



Deep learning searches for vector-like leptons at the LHC and electron/muon colliders

António P. Morais^{1,2,a} , António Onofre^{3,b}, Felipe F. Freitas^{1,c}, João Gonçalves^{1,d}, Roman Pasechnik^{4,e}, Rui Santos^{5,6,f}

¹ Departamento de Física da Universidade de Aveiro and Centre for Research and Development in Mathematics and Applications (CIDMA), 3810-183 Aveiro, Portugal

² Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland

³ Departamento de Física da Universidade do Minho, 4710-057 Braga, Portugal

⁴ Department of Physics, Lund University, Sölvegatan 14A, 223-62 Lund, Sweden

⁵ Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal

⁶ ISEL-Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, 1959-007 Lisbon, Portugal

Received: 5 August 2022 / Accepted: 12 February 2023 / Published online: 21 March 2023
© The Author(s) 2023

Abstract The discovery potential of both singlet and doublet vector-like leptons (VLLs) at the Large Hadron Collider (LHC) as well as at the not-so-far future muon and electron machines is explored. The focus is on a single production channel for LHC direct searches while double production signatures are proposed for the leptonic colliders. A Deep Learning algorithm to determine the discovery (or exclusion) statistical significance at the LHC is employed. While doublet VLLs can be probed up to masses of 1 TeV, their singlet counterparts have very low cross sections and can hardly be tested beyond a few hundreds of GeV at the LHC. This motivates a physics-case analysis in the context of leptonic colliders where one obtains larger cross sections in VLL double production channels, allowing to probe higher mass regimes otherwise inaccessible even to the LHC high-luminosity upgrade.

1 Introduction

The Standard Model (SM) of particle physics has, so far, served as the guide towards unraveling the nature of all subatomic phenomena and its success is now indisputable. However, it is clear that it lacks some of the necessary ingredients

to fully describe nature, such as an explanation for neutrino masses, as firstly observed in neutrino oscillations by experiments like the Super-Kamiokande [1] or a particle candidate to explain experimental evidences for the existence of Dark Matter (DM) [2]. Recently, other discrepancies have also come into the forefront, such as the measurement of the anomalous magnetic moment of the muon at Fermilab [3], indicating a tension of 4.2σ with the SM predictions, as well as the recent results coming from the LHCb experiment, where hints of lepton flavour universality violation have been reported at a 3.1σ significance [4]. Naturally, such deviations do not meet the 5σ requirement to indicate the existence of new physics (NP) and more data still needs to be collected if the experimental deviations are confirmed. Together with the theoretical calculations in the framework of the SM, an explanation of anomalies requires the addition of NP.

Finding well motivated NP scenarios that simultaneously address all the aforementioned questions is of utmost importance and, in particular, Grand Unified theories (GUTs) provide interesting avenues to follow. These classes of models typically predict new particles present at the TeV scale (or not too far from it), that may be probed at current and future colliders. Therefore, the study of simplified frameworks that can later be mapped to the low-energy limit of a certain UV complete framework can help in constraining their viable parameter space. In particular, some of the authors have proposed a GUT framework, which unifies all matter, Higgs and gauge sectors within the E_8 group, either at a conventional high-scale scenario [5,6], or a low-scale one [7,8]. It turns out that, as a common prediction of any of such proposals, the presence of new doublet Vector-Like Lepton (VLL) particles

^a e-mail: aapmorais@ua.pt (corresponding author)

^b e-mail: onofre@fisica.uminho.pt

^c e-mail: felipefreitas@ua.pt

^d e-mail: jpedropino@ua.pt

^e e-mail: Roman.Pasechnik@hep.lu.se

^f e-mail: rasantos@fc.ul.pt

at the TeV scale or below is expected. A phenomenological analysis of doublet type VLLs, in the context of GUT model, was extensively discussed by some of the authors in a previous work [9]. Naturally, such description was bounded by the constraints and details of the underlying UV theory and as such, the main VLL features can only be fully captured in more simple phenomenological models, without the vast particle content and constraints of GUT frameworks. Additionally, singlet-type VLLs, which are not predicted in the context of the previously considered GUT, can also be constructed and represent a significant part of this work. Studying the differences between both types of VLL is of relevance not only in the context of the LHC but also at future proposed leptonic colliders. Furthermore, if the anomaly in the measurement of the magnetic moment of the muon is confirmed, which can happen in about one year from the date this article is being written, a possible explanation can arise in the form of TeV scale VLLs. In this context, we refer to [10–15] for further recent VLLs phenomenological studies.

With the above arguments in mind we study the phenomenological viability of simple extensions of the SM enlarged with a single VLL and a right-handed neutrino. For completeness, we consider two scenarios, the $SU(2)_L$ doublet and the singlet VLLs. Due to their vector-like nature, they escape constraints from the fourth generation searches [16], while non-vanishing couplings to the SM can induce contributions not only to the muon $(g - 2)_\mu$ anomaly, but also opening tightly constrained channels such as $\mu \rightarrow e\gamma$ [17].

This article is organized as follows. In Sect. 2, we introduce the simplified models, presenting the relevant theoretical details that will serve as the basis behind the subsequent numerical simulations. In Sect. 3 we introduce the numerical methodology for the studies we conduct in this work. In particular, in Sect. 3.1, we discuss the impact of the main flavour constraints associated with the addition of a VLL, while in Sect. 3.2 we delve into collider analysis, both in the context of the LHC and future electron (linear) or muon colliders. Finally, in Sect. 4 we present our main results and conclude in Sect. 5.

2 The models

In this article, we study collider phenomenology of both $SU(2)_L$ doublet and singlet VLL extensions of the SM accompanied with new right-handed neutrinos. We present the NP and SM fields' quantum numbers in Tables 1 and 2 respectively.

The $SU(2)_L$ doublets are defined as

$$Q_L^i = \begin{bmatrix} u_L \\ d_L \end{bmatrix}^i, \quad L^i = \begin{bmatrix} \nu_L \\ e_L \end{bmatrix}^i,$$

Table 1 Quantum numbers of the new exotic fermions. We adopt the notation of $E_{L,R}$ being an $SU(2)$ doublet, while $\mathcal{E}_{L,R}$ is a singlet, and ν_R is the new right-handed neutrino

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
$E_{L,R}$	1	2	$-1/2$
$\mathcal{E}_{L,R}$	1	1	1
ν_R	1	1	0

Table 2 Quantum numbers for the SM fields. Here, Q_L, d_R, u_R are the quark fields, L and e_R are the lepton fields, and ϕ is the Higgs doublet

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	# of generations
Q_L	3	2	$1/6$	3
L	1	2	$-1/2$	3
d_R	3	1	$-1/3$	3
u_R	3	1	$2/3$	3
e_R	1	1	-1	3
ϕ	1	2	$1/2$	1

$$E_{L,R} = \begin{bmatrix} \nu'_{L,R} \\ e'_{L,R} \end{bmatrix}, \quad \phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}, \tag{1}$$

where i is a generation index, $i = 1, 2, 3$. The most general and renormalizable Lagrangian density of the doublet VLL model, consistent with the symmetries in Table 1, reads as

$$\begin{aligned} \mathcal{L}_d = & \left(\Theta_i \bar{E}_L e_R^i \phi + \Upsilon \bar{E}_L \nu_R \tilde{\phi} + \Sigma_i \bar{L}^i \nu_R \tilde{\phi} \right. \\ & \left. + \Omega \bar{E}_R \nu_R \tilde{\phi} + \Pi_{ij} \bar{L}^i e_R^j \phi + \text{H.c.} \right) + \\ & + M_E \bar{E}_L E_R + M_{LE}^i \bar{L}^i E_R + \frac{1}{2} M_{\nu_R} \bar{\nu}_R \nu_R, \end{aligned} \tag{2}$$

