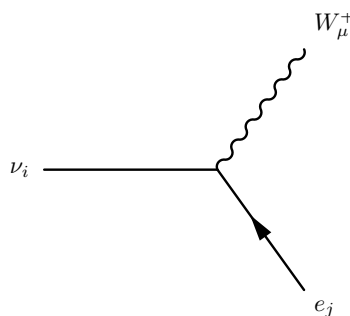
$$\begin{aligned} & \frac{i}{2} (g_1 \sin \theta_W + g_2 \cos \theta_W) \left(- \sum_{a=1}^6 U_{\nu,ja}^* U_{\nu,ia} + \sum_{a=1}^3 U_{\nu,j6+a}^* U_{\nu,i6+a} \right) (\gamma^\mu \cdot P_L) \quad (\text{A.9}) \\ & + -\frac{i}{2} (g_1 \sin \theta_W + g_2 \cos \theta_W) \left(- \sum_{a=1}^6 U_{\nu,ia}^* U_{\nu,ja} + \sum_{a=1}^3 U_{\nu,i6+a}^* U_{\nu,j6+a} \right) (\gamma^\mu \cdot P_R) \end{aligned}$$

– Neutrinos — Charged Leptons — W^- boson vertex



$$\begin{aligned} & -i \frac{1}{\sqrt{2}} g_2 \sum_{a=1}^6 U_{\nu,ja}^* U_{L,ia}^e (\gamma^\mu \cdot P_L) \\ & + -i \frac{1}{\sqrt{2}} g_2 \sum_{a=1}^3 U_{R,i3+a}^{e,*} U_{\nu,j6+a} (\gamma^\mu \cdot P_R) \end{aligned} \quad (\text{A.10})$$

– Neutrinos — Charged Leptons — W^+ boson vertex



$$\begin{aligned} & -i \frac{1}{\sqrt{2}} g_2 \sum_{a=1}^6 U_{L,ja}^{e,*} U_{\nu,ia} (\gamma^\mu \cdot P_L) \\ & + -i \frac{1}{\sqrt{2}} g_2 \sum_{a=1}^3 U_{\nu,i6+a}^* U_{R,j3+a}^e (\gamma^\mu \cdot P_R) \end{aligned} \quad (\text{A.11})$$

B Kinematic and angular variable for all topologies with detector effects

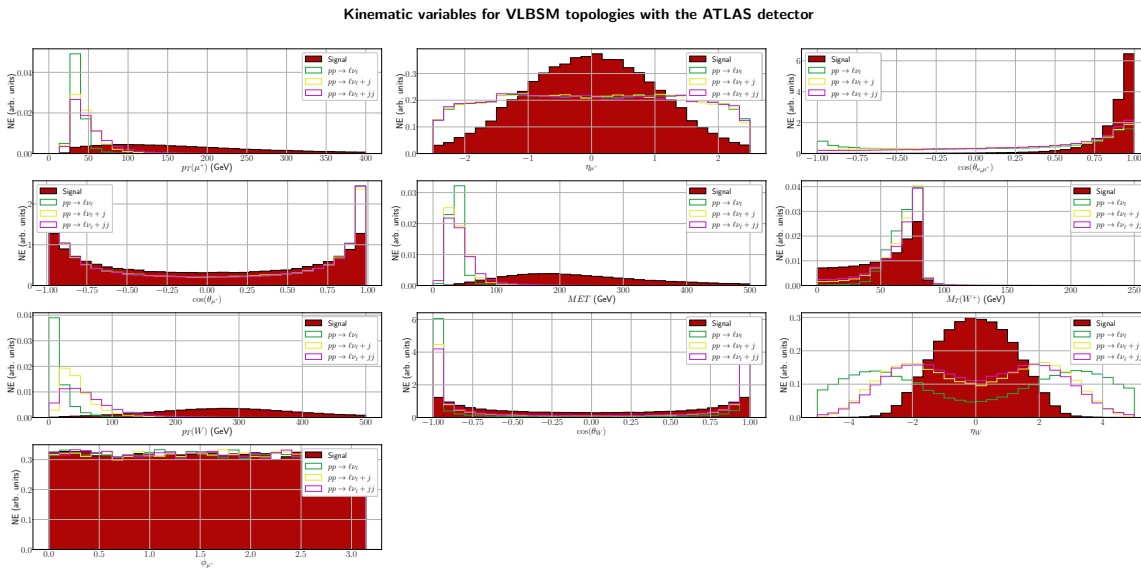


Figure 20. Dimensionless (angular) and dimension-full (kinematic) observables at lab reference frame for VLBSM channel (solid red), with $pp \rightarrow \ell\nu_\ell$ (green line), $pp \rightarrow \ell\nu_\ell + j$ (yellow line) and $pp \rightarrow \ell\nu_\ell + jj$ (purple line) backgrounds where it is considered 30 bins for all histograms. From top left to bottom right, we have distributions for transverse momentum μ^+ , pseudo-rapidity μ^+ , $\cos(\theta_{\mu^+})$, $\cos(\theta_{\nu_\mu^+ \mu^+})$, MET, transverse mass for W , transverse momentum for W , $\cos(\theta_W)$, pseudo-rapidity for W and azimuthal angle for μ^+ . Note that ν is the reconstructed missing energy vector, according to the discussion in section 3.

Kinematic variables for ZA topologies with the ATLAS detector

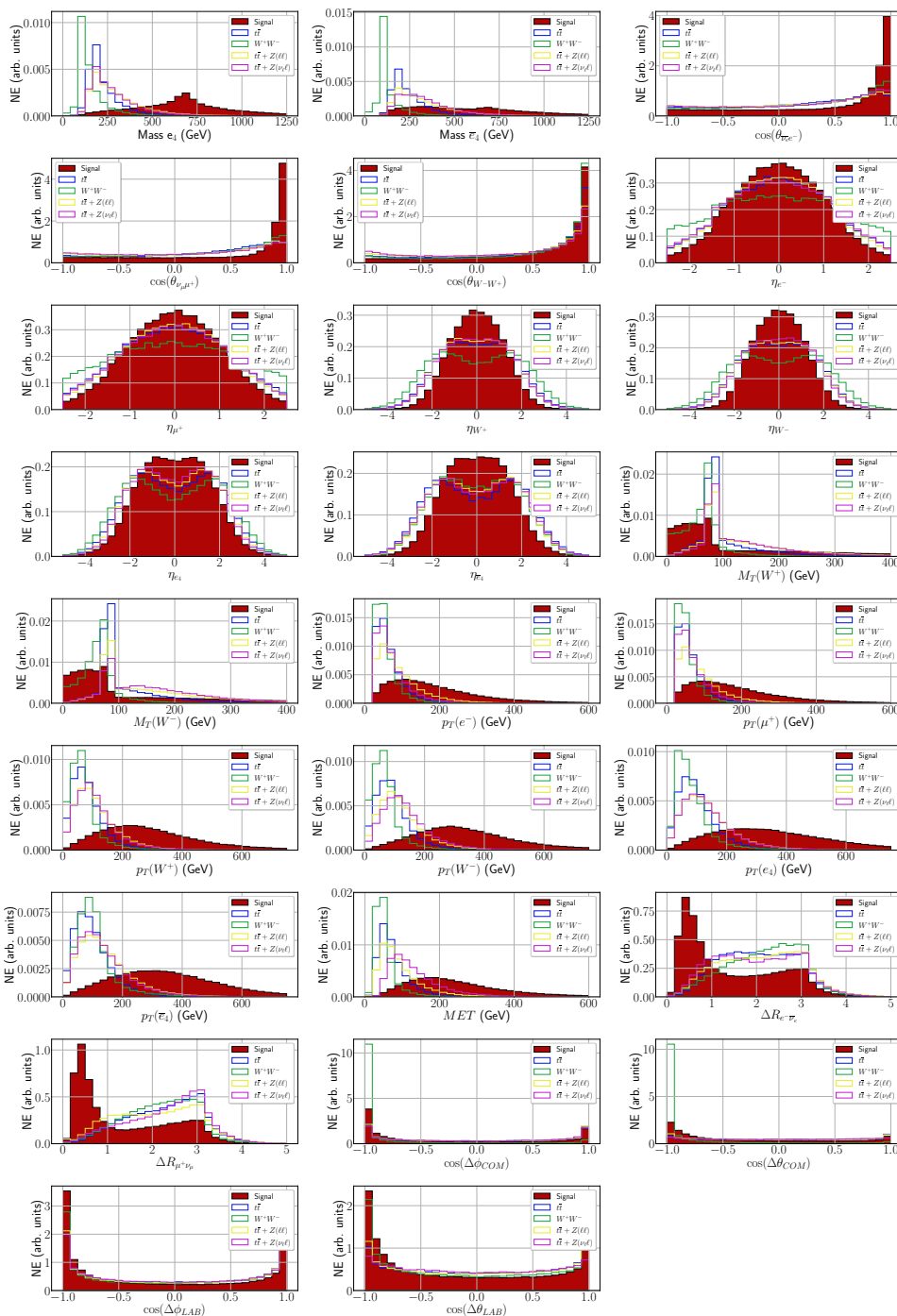


Figure 21. Dimensionless (angular) and dimension-full (kinematic) observables at lab reference frame for ZA channel (solid red), with $t\bar{t}$ (blue line), W^+W^- (green line), $t\bar{t} + Z(\ell^+\ell^-)$ (yellow line) and $t\bar{t} + Z(\nu\ell)$ (purple line) backgrounds where it is considered 30 bins for all histograms. From top left to bottom right, we have distributions for e_4 and \bar{e}_4 mass, $\cos(\theta_{\bar{\nu}_e e^-})$, $\cos(\theta_{\nu\mu^+})$, $\cos(\theta_{W^+W^-})$, pseudo-rapidity for e^- , μ^+ , W^+ , W^- , e_4 and \bar{e}_4 , transverse mass for W^+ and W^- , transverse momentum for e^- , μ^+ , W^+ , W^- , e_4 and \bar{e}_4 , MET, $\Delta R_{e^-\bar{\nu}_e}$, $\Delta R_{\mu^+\nu_\mu}$, $\cos(\Delta\phi)$ and $\cos(\Delta\theta)$ in the lab and e_4/\bar{e}_4 CM frames. Note that ν is the reconstructed missing energy vector, according to the discussion in section 3.