where the Yukawa couplings are denoted by $\Theta, \Upsilon, \Sigma, \Omega$ and Π , whereas bi-linear mass terms are indicated by M_E, M_{LE} and M_{ν_R} . For the singlet VLL case, the Lagrangian is written as

$$\begin{aligned} \mathcal{L}_s = & \left(\theta_i \bar{L}^i \mathcal{E}_R \phi + \sigma^i \bar{L}^i \nu_R \tilde{\phi} + \pi_{ij} \bar{L}^i e_R^j \phi + \text{H.c.} \right) \\ & + M_E \bar{\mathcal{E}}_L \mathcal{E}_R + \frac{1}{2} M_{\nu_R} \bar{\nu}_R \nu_R, \end{aligned} \tag{3}$$

where π, σ and θ are the corresponding Yukawa couplings. The mass matrices for both the singlet and the doublet scenarios, expressed in the $\{e_L^1, e_L^2, e_L^3, e'_L\} \otimes \{e_R^1, e_R^2, e_R^3, e'_R\}$ basis, take the following form

$$\begin{aligned} M_L^{\text{doublet}} &= \begin{bmatrix} \left(\frac{v\Pi}{\sqrt{2}} \right)_{3 \times 3} & (M_{LE})_{3 \times 1} \\ \left(\frac{v\Theta}{\sqrt{2}} \right)_{1 \times 3} & (M_E)_{1 \times 1} \end{bmatrix}, \\ M_L^{\text{singlet}} &= \begin{bmatrix} \left(\frac{v\pi}{\sqrt{2}} \right)_{3 \times 3} & \left(\frac{v\theta}{\sqrt{2}} \right)_{3 \times 1} \\ (0)_{1 \times 3} & (M_E)_{1 \times 1} \end{bmatrix}, \end{aligned} \tag{4}$$

with $v = 246$ GeV the electroweak symmetry breaking (EWSB) Higgs doublet VEV $(0, v)^T$, and whose eigenstates describe both the chiral SM-like charged leptons as well as the new VLLs.

For the neutrino sector the mass matrices can be expressed as

$$\begin{aligned}
 M_{\nu}^{\text{doublet}} &= \begin{bmatrix} (0)_{3 \times 3} & \left(\frac{\sigma v}{\sqrt{2}}\right)_{3 \times 1} & (0)_{3 \times 1} & (M_{LE})_{3 \times 1} \\ \left(\frac{\sigma v}{\sqrt{2}}\right)_{1 \times 3} & M_{\nu R} & \left(\frac{\Upsilon v}{\sqrt{2}}\right)_{1 \times 1} & \left(\frac{\Omega v}{\sqrt{2}}\right)_{1 \times 1} \\ (0)_{1 \times 3} & \left(\frac{\Upsilon v}{\sqrt{2}}\right)_{1 \times 1} & (0)_{1 \times 1} & (M_E)_{1 \times 1} \\ (M_{LE})_{1 \times 3} & \left(\frac{\Omega v}{\sqrt{2}}\right)_{1 \times 1} & (M_E)_{1 \times 1} & (0)_{1 \times 1} \end{bmatrix}, \\
 M_{\nu}^{\text{singlet}} &= \begin{bmatrix} (0)_{3 \times 3} & \left(\frac{\sigma v}{\sqrt{2}}\right)_{3 \times 1} \\ \left(\frac{\sigma v}{\sqrt{2}}\right)_{1 \times 3}^\dagger & (M_{\nu R})_{1 \times 1} \end{bmatrix}. \tag{5}
 \end{aligned}$$

where the $\{v_L^1, v_L^2, v_L^3, \nu_R, v_L', v_R'\} \otimes \{v_L^1, v_L^2, v_L^3, \nu_R, v_L', v_R'\}$ basis was used for the doublet scenario whereas that of the singlet model was chosen as $\{v_L^1, v_L^2, v_L^3, \nu_R\} \otimes \{v_L^1, v_L^2, v_L^3, \nu_R\}$.

In this article, one considers that the lightest beyond-the-SM (BSM) neutrino is in the KeV range and sterile, acting as missing energy in the detector. Therefore, while the neutrino mass matrices above have the standard type-I seesaw form, such a light sterile neutrino implies that $M_{\nu R}$ is equally small, which means that Σ, Υ and Ω (for the doublet scenario) as well as σ (for the singlet scenario), need to be very small in order for both models to be consistent with the active neutrinos mass scale. While a detailed discussion for the neutrino mass generation mechanism is beyond the scope of this article, it is not a difficult task to minimally extend both the doublet and singlet scenarios, e.g. by introducing the dimension-five Weinberg operator, without affecting our analysis and conclusions [18]. However, the fact that Σ, Υ, Ω and σ need to be very small is calling for the introduction of an approximate global lepton number symmetry, which, for convenience, we define as $U(1)_e \times U(1)_\mu \times U(1)_\tau$, such that only leptons transform non-trivially according to the following quantum numbers:

Such an approximate symmetry constrains the model parameters as follows:

- Σ, Υ, Ω (for the doublet case) and σ (for the singlet case) must be tiny in consistency with the smallness of the active neutrino masses;
- The charged lepton-sector Yukawa couplings Π (for the doublet model) and π (for the singlet model), are approximately diagonal and off-diagonal elements can be ignored in our analysis;

Table 3 Quantum numbers under the approximate lepton number symmetry

Field	$U(1)_e$	$U(1)_\mu$	$U(1)_\tau$
$E_{L,R}$	0	1	0
$\mathcal{E}_{L,R}$	0	1	0
ν_R	0	0	0
L^1	1	0	0
L^2	0	1	0
L^3	0	0	1
e_R^1	1	0	0
e_R^2	0	1	0
e_R^3	0	0	1

- The VLLs solely couple to muons such that only $\Theta_2, M_{LE,2}$ (for the doublet scenario) and θ_2 (singlet scenario) are sizeable and relevant for our numerical studies.

In what follows, we are essentially interested in lepton couplings to both gauge and Higgs bosons. In particular, the focus is on the $W\nu E(\mathcal{E}), Z^0 E(\mathcal{E})\ell$ and $HE(\mathcal{E})\ell$ vertices. All numerical computations are performed in the mass basis such that one needs to rotate the gauge eigenstates via the bi-unitary transformations

$$\begin{aligned}
 \text{Leptons : } M_L^{\text{diag}} &= (U_L^e)^\dagger M_L (U_R^e), \\
 \text{Neutrinos : } M_\nu^{\text{diag}} &= (U_\nu)^\dagger M_\nu (U_\nu). \tag{6}
 \end{aligned}$$

The list of all Feynman rules that are relevant to the numerical analysis are shown in Appendix A.

Before presenting the methodology behind the numerical calculations, it is instructive to revisit the main experimental constraints for VLLs as well as the most relevant couplings for this discussion. First, let us recall that there are not many direct searches for VLLs, at least, in comparison with their quark counterparts. In fact, LEP searches have lead to a lower bound on the VLLs mass of $M_{\text{VLL}} > 100.6$ GeV [19], while the most restrictive constraints are coming from CMS where direct searches for doublet VLLs with sizeable couplings to taus are excluded at 95% Confidence Level (CL) in the 120–790 GeV mass range [20]. Additionally, there are searches for charged lepton resonances which constrain the mass of a hypothetical VLL to be above a few hundred GeV, depending on the underlying assumptions (114–176 GeV for VLL and 100–468 GeV for a heavy lepton in the type-III inverse seesaw model [21], as well as 80–210 GeV in Ref. [22]). However, for the model under consideration, the $[U(1)]^3$ lepton number symmetry dictates that the relevant couplings are those to muons such that the most restrictive search in [20] does not apply.