Kinematic variables for VBF topologies with the ATLAS detector

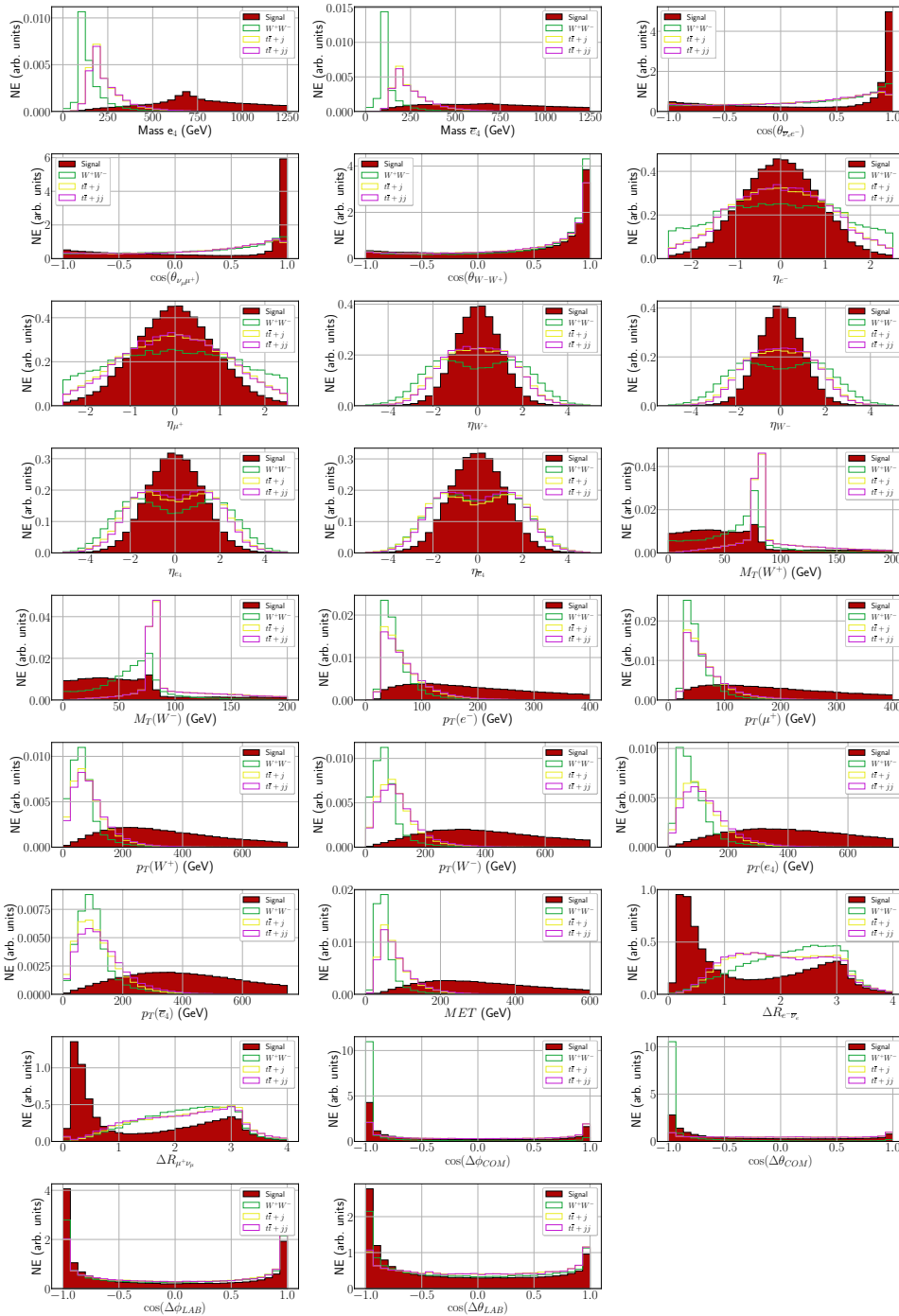


Figure 22. Dimensionless (angular) and dimension-full (kinematic) observables at lab reference frame for VBF channel (solid red), with W^+W^- (green line), $t\bar{t} + j$ (yellow line) and $t\bar{t} + jj$ (purple line) backgrounds where it is considered 30 bins for all histograms. From top to bottom left to bottom right, we have distributions for e_4 and \bar{e}_4 mass, $\cos(\theta_{\bar{\nu}_e e^-})$, $\cos(\theta_{\nu_\mu \mu^+})$, $\cos(\theta_{W^+ W^-})$, pseudo-rapidity for e^- , μ^+ , W^+ , W^- , e_4 and \bar{e}_4 , transverse mass for W^+ and W^- , transverse momentum for e^- , μ^+ , W^+ , W^- , e_4 and \bar{e}_4 , MET, $\Delta R_{e^- \bar{\nu}_e}$, $\Delta R_{\mu^+ \nu_\mu}$, $\cos(\Delta\phi)$ and $\cos(\Delta\theta)$ in the lab and e_4/\bar{e}_4 CM frame. Note that ν is the reconstructed missing energy vector, according to the discussion in section 3.

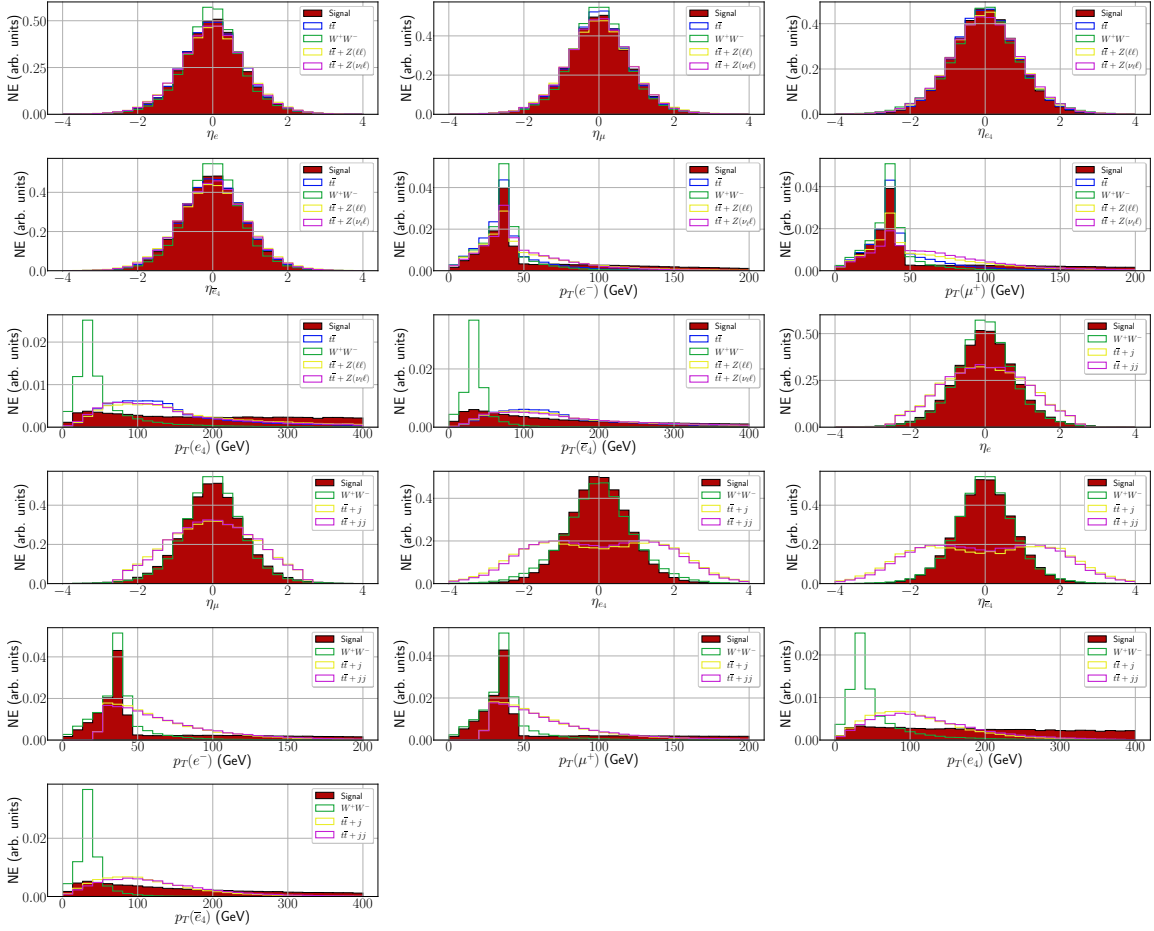


Figure 23. Dimensionless (angular) and dimension-full (kinematic) observables at the W^\pm reference frame for VBF and ZA channels (solid red) where it is considered 30 bins for all histograms. VBF channel correspond to plots with 3 distinct backgrounds, while ZA has 4. The same backgrounds from previous plots also applies here. From top left to bottom right, for both channels, we have distributions for pseudo-rapidity for e^- , μ^+ , e_4 and \bar{e}_4 and transverse momentum for e^- , μ^+ , e_4 and \bar{e}_4 . Note that ν is the reconstructed missing energy vector, according to the discussion in section 3.