The left-mixing matrix that rotates from the flavour to the mass basis can be parametrized by a single mixing angle, α ,

as follows

$$U_L^e = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & 0 & -\sin \alpha \\ 0 & 0 & 1 & 0 \\ 0 & -\sin \alpha & 0 & -\cos \alpha \end{bmatrix}. \tag{7}$$

With the $U(1)_e \times U(1)_\mu \times U(1)_\tau$ approximate symmetry the mass matrices for both the doublet and singlet cases take a simple form

$$M_L^{\text{doublet}} = \begin{bmatrix} \frac{v\Pi_{11}}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{v\Pi_{22}}{\sqrt{2}} & 0 & M_{LE,2} \\ 0 & 0 & \frac{v\Pi_{33}}{\sqrt{2}} & 0 \\ 0 & \frac{v\Theta_2}{\sqrt{2}} & 0 & M_E \end{bmatrix},$$

$$M_L^{\text{singlet}} = \begin{bmatrix} \frac{v\Pi_{11}}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{v\Pi_{22}}{\sqrt{2}} & 0 & \frac{v\theta_2}{\sqrt{2}} \\ 0 & 0 & \frac{v\Pi_{33}}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & M_E \end{bmatrix}. \tag{8}$$

Besides lepton mixing terms one also considers a neutrino mixing with the following structure,

$$U_\nu^{\text{doublet}} = \begin{bmatrix} (U_{\text{PMNS}})_{11} & (U_{\text{PMNS}})_{12} & (U_{\text{PMNS}})_{13} & 0 & 0 & 0 \\ (U_{\text{PMNS}})_{21} & (U_{\text{PMNS}})_{22} & (U_{\text{PMNS}})_{23} & 0 & 0 & 0 \\ (U_{\text{PMNS}})_{31} & (U_{\text{PMNS}})_{32} & (U_{\text{PMNS}})_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & (U_\nu)_{55} & (U_\nu)_{56} \\ 0 & 0 & 0 & 0 & (U_\nu)_{65} & (U_\nu)_{66} \end{bmatrix},$$

$$U_\nu^{\text{singlet}} = \begin{bmatrix} (U_{\text{PMNS}})_{11} & (U_{\text{PMNS}})_{12} & (U_{\text{PMNS}})_{13} & 0 \\ (U_{\text{PMNS}})_{21} & (U_{\text{PMNS}})_{22} & (U_{\text{PMNS}})_{23} & 0 \\ (U_{\text{PMNS}})_{31} & (U_{\text{PMNS}})_{32} & (U_{\text{PMNS}})_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

where the 3×3 SM block is fixed by the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [23]. Technically, and considering only such a block, the PMNS matrix must in rigour be defined as $U_{\text{PMNS}} = U_\nu \cdot U_L^\dagger$ such that the actual mixing is instead given as $(U_{\text{PMNS}})_{i1} = (U_\nu)_{i1}$, $(U_{\text{PMNS}})_{i3} = (U_\nu)_{i3}$ and $(U_{\text{PMNS}})_{i2} = (U_\nu)_{i2} \cos \alpha$ with $i = 1, 2, 3$. However, the benchmark scenarios considered in Sect. 4 are such that $\cos \alpha \approx 1$ for both the doublet and the singlet models and therefore the parametrization in Eq. (9) is consistent with a realistic lepton mixing.

The 2×2 block in the doublet case represents the mixing between the ν_5 and ν_6 mass eigenstates. These elements do not contribute to the interaction vertices used in our analysis and therefore their size and numerical values are not relevant

for the discussion. For consistency, we assume a generic $\mathcal{O}(1)$ mixing. It also follows from the approximate lepton number symmetry that the mixing between the right-handed neutrino and the remaining neutral leptons is negligible and can be ignored. This means that $(U_\nu)_{44} = 1$, $(U_\nu)_{4j} = (U_\nu)_{j4} = 0$ and the couplings between the SM and BSM neutrinos can be neglected.

Let us note that if one requires consistency with the recently observed muon $(g - 2)_\mu$ anomaly [3], the preferred values of the cosine of the mixing angle α must not be far from unity. On the other hand, a small α (or equivalently, small Yukawa parameters θ_2 and Θ_2) is preferable in order to keep tree-level modifications to the Feynman rules for the $H\mu^+\mu^-$ and $Z^0\mu^+\mu^-$ vertices under control (a list of the rules are shown in Appendix A). Indeed, a large mixing angle would make the values of the branching fractions $\text{BR}(Z^0 \rightarrow \mu^+\mu^-)$ and $\text{BR}(H \rightarrow \mu^+\mu^-)$ unacceptably large. Furthermore, and noting that the VLL only couples to muons, one can see from [24] that EW global fits constrain the mixing angle to be less than 0.034, which we impose in our numerical analysis. With such a small mixing, lepton flavour violation observables can be kept under control and the VLL single production cross-section is, for the singlet model, dominated by the $W\nu_\ell\mathcal{E}$ vertex and, for the doublet model, by the $W\nu'_L E$ vertex, as we discuss in Sect. 3.2. It follows from the mixing structure in Eq. (9) (see also the Feyn-

man rules in Eqs. (A2), (A3), (A6) and (A7)) that interactions involving right-handed neutrinos are absent in both scenarios. This is consistent with the approximate lepton number symmetry, where one can assume that the smallness of the σ and Σ Yukawa couplings, results in a negligible $W\nu_R\mathcal{E}$ and $W\nu_R E$ interaction strength. Therefore, in our numerical analysis, one can ignore such interactions by safely setting it to zero.

Another important element to take into consideration is the VLL decay width. In particular, it can be expressed as

$$\Gamma(E(\mathcal{E}) \rightarrow W\nu_\ell) = \frac{g^2 U_{\text{MIX}}^2 m_{E,\mathcal{E}}^2}{64\pi M_W^2} \left(1 - \frac{M_W^2}{m_{E,\mathcal{E}}^2} \right)^2$$

$$\times \left(1 + 2 \frac{M_W^2}{m_{E,\mathcal{E}}^2} \right). \tag{10}$$

Here, U_{MIX}^2 is a non-trivial combination of the neutrino and charged lepton mixing matrices whose specific structure can be determined from the Feynman rules in Eqs. (A2), (A3), (A6) and (A7). One notes that such a decay width grows as a power law in m_E , becoming rather significant for large masses as shown in Fig. 1. For example, taking the case where $U_{\text{MIX}} = 1$, one can observe that for a mass of 2 TeV, the decay width already surpasses the mass itself with $\Gamma = 2705.4$ GeV. Such an effect can be mitigated for smaller mixing angles as demonstrated in Fig. 1. Indeed, a decay width greater than its mass leads to conceptual issues with respect to the interpretation of such large-width fields as particles. In fact, within the context of Quantum Field Theory, a large value for the width implies that the particle is highly non-local and the narrow-width approximation is no longer valid calling for a more robust framework to describe it (see e.g. [25–27]). Such a scenario is not relevant for the charged VLL as the mixing element U_{MIX} must be small due to EW constraints involving Higgs and Z boson decays. On the other hand, the VLL neutrino partners that reside in the same doublet representation can also decay via a W boson and a charged lepton. In the limit of $m_{\nu'_L} \gg m_\ell$, the expression for the decay width is identical to that of the VLL (with the appropriate changes),

$$\Gamma(\nu'_L \rightarrow W\ell) = \frac{g^2 U_{\text{MIX}}^2 m_{\nu'_L}^2}{64\pi M_W^2} \left(1 - \frac{M_W^2}{m_{\nu'_L}^2} \right)^2 \left(1 + 2 \frac{M_W^2}{m_{\nu'_L}^2} \right), \tag{11}$$

where U_{MIX} can become sizable. Indeed, taking $\ell = \mu$, then, in accordance with the Feynman rules, the mixing elements coming from the left-chiral couplings are given as $\cos \alpha [U_v^{\text{doublet}}]_{55}$. Both $\cos(\alpha)$ and $[U_v^{\text{doublet}}]_{55}$ are of order $\mathcal{O}(1)$ and therefore the $\nu'_L \rightarrow W\ell$ decay width becomes relevant for large masses. However, notice that the neutrino mass is not independent of its charged VLL counterpart and thus cannot be made arbitrarily small for the width not