C Neural Network models for different masses

Mass	ZA
200 GeV	<p><u>Layers</u> : 1 input + 2 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1,2,3) and none (for layer 4)</p> <p><u>Activation function</u> : tanh (for layers 1,2,3) and sigmoid (for layer 4)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1,2,3) in fan_in mode and uniform distribution (layer 4) in fan_avg mode</p>
486 GeV	<p><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p><u>Activation function</u> : elu (for layers 1 until 5) and sigmoid (for layer 6)</p>
677 GeV	<p><u>Layers</u> : 1 input + 1 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 and 2) and none (for layer 3)</p> <p><u>Activation function</u> : tanh (for layers 1 and 2) and sigmoid (for layer 3)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 and 2) in fan_in mode and uniform distribution (layer 3) in fan_avg mode</p>
868 GeV	<p><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p><u>Activation function</u> : elu (for layers 1 to 5) and sigmoid (for layer 6)</p> <p><u>Initializer</u> : RandomNormal (for layers 1 until 5) and VarianceScaling with uniform distribution (layer 6) in fan_avg mode</p>
1250 GeV	<p><u>Layers</u> : 1 input + 1 hidden + 1 output. Hidden and input layers with 512 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1,2) and none (for layer 3)</p> <p><u>Activation function</u> : tanh (for layers 1,2) and sigmoid (for layer 3)</p> <p><u>Initializer</u> : VarianceScaling, with uniform distribution (for layers 1,2) in fan_in mode and uniform distribution (layer 3) in fan_avg mode</p>

Table 10. Neural networks architectures employed for each mass of the lightest VLL for ZA channel. The architecture is determined following the implementation of an evolutive algorithm that maximizes the accuracy metric.

Mass	VBF
200 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 512 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 4) and none (for layer 5)</p> <p><u>Activation function</u> : elu (for layers 1 to 4) and sigmoid (for layer 5)</p> <p><u>Initializer</u> : VarianceScaling, with uniform distribution (for layers 1 to 4) in fan_in mode and uniform distribution (layer 5) in fan_avg mode</p>
486 GeV	<p><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 512 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p><u>Activation function</u> : relu (for layers 1 until 5) and sigmoid (for layer 6)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 5) in fan_in mode and uniform distribution (layer 6) in fan_avg mode</p>
677 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 and 4) and none (for layer 5)</p> <p><u>Activation function</u> : sigmoid (for layers 1 to 5)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 and 4) in fan_in mode and uniform distribution (layer 5) in fan_avg mode</p>
868 GeV	<p><u>Layers</u> : 1 input + 2 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 3) and none (for layer 4)</p> <p><u>Activation function</u> : sigmoid (for layers 1 to 4)</p> <p><u>Initializer</u> : VarianceScaling, with uniform distribution (for layers 1 and 3) in fan_in mode and uniform distribution (layer 3) in fan_avg mode</p>
1250 GeV	<p><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p><u>Activation function</u> : tanh (for layers 1 to 5) and sigmoid (for layer 6)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 5) in fan_in mode and uniform distribution (layer 3) in fan_avg mode</p>

Table 11. Neural networks architectures employed for each mass of the lightest VLL for VBF channel. The architecture is determined following the implementation of an evolutive algorithm that maximizes the accuracy metric.

Mass	VLBSM
200 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 2048 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 4) and none (for layer 5)</p> <p><u>Activation function</u> : Sigmoid (for layers 1 to 5)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 4) in fan_in mode and uniform distribution (layer 5) in fan_avg mode</p>
486 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 1024 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 4) and none (for layer 5)</p> <p><u>Activation function</u> : relu (for layers 1 until 4) and sigmoid (for layer 5)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (layer 1) in fan_in mode and uniform distribution (for layers 2 to 5) in fan_avg mode</p>
677 GeV	<p><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 and 5) and none (for layer 6)</p> <p><u>Activation function</u> : relu (for layers 1 to 5) and Sigmoid (layer 6)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 and 5) in fan_in mode and uniform distribution (layer 6) in fan_avg mode</p>
868 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 4) and none (for layer 5)</p> <p><u>Activation function</u> : tanh (for layers 1 to 4) and sigmoid (layer 5)</p> <p><u>Initializer</u> : RandomNormal (for layers 1 to 4) and VarianceScaling with uniform distribution (layer 5) in fan_avg mode</p>
1250 GeV	<p><u>Layers</u> : 1 input + 1 output. Input layers with 256 neurons, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (layer 1) and none (layer 2)</p> <p><u>Activation function</u> : elu (layer 1) and sigmoid (layer 2)</p> <p><u>Initializer</u> : VarianceScaling, with uniform distribution (layer 1) in fan_in mode and uniform distribution (layer 2) in fan_avg mode</p>

Table 12. Neural networks architectures employed for each mass of the lightest VLL for VLBSM. The architecture is determined following the implementation of an evolutive algorithm that maximises the accuracy metric.

Mass	ZA
200 GeV	<p><u>Layers</u> : 1 input + 1 output. Input layer with 256 neurons, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layer 1) and none (for layer 2)</p> <p><u>Activation function</u> : Sigmoid (for layers 1 and 2)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (layer 1) in fan_in mode and uniform distribution (layer 2) in fan_avg mode</p>
486 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 512 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 4) and none (for layer 5)</p> <p><u>Activation function</u> : sigmoid (for layers 1 until 5)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 4) in fan_in mode and uniform distribution (for layer 5) in fan_avg mode</p>
677 GeV	<p><u>Layers</u> : 1 input + 3 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 5 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 and 4) and none (for layer 5)</p> <p><u>Activation function</u> : sigmoid (for layers 1 to 5)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 and 4) in fan_in mode and uniform distribution (layer 5) in fan_avg mode</p>

Table 13. Neural networks architectures employed for each mass of the lightest VLL for ZA. The architecture is determined following the implementation of an evolutive algorithm that maximises the Asimov significance.