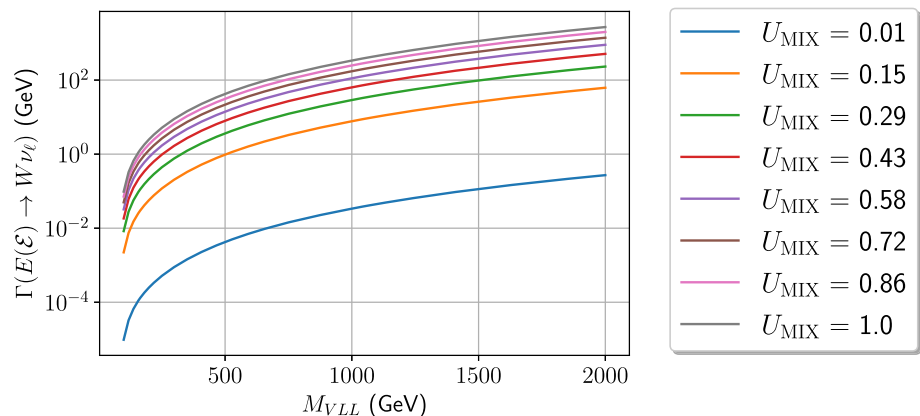
becoming too large. At leading order, both the VLL and the heavy neutrino masses are controlled by the M_E parameter in Eq. (2) making them almost degenerate with a small deviation expected to be induced at one-loop level [28]. Typically, such a mass splitting is of the order of a few hundred MeV [28], which must not play a sizeable role. With this in mind, and without loss of generality for this article’s analysis, we consider both the neutrino and its corresponding charged VLL mass degenerate, ensuring that $\Gamma(\nu'_L \rightarrow W\ell)$ is never unacceptably large. Although the widths must be kept under control, a sizeable value has interesting experimental implications. For instance, it results in wider distributions of kinematic variables such as the mass and the transverse momentum (p_T), where large tails can extend into phase space regions not populated by the SM background events.

3 VLLs: collider phenomenology and flavour constraints

3.1 Flavour constraints

It follows from interactions between the VLL and the muon that non-zero mixing elements can trigger Lepton Flavour Violation (LFV) interactions, in particular, those that relate to muon and tau decays. Therefore it is important to confront all points considered in our numerical analysis with the following LFV branching ratios (BR): $\text{BR}(\mu \rightarrow e\gamma)$, $\text{BR}(\tau \rightarrow e\gamma)$, $\text{BR}(\tau \rightarrow \mu\gamma)$, $\text{BR}(\mu^- \rightarrow e^-e^+e^-)$, $\text{BR}(\tau^- \rightarrow e^-e^+e^-)$, $\text{BR}(\tau^- \rightarrow \mu^-e^+\mu^-)$, $\text{BR}(\tau^- \rightarrow e^-e^+\mu^-)$, $\text{BR}(\tau^- \rightarrow \mu^-e^+\mu^-)$, $\text{BR}(\tau^- \rightarrow \mu^-e^+e^-)$, $\text{BR}(\tau^- \rightarrow \mu^-e^+\mu^-)$, $\text{BR}(\tau^- \rightarrow e^-e^+\mu^-)$, $\text{BR}(\tau^+ \rightarrow \pi^0e^+)$ and $\text{BR}(\tau^+ \rightarrow \pi^0\mu^+)$. For such a purpose we use the latest version of SPheno [29] to generate the Wilson coefficient cards, which are then passed to flavio [30] in order to compute the corresponding LFV observables. It is important to mention that SPheno also computes the BRs i.e. flavio merely works as an extra layer of added scrutiny, even if both programs’ numerical outputs (for the BRs) are well within each other’s error and are therefore compatible. We use the 90% CL experimental limits on the considered

Fig. 1 Decay width as a function of the VLL’s mass, both displayed in GeV. Each coloured line is representative of a different mixing element, whose numerical value is displayed in the box to the right of the plot



LFV observables as reported in the most recent issue of the PDG review [17].

Last but not least, let us comment that VLLs have long been proposed as a potential explanation for the anomalous magnetic moment of the muon. However, it follows from the smallness of the α mixing angle that neither the singlet nor doublet models discussed in this article can successfully accommodate such an anomaly as couplings of muons to VLLs and to Higgs or Z bosons are too small. This was numerically verified and consistent with the results obtained e.g. in [31]. Therefore, it is not sufficient to extend the SM with one generation of VLLs to explain the muon $g - 2$ anomaly and additional new physics with less constrained couplings is necessary. A possibility is to extend the scalar sector as proposed by various authors [32–37].

3.2 Collider analysis: LHC and muon colliders

The analysis techniques and Deep Learning algorithms used in the current study for both the doublet and the singlet models were first implemented in [9] and [37], respectively. Starting with the doublet case, the channel we are probing is characterised by a final state with 4 light jets, originating from the decays of two W bosons, a charged lepton, which we take to be the muon, and transverse missing energy (MET), whose origin results from the undetected neutrinos in the final state. At Leading Order (LO), the production diagram, at the LHC, can be seen in Fig. 2, where the heavy ν'_L neutrino belongs to a VLL doublet as shown in Eq. (1). For this signal topology the main irreducible backgrounds are the double top-quark production, $t\bar{t}$, W +jets (we consider up to 4 jets) and Diboson plus jets. Notice that for the $t\bar{t}$ channel we consider that one of the W 's decays in its hadronic channel with the other one decaying into a muon and a neutrino. All backgrounds are generated at LO precision. For the singlet scenario, the final state topology is characterized by one isolated charged lepton from the W boson decay, and transverse missing energy associated with the undetected neutrinos. In particular, we focus on the channel where the charged lepton is a muon. For this signal topology, we consider the irreducible background together with all physics processes that lead to the production of, at least, one muon and up to two jets in the final state, $pp \rightarrow \mu^- \bar{\nu}_\mu$ (+jets).¹

For the case of the VLL singlet only one BSM coupling contributes to the decay chain, in particular, the one containing a muon-specific neutrino, whereas for the doublet scenario the couplings involving both the charged and neutral components of the VLL doublet are present. Taking

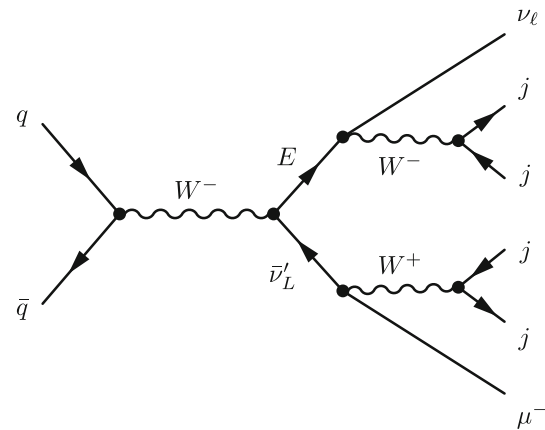


Fig. 2 Leading-order Feynman diagram for single production of the doublet VLL. Here, q and \bar{q} correspond to quarks originating from the colliding protons, E represents the VLL and ν'_L denotes the VLL doublet partner. Both W 's in the hadronic channel with j indicating first and second generation chiral quarks

into account the Feynman rules shown in Appendix A, the strength of the relevant left and right chiral charged current projections read as

$$\begin{aligned}
 g_L^{E\nu_\mu W} &= \frac{g}{\sqrt{2}} \sin \alpha \left[U_v^{\text{doublet}} \right]_{22}, \\
 g_L^{\mathcal{E}\nu_\mu W} &= \frac{g}{\sqrt{2}} \sin \alpha \left[U_v^{\text{singlet}} \right]_{22}, \\
 g_R^{\nu'_L \mu^- W^+} &= \frac{g}{\sqrt{2}} \left[U_v^{\text{doublet}} \right]_{55} \left[U_R^e \right]_{24}, \\
 g_L^{\nu'_L \mu^- W^+} &= \frac{g}{\sqrt{2}} \sin \alpha \left[U_v^{\text{doublet}} \right]_{55}, \\
 g_L^{E\nu'_L W} &= \frac{g}{\sqrt{2}} \cos \alpha \left[U_v^{\text{doublet}} \right]_{55}, \\
 g_R^{E\nu'_L W} &= -\frac{g}{\sqrt{2}} \left[U_R^e \right]_{24} \left[U_v^{\text{doublet}} \right]_{55}.
 \end{aligned} \tag{12}$$

This in turn means that the cross section for the doublet case will be much larger than the one for the singlet case, because the strength of the dominant contribution for the former, in particular the coupling between the VLL and its neutrino counterpart, is proportional to $\cos \alpha$ while for the latter all couplings are controlled by $\sin \alpha$. Notice that the single production channel in Fig. 3 is valid for both the doublet and singlet scenarios where, in the case of equal angles $g_L^{E\nu_\mu W} = g_L^{\mathcal{E}\nu_\mu W}$. Therefore, it is sufficient to study the singlet model and extract identical conclusions for the doublet case.

The analysis follows a well-defined guideline. Both models are first implemented in SARAH [38] to generate all interaction vertices as well as all relevant files that interface with Monte-Carlo generators, namely, MadGraph5 [39] for quark-level matrix-element calculations and Pythia8 [40] for hadronization and showering leading to final state parti-

¹ In more practical terms, this background is simulated in MadGraph via the commands `generate p p > mu+ vm && add process p p > mu+ vm j && add process p p > mu+ vm j j`.

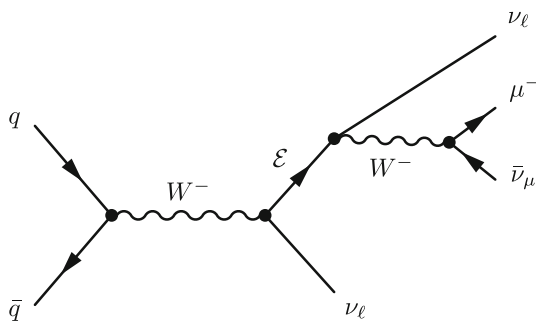


Fig. 3 Leading-order Feynman diagram for VLL single production. q and \bar{q} corresponds to quarks originating from the initial protons and ν_ℓ are the SM neutrinos

cles. In MadGraph5 we simulate proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 14$ TeV, for a total of 250k events for the signal and backgrounds. We also employ the default LO parton-distribution function NNPDF2.3 [41] which fixes the evolution of the strong coupling constant, α_s . Fast-simulation of the ATLAS detector is conducted, with Delphes [42]. All kinematic distributions are extracted with the help of ROOT [43]. At this stage, we also impose selection criteria, to maximise the signal significance. In particular, all events must satisfy

$$\begin{aligned}
 p_T(\mu^-) &> 25 \text{ GeV}, \\
 \text{MET} &> 15 \text{ GeV}, \\
 |\eta(\mu^-)| &\leq 2.5 \text{ and} \\
 |\eta(j)| &\leq 2.5.
 \end{aligned}$$

where η represents the muon pseudo-rapidity.² Additionally, for the doublet topology, which involves jets in the final state, we require them to be tagged as originating from light chiral quarks, i.e., they cannot be tagged as b-jets. Note that the signal production represented in Fig. 3, is characterized by having no jets in the final states. While it is true that the lack of no jet candidates at any given event in a hadron collider is perhaps too unrealistic, the absence of such selection can also cause complications in the Deep Learning algorithms used later in the analysis, since signal/background classes are incredibly unbalanced and may lead to over-fitting problems.

The kinematics of final states for both signal and background events are translated into tabular datasets and used as inputs for neural models, whose job is to separate the signal from the background. The neural network is constructed using Keras [44]. For classification of the singlet model, we choose a set of 5 observables computed in the laboratory frame. Low-level features include the muon observables $\cos(\theta_{\mu^-})$, $\eta(\mu^-)$, $\phi(\mu^-)$ and $p_T(\mu^-)$ as well as the MET.

² The pseudorapidity of a particle is defined as $\eta = -\ln(\tan(\theta/2))$, where θ is its polar angle.

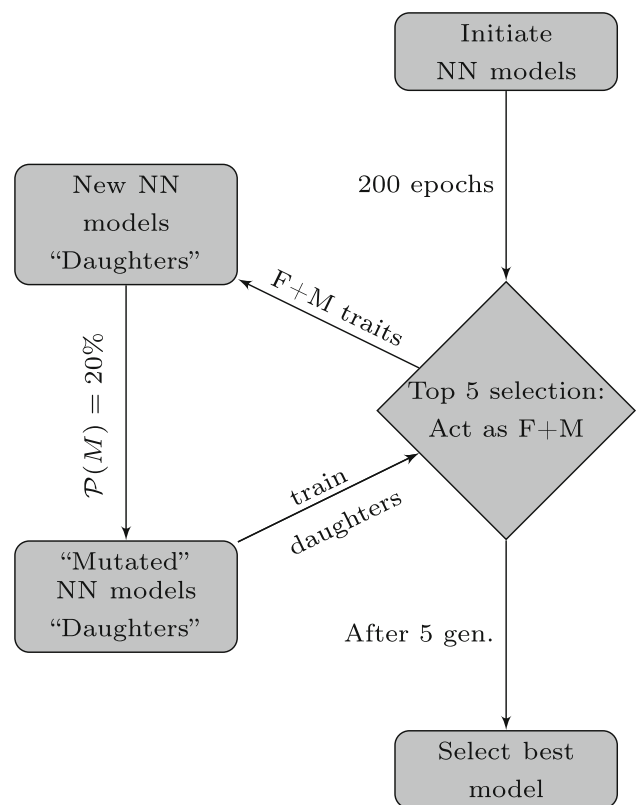


Fig. 4 Flowchart representative of all iterative steps involved in the genetic algorithm that we employ in this work

For the doublet model, a richer final state is present and, as such, one can compute a much more complete list of observables. Do note that, for this physics scenario, the final states contain 4 light-jets, which implies that one cannot distinguish between the jets originating from W^+ and those originating from W^- . Therefore, for reconstructed observables, we consider all possible combinations of jets that can be used. These are $C = (j_1, j_2), (j_1, j_3), (j_1, j_4), (j_2, j_3), (j_2, j_4)$ and (j_3, j_4) , where we define j_1 as the leading jet, with the highest p_T , and j_4 as the subleading jet, with the lowest p_T . The full list of observables that are used in the training are shown in Table 4.

To optimize the neural network architecture, we employ a genetic algorithm following the same steps as described in [9, 37] and schematically represented in the diagram of Fig. 4. The algorithm begins by first generating an arbitrary number of neural networks, whose architecture is determined by randomly pooling a list of predefined hyperparameters (number of layers, activation functions, regularisers, etc.). We then train each individual network over the data for a given number of epochs. From the trained networks, we pick the top five best performing ones. From these, we create Father–Mother pairs, where we combine 50% of the Father’s traits (that is, its hyperparameters) and 50% of the Mother traits to construct new neural architectures, which we dub as Daughters.