Mass	VBF
200 GeV	<p style="text-align: center;"><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p style="text-align: center;"><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p style="text-align: center;"><u>Activation function</u> : tanh (for layers 1 to 5) and sigmoid (layer 6)</p> <p style="text-align: center;"><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 5) in fan_in mode and uniform distribution (layer 6) in fan_avg mode</p>
486 GeV	<p style="text-align: center;"><u>Layers</u> : 1 input + 2 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p style="text-align: center;"><u>Regularizer</u> : L2 (for layers 1 to 3) and none (for layer 4)</p> <p style="text-align: center;"><u>Activation function</u> : sigmoid (for layers 1 to 4)</p> <p style="text-align: center;"><u>Initializer</u> : VarianceScaling, with uniform distribution (for layers 1 to 3) in fan_in mode and uniform distribution (layer 4) in fan_avg mode</p>
677 GeV	<p style="text-align: center;"><u>Layers</u> : 1 input + 4 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p style="text-align: center;"><u>Regularizer</u> : L2 (for layers 1 to 5) and none (for layer 6)</p> <p style="text-align: center;"><u>Activation function</u> : sigmoid (for layers 1 to 5)</p> <p style="text-align: center;"><u>Initializer</u> : VarianceScaling, with uniform distribution (for layers 1 and 5) in fan_in mode and uniform distribution (layer 6) in fan_avg mode</p>

Table 14. Neural networks architectures employed for each mass of the lightest VLL for VBF. The architecture is determined following the implementation of an evolutive algorithm that maximises the Asimov significance.

Mass	VLBSM
200 GeV	<p><u>Layers</u> : 1 input + 2 hidden + 1 output. Hidden and input layers with 512 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 3) and none (for layer 4)</p> <p><u>Activation function</u> : relu (for layers 1 until 3) and sigmoid (layer 4)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 3) in fan_in mode and uniform distribution (layer 4) in fan_avg mode</p>
486 GeV	<p><u>Layers</u> : 1 input + 2 hidden + 1 output. Hidden and input layers with 256 neurons each, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (for layers 1 to 3) and none (layer 4)</p> <p><u>Activation function</u> : sigmoid (for layers 1 to 4)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (for layers 1 to 3) in fan_in mode and uniform distribution (layer 4) in fan_avg mode</p>
677 GeV	<p><u>Layers</u> : 1 input + 1 output. Input layer with 256 neurons, output layer with 3 neurons</p> <p><u>Regularizer</u> : L2 (layer 1) and none (layer 2)</p> <p><u>Activation function</u> : relu (layer 1) and sigmoid (layer 2)</p> <p><u>Initializer</u> : VarianceScaling, with normal distribution (layer 1) in fan_in mode and uniform distribution (layer 2) in fan_avg mode</p>

Table 15. Neural networks architectures employed for each mass of the lightest VLL for VLBSM. The architecture is determined following the implementation of an evolutive algorithm that maximises the Asimov significance.

D Hierarchical clustering for the neural networks

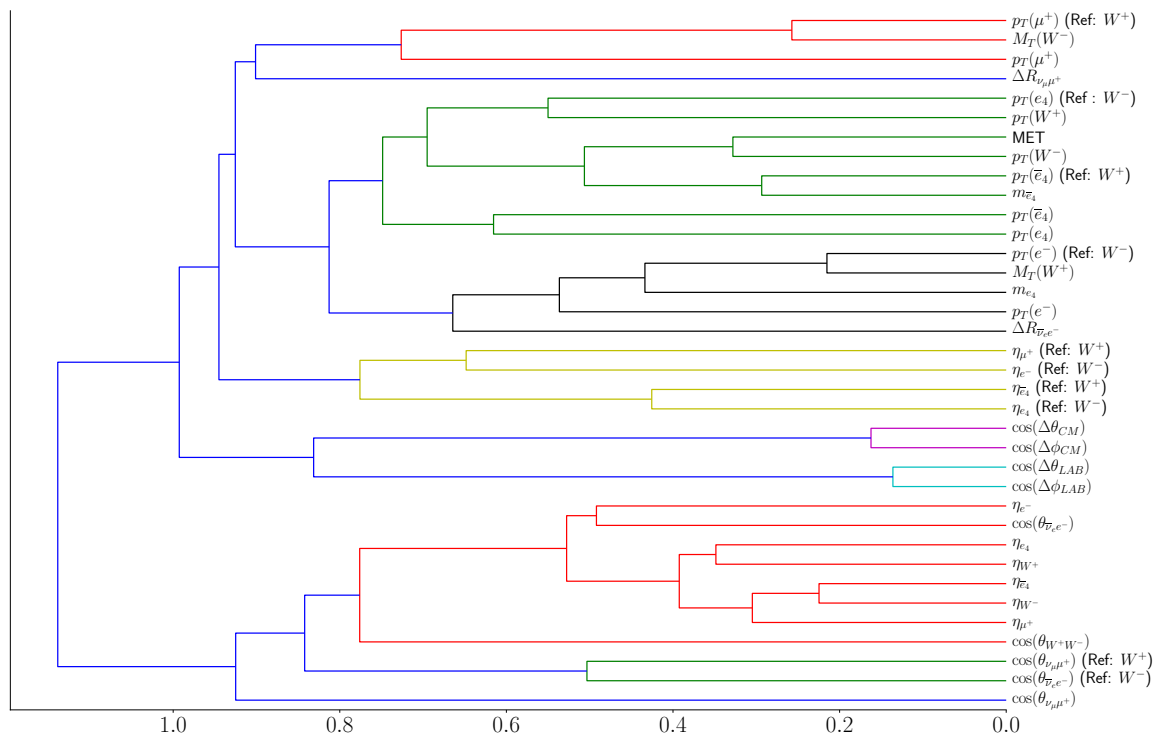


Figure 24. Hierarchical clustering of the various kinematic/angular variables utilized by the neural network model for a light VLL of 200 GeV and a ZA topology signal, under the Asimov metric. This dendrograms show how the NN groups the input features in different clusters. The dendrogram illustrates how each cluster is composed by drawing a U-shaped link between a non-singleton cluster and its children. The length of the two legs of the U-link represents the distance between the child clusters.