Table 4 Angular and kinematic distributions for the analysis of the doublet production topology. All observables are computed in the laboratory frame of reference. To simplify notation, we define (j_a, j_b) as

	Dimension-full	Dimensionless
Doublet VLL	$p_T(\mu^-), p_T(j_1), p_T(j_2), p_T(j_3)$ MET, $p_T(j_4), E(\mu^-), E(j_1), E(j_2),$ $E(j_3), E(j_4), M(j_1, j_2), M(j_1, j_3),$ $M(j_a, j_b), p_T(j_a, j_b), M(j_a, j_b, \mu^-),$ $p_T(j_a, j_b, \mu^-), M(j_a, j_b, \text{MET}), p_T(j_a, j_b, \text{MET})$	$\cos(\theta_{j_a j_b}), \cos(\theta_{j_a \mu^-}),$ $\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4),$ $\eta(\mu^-), \phi(j_1), \phi(j_2), \phi(j_3),$ $\Delta R(j_a, j_b), \Delta R(j_a, \mu^-), \Delta\phi(j_a, j_a),$ $\eta(j_4), \eta(j_a, j_b), \phi(j_a, j_b),$ $\Delta\phi(j_a, \mu^-), \Delta\theta(j_a, j_b), \Delta\theta(j_a, \mu^-).$ $\eta(j_a, j_b, \mu^-), \eta(j_a, j_b, \mu^-)$

We also consider the possibility of mutation, that is, after the Daughter networks have been built, we consider that the hyperparameters may change to another, with the probability of $\mathcal{P}(M) = 20\%$. We then train the Daughter networks for a given number of epochs and repeat the procedure for a set of generations. Finally, we select the best performing network based on some metric of choice. In our analysis, the algorithm is designed to maximise the Asimov statistical significance, based on a earlier work of Elwood and Adam in [45]. The Asimov metric is defined as [46]

$$\mathcal{Z}_A = \left[2 \left((s + b) \ln \left(\frac{(s + b)(b + \sigma_b^2)}{b^2 + (s + b)\sigma_b^2} \right) - \frac{b^2}{\sigma_b^2} \ln \left(1 + \frac{\sigma_b^2 s}{b(b + \sigma_b)} \right) \right) \right]^{1/2}, \tag{13}$$

where s is the number of signal events, b is the number of background events and σ_b is the uncertainty of the background. As part of the optimization procedure, we consider the same list of hyperparameters as in our previous work [9]:

- number of hidden layers: 1–5
- 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. The parameters are chosen randomly from the previous lists. Each network is trained up 200 epochs and if any improvement is not observed by at least 5 epochs, then the training stops. Note that a subset of the neural network properties are not subject to optimisation but instead remain fixed during the runs. Namely, we consider the following

- The input data of the NNs are standardised, that is, input vectors have a mean of zero and a standard deviation of 1. All observables were extracted from the ROOT files and outputted into dataframes. The data is reshuffled and then divided into training samples (80% of the total data)

corresponding to all jet combinations, as described in the text. Variables that involve combinations of final states, e.g. $p_T(j_a, j_b, \mu^-)$, are reconstructed from the states indicated inside the parenthesis

and validation samples (the remainder 20%). We also consider cross-validation with a five-fold scheme during training. The statistical significance is computed based on the trained NN predictions of the validation data.

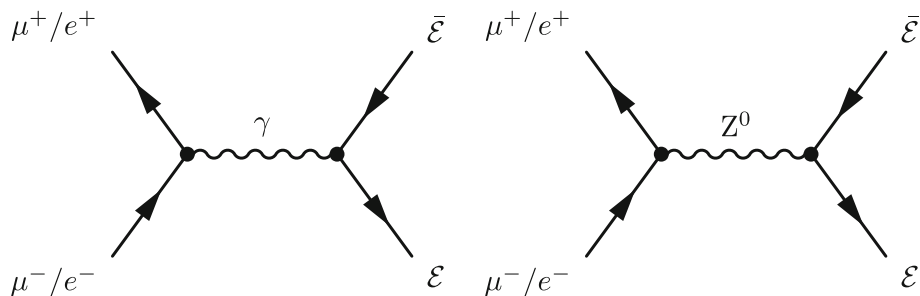
- We employ a cyclic learning rate during the training phase with 0.01 initial value and maximal value of 0.1.
- The output of the NN is a vector of probabilities such that we define a signal for an output with probability greater than 0.5, otherwise it is considered as a background.
- Batch size of 32,768.
- The best NN is selected based on the Asimov metric. The loss function for this case is defined as the inverse of Eq. (13), such that when the loss function is minimised, the Asimov significance is maximised. We also consider a fixed value of the background uncertainty to $\sigma_b = 10^{-1}$.

An additional consideration is the fact that our datasets are unbalanced, which is in part a result of the selection criteria imposed on the final states, reducing the allowed phase space for both signal and backgrounds. Unbalanced datasets can lead to overfitted networks and, as such, must be properly dealt with. In this work, we have utilised the Synthetic Minority Oversampling Technique (SMOTE) [47], which oversamples the minority classes of our training data. Do note that this algorithm is only employed to the training dataset and we do not perform any re-sampling of the validation samples.

With the LHC being a proton collider, production of coloured particles by the strong interaction is heavily favoured, compared to identical processes involving electroweak bosons. As VLLs are colour singlets, they can only be produced via this last interaction at LO. It is then worth mentioning that, in addition to the obvious motivation of studying VLL production at the LHC already during RUN3 and, later on, in its high luminosity phase (HL-LHC), it is of utmost importance to understand the sensitivity with which these new particles can be probed, at future colliders.

In particular, there has been an active discussion within the community about e^+e^- colliders like the Compact Linear Collider (CLIC) [48, 49] or the International Linear Collider (ILC) [50], and more recently, on the possibility of building a $\mu^-\mu^+$ collider [51, 52]. Besides offering cleaner environ-

Fig. 5 LO Feynman diagrams for VLL double production at a lepton/anti-lepton collider via exchange of either a photon or a Z^0 boson



ments, when compared to hadronic machines, production via electroweak processes is favoured, hence, VLLs should have a higher chance of being observed, in case they exist, in this type of machines. For this purpose, we perform numerical and analytical computations with `FeynCalc` [53] for the pair-production of VLLs whose tree-level diagrams can be seen in Fig. 5. As previously stated, singlets have lower couplings compared to doublets, hence we discuss the prospects for VLL discovery at lepton colliders for the singlet scenario and not the doublet, as this can be seen as the worst-case scenario. Do note, however, that the conclusions one takes from the singlet model can be easily generalized to the doublet model. That is, the only difference between the collider studies for the two scenarios are the values of the couplings.

In conclusion, the single VLL production is favoured at the LHC for the doublet case, due mainly to the strength of the couplings. Double production of VLLs has the obvious disadvantage of the need to produce two heavy states and a more elaborate final state to detect. Additionally, as it was shown in previous works [9,37], double production topologies are sub-leading when compared with single production ones and, as such, are not considered in this work. If one wants to probe the singlet model, and assuming that it cannot be done at the LHC, the future lepton colliders, where single production is precluded at LO, may give us enough VLLs in the double production channel to test the singlet VLL scenario.