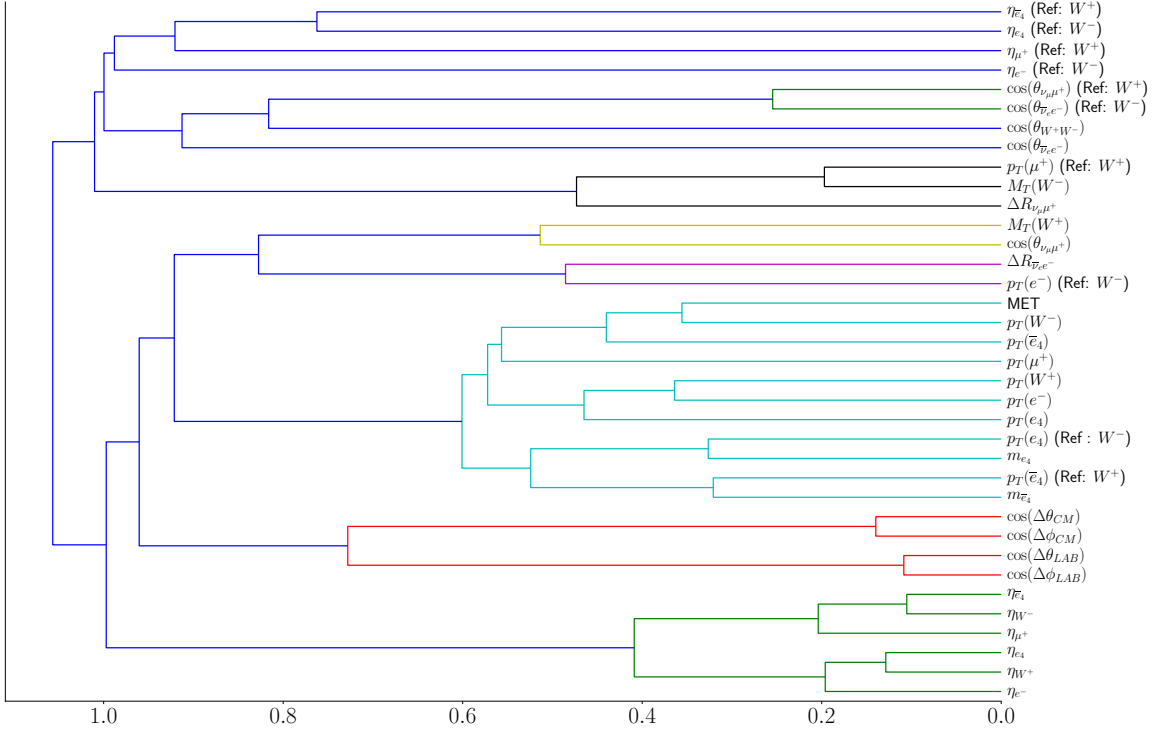


Figure 25. Hierarchical clustering of the various kinematic/angular variables utilized by the neural network model for a light VLL of 200 GeV and VBF topology signal, under the Asimov metric.