4 Results

In this section, we present a numerical analysis focusing on the LHC and future lepton colliders. In Eqs. (14) and (15) we show two distinct viable scenarios (one for the doublet and one for the singlet model), where all couplings are fixed and only the VLL mass varies. The colour coding, black for the doublet model and green for the singlet case, merely serves to distinguish each of the benchmark scenarios and identify them throughout the text whenever necessary. We first consider a benchmark point where we set the VLL mass to $M_{VLL} = 700$ GeV. The remaining BSM neutrinos present in the doublet model share, at LO, the same mass with their $SU(2)_L$ doublet VLL counterpart and, as noted in

Eq. (12), these neutrinos efficiently decay into muons and W^\pm bosons, with an interaction strength of the order of the weak gauge coupling, where, for the considered benchmarks, $[U_\nu^{\text{doublet}}]_{55} \approx [U_R^e]_{24} \approx 1$. Indeed, the right chiral component of the coupling is the dominating factor, as the left-coupling is suppressed by a factor of $\sin(\alpha)$, which is of $\mathcal{O}(10^{-2})$ in our numerical analysis. The selected benchmark point, represented in black, for the doublet model, compatible with all flavour observables discussed above reads as

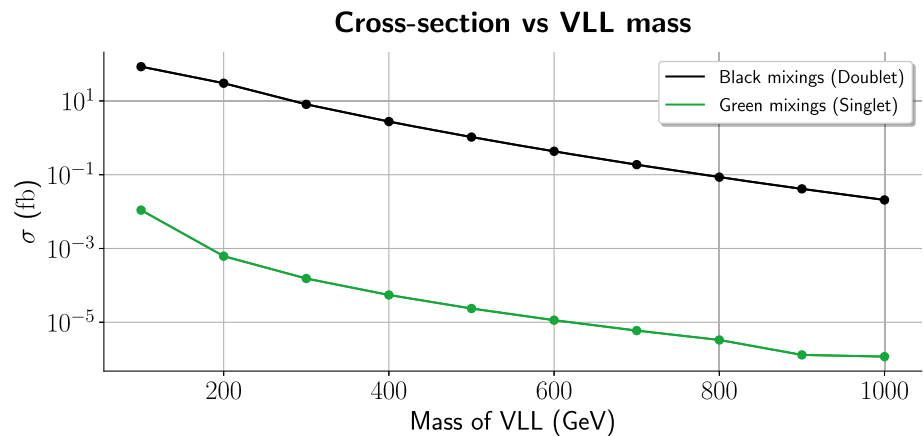
$$\begin{aligned}
 U_L^e &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.9997545 & 0 & -0.02215850 \\ 0 & 0 & 1 & 0 \\ 0 & -0.02215850 & 0 & 0.9997545 \end{bmatrix}, \\
 U_R^e &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.999842 & 0 & 0.0177691 \\ 0 & 0 & -1 & 0 \\ 0 & 0.0177691 & 0 & 0.999842 \end{bmatrix}, \\
 U_\nu &= \begin{bmatrix} U_{\text{PMNS}} & 0 \\ 0 & U_\nu^{\text{BSM}} \end{bmatrix}, \\
 U_\nu^{\text{BSM}} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.960772 & -0.277338 \\ 0 & 0.277338 & -0.960772 \end{bmatrix}, \tag{14}
 \end{aligned}$$

whereas for the singlet scenario, in green, one has

$$\begin{aligned}
 U_L^e &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.9995324 & 0 & -0.03057725 \\ 0 & 0 & 1 & 0 \\ 0 & -0.03057725 & 0 & 0.9995324 \end{bmatrix}, \\
 U_R^e &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1.57282 \times 10^{-5} \\ 0 & 0 & -1 & 0 \\ 0 & -1.57282 \times 10^{-5} & 0 & -1 \end{bmatrix}, \\
 U_\nu &= \begin{bmatrix} U_{\text{PMNS}} & 0 \\ 0 & 1 \end{bmatrix}. \tag{15}
 \end{aligned}$$

The neutrino mixing is identical in the two considered cases and therefore is always represented in black.

Fig. 6 Production cross section for both doublet and singlet topologies, in femtobarn (fb), as a function of the VLL's mass, in GeV. The y-axis is shown in logarithmic scale. The black and green curves correspond to the same colour mixing scenarios in Eqs. (14) and (15)



It follows from Eq. (12) and the numerical values shown above that, upon applying the selection criteria described in Sect. 3.2, the signal and background cross sections are

$$\begin{aligned}
 \sigma(\text{doublet}) &: 0.19 \text{ fb}, \\
 \sigma(\text{singlet}) &: 5.94 \times 10^{-6} \text{ fb}, \\
 \sigma(pp \rightarrow \mu^- \bar{\nu}_\mu + \text{up to 2 jets}) &: 4.79 \times 10^6 \text{ fb}, \\
 \sigma(t\bar{t}) &: 13616.16 \text{ fb}, \\
 \sigma(W + \text{jets}) &: 102448.54 \text{ fb}, \\
 \sigma(\text{Diboson} + \text{jets}) &: 296.94 \text{ fb},
 \end{aligned} \tag{16}$$

It is evident that the doublet model cross section is far larger than that of the singlet. This is indeed expected as the dominant contribution in the singlet model contains a $\sin \alpha$ suppression factor as one can see in Eq. (12). The dependency of the production cross section with the VLL mass is shown in Fig. 6. One can now estimate the total expected number of events, given the cross section above, after event selection. Assuming the target luminosity of the HL-LHC, $\mathcal{L} = 3000 \text{ fb}^{-1}$ and that the expected number of events is given by $N = \sigma \mathcal{L}$, we have for the doublet case $N = 570.0$ events, while for the singlet we obtain $N = 0.01782$ events. For the singlet model, this implies that the production cross section is not large enough to generate one event at the LHC. For this reason it is meaningless to present the statistical significance for the singlet model for such a heavy VLL. However, we see from Fig. 6 that, for lighter singlet VLLs, in particular for their masses of 100 and 200 GeV, one has $N \approx 30$ and $N \approx 3$, respectively. Thus, it becomes possible to produce them at the HL-LHC motivating a further full analysis.

The kinematic features used in the Deep Learning analysis can be seen in Figs. 10 and 11 of Appendix B for the doublet and singlet models, respectively. For the singlet model, the main variables allowing for a good discrimination between the signal and the background are the transverse momentum of the muon i.e., $p_T(\mu^-)$ as well as the MET distribution. These distributions are characterized by long tails at higher energies where the SM background is no longer present. On the other hand, the angular distributions for the cosine of the

polar angle, as well as the azimuthal angle, offer the least discriminating power. This follows from the fact that both signal and background have a similar shape. The pseudo-rapidity distributions can also be used in the discrimination since signal events tend to peak around $\eta \sim 0$ whereas the SM backgrounds spread out over the entire $|\eta| \leq 2.5$ region. In particular, for the $\eta(\mu^-)$ distributions, the SM backgrounds spread uniformly in the entire range. For the doublet model, we note that the kinematic distributions offer the best discriminating power, with both the p_T and mass distributions peaking at higher energies than those of the main irreducible backgrounds. While $\Delta\theta$ distributions closely follow the SM background prediction, the $\Delta\phi$ and ΔR distributions are distinct from those of the SM, with $\Delta\phi$ possessing a double peak structure near zero, whereas ΔR have its maximum at zero.

With this information at hand, one can construct multi-dimensional distributions to be used as inputs for a neural network that solves a classification task. This is done via an evolution algorithm to optimise the various hyper-parameters of the neural model, whose metric to be maximised is the Asimov statistical significance. For completeness, we present our results for different statistical models, some more conservative than others. In this article we use the same measures as in [9,37], which include:

1. The Asimov significance, \mathcal{Z}_A , with 1% systematic uncertainty,³ as defined in Eq. (13);
2. A less conservative version of the Asimov significance, which we dub as $\mathcal{Z}(< 1\%)$. In this measure, we assume

³ The dominant backgrounds of the analysis ($t\bar{t}$ and W +jets) have currently their cross sections measured with $\mathcal{O}(2\%)$ precision with an important contribution from the integrated luminosity measurement which strongly dominates the total uncertainty. When preparing for the European Strategy for Particle Physics reports that both ATLAS and CMS contributed to, a range of 1–1.5% was considered as a guideline for the luminosity uncertainty of the HL-LHC [54]. Furthermore, in a recent update [55] this recommendation was further improved to 0.83%, based on the full run-II accumulated dataset. As such, a global of 1% in our backgrounds can be seen as a realistic target for both run-III and the HL phase of the LHC.