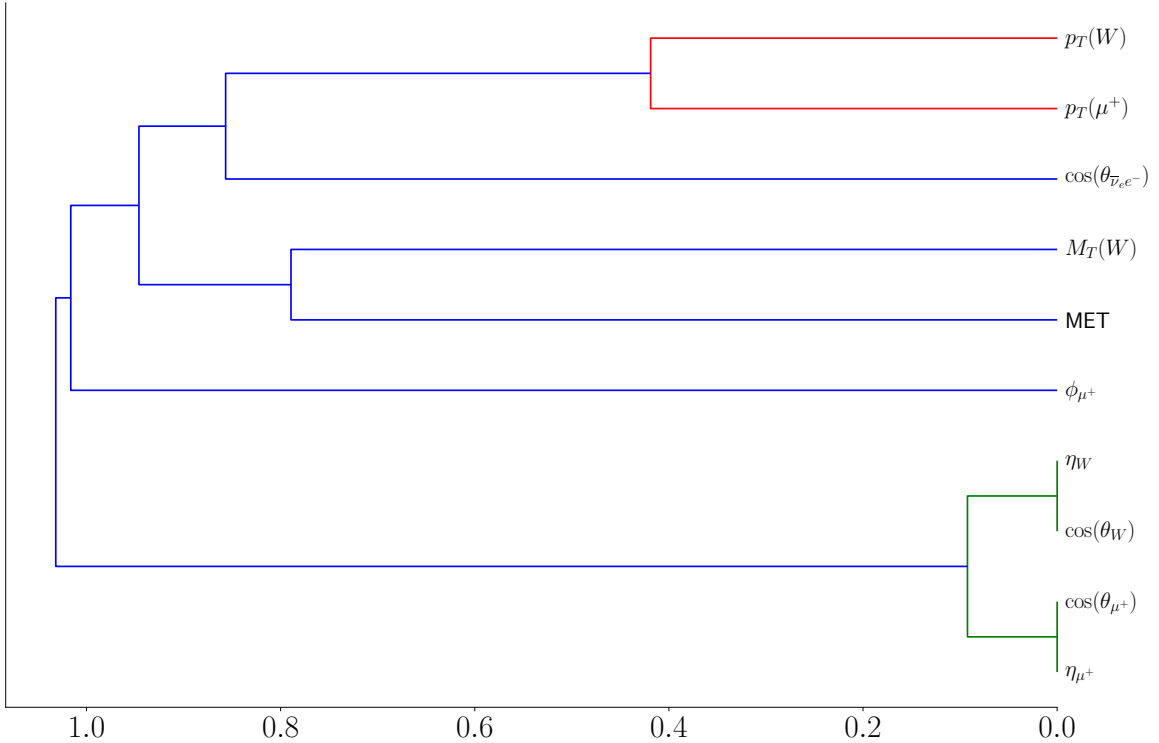


Figure 26. Hierarchical clustering of the various kinematic/angular variables utilized by the neural network model for a light VLL of 200 GeV and VLBSM topology signal, under the Asimov metric.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [[arXiv:1207.7235](https://arxiv.org/abs/1207.7235)] [[INSPIRE](#)].
- [2] UA1 collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122** (1983) 103 [[INSPIRE](#)].
- [3] GARGAMELLE NEUTRINO collaboration, *Observation of neutrino like interactions without muon or electron in the Gargamelle neutrino experiment*, *Phys. Lett. B* **46** (1973) 138 [[INSPIRE](#)].
- [4] F.J. Hasert et al., *Search for elastic ν_μ electron scattering*, *Phys. Lett. B* **46** (1973) 121 [[INSPIRE](#)].
- [5] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626 [[hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002)] [[INSPIRE](#)].
- [6] R.H. Parker, C. Yu, W. Zhong, B. Estey and H. Müller, *Measurement of the fine-structure constant as a test of the Standard Model*, *Science* **360** (2018) 191 [[arXiv:1812.04130](https://arxiv.org/abs/1812.04130)] [[INSPIRE](#)].
- [7] D. Hanneke, S. Hoogerheide and G. Gabrielse, *Cavity control of a single-electron quantum cyclotron: measuring the electron magnetic moment*, *Phys. Rev. A* **83** (2011) 052122 [[arXiv:1009.4831](https://arxiv.org/abs/1009.4831)] [[INSPIRE](#)].
- [8] SUPER-KAMIOKANDE collaboration, *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [[hep-ex/9807003](https://arxiv.org/abs/hep-ex/9807003)] [[INSPIRE](#)].
- [9] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279 [[hep-ph/0404175](https://arxiv.org/abs/hep-ph/0404175)] [[INSPIRE](#)].
- [10] D. Tong, *String theory*, [arXiv:0908.0333](https://arxiv.org/abs/0908.0333) [[INSPIRE](#)].
- [11] S.P. Martin, *A supersymmetry primer*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1 [*Adv. Ser. Direct. High Energy Phys.* **21** (2010) 1] [[hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356)] [[INSPIRE](#)].
- [12] F. Ferreira, S. Fichet and V. Sanz, *On new physics searches with multidimensional differential shapes*, *Phys. Lett. B* **778** (2018) 35 [[arXiv:1702.05106](https://arxiv.org/abs/1702.05106)] [[INSPIRE](#)].
- [13] A. Alves and F.F. Freitas, *Towards recognizing the light facet of the Higgs boson*, *Mach. Learn. Sci. Tech.* **1** (2020) 045025 [[arXiv:1912.12532](https://arxiv.org/abs/1912.12532)] [[INSPIRE](#)].
- [14] J.E. Camargo-Molina, A.P. Morais, A. Ordell, R. Pasechnik, M.O.P. Sampaio and J. Wessén, *Reviving trinification models through an E_6 -extended supersymmetric GUT*, *Phys. Rev. D* **95** (2017) 075031 [[arXiv:1610.03642](https://arxiv.org/abs/1610.03642)] [[INSPIRE](#)].
- [15] J.E. Camargo-Molina, A.P. Morais, A. Ordell, R. Pasechnik and J. Wessén, *Scale hierarchies, symmetry breaking and particle spectra in SU(3)-family extended SUSY trinification*, *Phys. Rev. D* **99** (2019) 035041 [[arXiv:1711.05199](https://arxiv.org/abs/1711.05199)] [[INSPIRE](#)].
- [16] A.P. Morais, R. Pasechnik and W. Porod, *Grand unified origin of gauge interactions and families replication in the Standard Model*, [arXiv:2001.04804](https://arxiv.org/abs/2001.04804) [[INSPIRE](#)].

- [17] A.P. Morais, R. Pasechnik and W. Porod, *Prospects for new physics from gauge left-right-colour-family grand unification hypothesis*, *Eur. Phys. J. C* **80** (2020) 1162 [[arXiv:2001.06383](#)] [[INSPIRE](#)].
- [18] ATLAS collaboration, *Search for squarks and gluinos in final states with jets and missing transverse momentum using 139 fb^{-1} of $\sqrt{s} = 13\text{ TeV}$ pp collision data with the ATLAS detector*, *PoS(EPS-HEP2019)605* (2020) [[arXiv:2010.14293](#)] [[INSPIRE](#)].
- [19] ATLAS collaboration, *Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum*, *JHEP* **12** (2019) 060 [[arXiv:1908.03122](#)] [[INSPIRE](#)].
- [20] ATLAS collaboration, *Search for pair production of higgsinos in final states with at least three b-tagged jets in $\sqrt{s} = 13\text{ TeV}$ pp collisions using the ATLAS detector*, *Phys. Rev. D* **98** (2018) 092002 [[arXiv:1806.04030](#)] [[INSPIRE](#)].
- [21] CMS collaboration, *Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$ in final states with jets and missing transverse momentum*, *JHEP* **05** (2018) 025 [[arXiv:1802.02110](#)] [[INSPIRE](#)].
- [22] CMS collaboration, *Search for new phenomena in final states with two opposite-charge, same-flavor leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13\text{ TeV}$* , *JHEP* **03** (2018) 076 [[arXiv:1709.08908](#)] [[INSPIRE](#)].
- [23] CMS collaboration, *Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$* , *JHEP* **03** (2018) 166 [[arXiv:1709.05406](#)] [[INSPIRE](#)].
- [24] I. Dorsner, S. Fajfer and I. Mustac, *Light vector-like fermions in a minimal SU(5) setup*, *Phys. Rev. D* **89** (2014) 115004 [[arXiv:1401.6870](#)] [[INSPIRE](#)].
- [25] S. Raby and A. Trautner, *Vector-like chiral fourth family to explain muon anomalies*, *Phys. Rev. D* **97** (2018) 095006 [[arXiv:1712.09360](#)] [[INSPIRE](#)].
- [26] M. Crispim Romao, S.F. King and G.K. Leontaris, *Non-universal Z' from fluxed GUTs*, *Phys. Lett. B* **782** (2018) 353 [[arXiv:1710.02349](#)] [[INSPIRE](#)].
- [27] A. Karozas, G.K. Leontaris, I. Tavellaris and N.D. Vlachos, *On the LHC signatures of $SU(5) \times U(1)'$ F-theory motivated models*, [arXiv:2007.05936](#) [[INSPIRE](#)].
- [28] Z. Poh and S. Raby, *Vectorlike leptons: muon $g - 2$ anomaly, lepton flavor violation, Higgs boson decays, and lepton nonuniversality*, *Phys. Rev. D* **96** (2017) 015032 [[arXiv:1705.07007](#)] [[INSPIRE](#)].
- [29] E. Megias, M. Quirós and L. Salas, *$g_\mu - 2$ from vector-like leptons in warped space*, *JHEP* **05** (2017) 016 [[arXiv:1701.05072](#)] [[INSPIRE](#)].
- [30] J.E. Camargo-Molina, A.P. Morais, R. Pasechnik and J. Wessén, *On a radiative origin of the Standard Model from trinification*, *JHEP* **09** (2016) 129 [[arXiv:1606.03492](#)] [[INSPIRE](#)].
- [31] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens and O. Ruchayskiy, *Sterile neutrino Dark Matter*, *Prog. Part. Nucl. Phys.* **104** (2019) 1 [[arXiv:1807.07938](#)] [[INSPIRE](#)].
- [32] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].

- [33] F. Staub, *SARAH 4: a tool for (not only SUSY) model builders*, *Comput. Phys. Commun.* **185** (2014) 1773 [[arXiv:1309.7223](#)] [[INSPIRE](#)].
- [34] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *UFO — the Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201 [[arXiv:1108.2040](#)] [[INSPIRE](#)].
- [35] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [36] DELPHES 3 collaboration, *DELPHES 3, a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](#)] [[INSPIRE](#)].
- [37] S. Höche et al., *Matching parton showers and matrix elements*, in *HERA and the LHC: a workshop on the implications of HERA for LHC physics*, CERN-2005-014, (2005), pg. 288 [[hep-ph/0602031](#)] [[INSPIRE](#)].
- [38] N.F. Bell, M.J. Dolan, L.S. Friedrich, M.J. Ramsey-Musolf and R.R. Volkas, *Electroweak baryogenesis with vector-like leptons and scalar singlets*, *JHEP* **09** (2019) 012 [[arXiv:1903.11255](#)] [[INSPIRE](#)].
- [39] A. Falkowski, D.M. Straub and A. Vicente, *Vector-like leptons: Higgs decays and collider phenomenology*, *JHEP* **05** (2014) 092 [[arXiv:1312.5329](#)] [[INSPIRE](#)].
- [40] I. Garcia Garcia, K. Howe and J. March-Russell, *Natural Scherk-Schwarz theories of the weak scale*, *JHEP* **12** (2015) 005 [[arXiv:1510.07045](#)] [[INSPIRE](#)].
- [41] N. Kumar and S.P. Martin, *Vectorlike leptons at the Large Hadron Collider*, *Phys. Rev. D* **92** (2015) 115018 [[arXiv:1510.03456](#)] [[INSPIRE](#)].
- [42] R. Dermisek and A. Raval, *Explanation of the muon $g - 2$ anomaly with vectorlike leptons and its implications for Higgs decays*, *Phys. Rev. D* **88** (2013) 013017 [[arXiv:1305.3522](#)] [[INSPIRE](#)].
- [43] R. Dermisek, J.P. Hall, E. Lunghi and S. Shin, *Limits on vector-like leptons from searches for anomalous production of multi-lepton events*, *JHEP* **12** (2014) 013 [[arXiv:1408.3123](#)] [[INSPIRE](#)].
- [44] B. Holdom and M. Ratzlaff, *Neglected heavy leptons at the LHC*, *Phys. Rev. D* **90** (2014) 013015 [[arXiv:1405.4573](#)] [[INSPIRE](#)].
- [45] S.A.R. Ellis, R.M. Godbole, S. Gopalakrishna and J.D. Wells, *Survey of vector-like fermion extensions of the Standard Model and their phenomenological implications*, *JHEP* **09** (2014) 130 [[arXiv:1404.4398](#)] [[INSPIRE](#)].
- [46] B. Bhattacharjee, P. Byakti, A. Kushwaha and S.K. Vempati, *Unification with vector-like fermions and signals at LHC*, *JHEP* **05** (2018) 090 [[arXiv:1702.06417](#)] [[INSPIRE](#)].
- [47] J. Kawamura, S. Raby and A. Trautner, *Complete vector-like fourth family with $U(1)'$: a global analysis*, *Phys. Rev. D* **101** (2020) 035026 [[arXiv:1911.11075](#)] [[INSPIRE](#)].
- [48] P.N. Bhattiprolu and S.P. Martin, *Prospects for vector-like leptons at future proton-proton colliders*, *Phys. Rev. D* **100** (2019) 015033 [[arXiv:1905.00498](#)] [[INSPIRE](#)].
- [49] K. Kowalska and D. Kumar, *Road map through the desert: unification with vector-like fermions*, *JHEP* **12** (2019) 094 [[arXiv:1910.00847](#)] [[INSPIRE](#)].
- [50] K. Fujikawa, *A vector-like extension of the Standard Model*, *Prog. Theor. Phys.* **92** (1994) 1149 [[hep-ph/9411258](#)] [[INSPIRE](#)].

- [51] A. Crivellin, F. Kirk, C.A. Manzari and M. Montull, *Global electroweak fit and vector-like leptons in light of the Cabibbo angle anomaly*, [arXiv:2008.01113](#) [INSPIRE].
- [52] CMS collaboration, *Search for vector-like leptons in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **100** (2019) 052003 [[arXiv:1905.10853](#)] [INSPIRE].
- [53] S. Bhattacharya, P. Ghosh, N. Sahoo and N. Sahu, *Mini review on vector-like leptonic dark matter, neutrino mass, and collider signatures*, *Front. in Phys.* **7** (2019) 80 [[arXiv:1812.06505](#)] [INSPIRE].
- [54] D. Cogollo, F.F. Freitas, C.A.S. Pires, Y.M. Oviedo-Torres and P. Vasconcelos, *Deep learning analysis of the inverse seesaw in a 3-3-1 model at the LHC*, *Phys. Lett. B* **811** (2020) 135931 [[arXiv:2008.03409](#)] [INSPIRE].
- [55] F.F. Freitas, C.K. Khosa and V. Sanz, *Exploring the Standard Model EFT in VH production with machine learning*, *Phys. Rev. D* **100** (2019) 035040 [[arXiv:1902.05803](#)] [INSPIRE].
- [56] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D* **98** (2018) 030001 [INSPIRE].
- [57] *Computing resources webpage*, <http://gravitation.web.ua.pt/computing>.
- [58] F. Chollet, *Keras*, <https://github.com/fchollet/keras>, (2015).
- [59] M. Abadi et al., *TensorFlow: large-scale machine learning on heterogeneous distributed systems*, [arXiv:1603.04467](#) [INSPIRE].
- [60] A. Elwood and D. Krücker, *Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders*, [arXiv:1806.00322](#) [INSPIRE].
- [61] C. Darwin, *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*, John Murray, London, U.K. (1859).
- [62] N.V. Chawla, K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, *J. Artif. Intel. Res.* **16** (2002) 321 [[arXiv:1106.1813](#)].
- [63] M. D’Onofrio, *Prospects for LHC run 3 and HL-LHC*, [Fermilab seminar](#), U.S.A., October 2019.