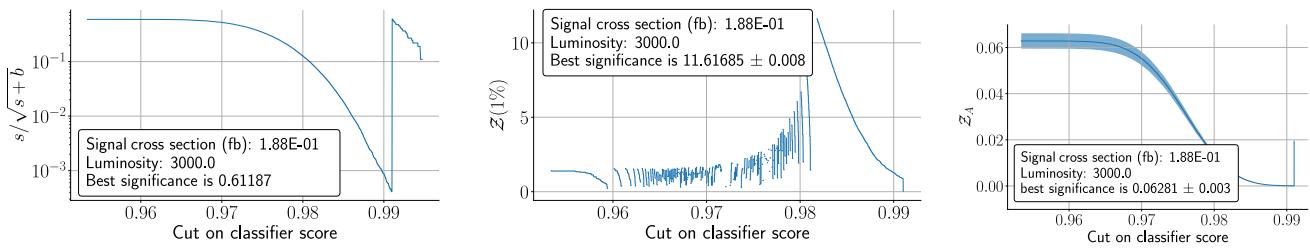
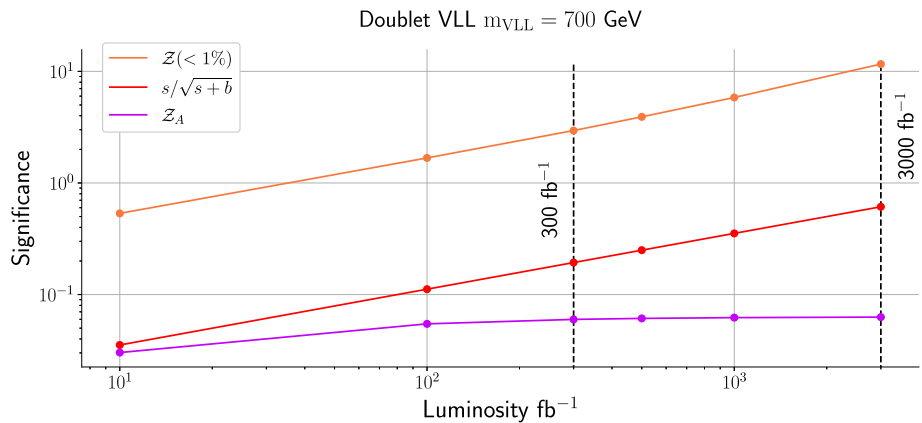


Fig. 7 Statistical significance as a function of the classifier score given by a neural network for different metrics, assuming a doublet-type VLL with $M_{VLL} = 700$ GeV and a collider luminosity $\mathcal{L} = 3000$ fb^{-1} . The computation of the statistical significance is made using the best neural

network that the evolution algorithm found. From left to right we plot the significance $s/\sqrt{s+b}$ in **a**, the adapted Asimov significance where we assume a background uncertainty of 10^{-3} in **b**, and the Asimov significance with systematics of 1% in **c**

Fig. 8 Statistical significance as a function of the collider luminosity, for a fixed doublet VLL mass of 700 GeV (mixing in Eq. (14)). Target luminosities of the HL-LHC (3000 fb^{-1}) and Run III (300 fb^{-1}) are marked with vertical dashed black lines. Both axes are shown in logarithmic scale



that backgrounds are known with an error of 10^{-3} . Of all measures, this is the most lenient one and typically offers the most significant results;

3. The more traditional metric, $s/\sqrt{s+b}$.

With this in mind, we compute the significances, for the doublet scenario, in a wide range of masses, from 100 to 1000 GeV, in steps of 100 GeV. In particular, we plot the various metrics as a function of the neural network score in Fig. 7, taking a VLL mass equal to 700 GeV, for illustration purposes. For this mass point and, for an integrated luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$, we obtain

$$\begin{aligned} s/\sqrt{s+b} &= 0.61\sigma, \\ \mathcal{Z}(< 1\%) &= 11.62\sigma, \\ \mathcal{Z}_A &= 0.063\sigma, \end{aligned} \tag{17}$$

where we note that we can exclude (or claim an hypothetical discovery) for a VLL with such a mass, since the $\mathcal{Z}(< 1\%)$ metric offers a statistical significance larger than 5σ . However, one must keep in mind that this metric is the least conservative of the those considered in this work. On the other hand, the Asimov metric is the stringiest and most conservative one resulting in a tiny significance for this particular point. It is also important to study the role of different values of the luminosity. As such, in Fig. 8, we show the evolution of the significance as a function of the collider’s luminosity for

a VLL with a mass of 700 GeV. In particular, we highlight with dashed black vertical lines the target luminosities at Run III (300 fb^{-1}) and at the HL-LHC (3000 fb^{-1}). Focusing on Run III, we obtain

$$\begin{aligned} s/\sqrt{s+b} &= 0.19\sigma, \\ \mathcal{Z}(< 1\%) &= 2.95\sigma, \\ \mathcal{Z}_A &= 0.059\sigma, \end{aligned} \tag{18}$$

such that, for a 700 GeV VLL, one does not expect any significant excess and therefore can not extract conclusions.

For an inclusive picture we perform a mass scan in the range $M_{VLL} \in [100, 1000]$ GeV, whose numerical values for the couplings are fixed to those shown in Eq. (14). On the other hand, it follows from the discussion in the previous paragraphs that only two example points are selected for the singlet model, both featuring the green mixing matrices in Eq. (15). We also vary the mass of the heavy left-handed neutrinos, present in the doublet model, such that they share the same mass with their charged $SU(2)_L$ doublet counterpart. In Table 5 our results for the scan are summarized, where the VLL masses and the calculated significance for both Run III and the HL-LHC upgrade are shown.

We notice that, both for $\mathcal{L} = 3000 \text{ fb}^{-1}$ or $\mathcal{L} = 300 \text{ fb}^{-1}$ as well as any of the mass values, we are able to obtain significances greater than 5σ if the simplified Asimov metric and $s/\sqrt{s+b}$ are considered. These results may indeed

Table 5 Signal significance for VLL single-production at the LHC calculated with an evolution algorithm that maximises the Asimov metric. The last two rows represent the singlet model whereas the remaining ten are benchmark points of the doublet model. The colour coding is the same as in Eqs. (14) and (15)

Mass of VLL (GeV)	300 fb ⁻¹			3000 fb ⁻¹		
	$s/\sqrt{s+b}$	$\mathcal{Z}(< 1\%)$	\mathcal{Z}_A	$s/\sqrt{s+b}$	$\mathcal{Z}(< 1\%)$	\mathcal{Z}_A
100	75.32	81.76	25.01	238.17	258.55	26.13
200	28.97	29.92	9.37	91.62	94.61	9.81
300	8.02	8.09	2.56	25.35	25.58	2.68
400	2.76	2.77	0.88	8.73	8.76	0.92
500	1.05	2.70	0.33	3.32	11.19	0.35
600	0.43	2.82	0.14	1.37	11.43	0.14
700	0.19	2.95	0.060	0.61	11.62	0.063
800	0.086	1.58	0.027	0.27	7.94	0.029
900	0.042	2.98	0.013	0.13	11.57	0.014
1000	0.021	1.84	0.0066	0.066	11.14	0.0069
100	0.94	1.33	0.00094	2.98	4.22	0.0023
200	0.25	0.36	0.00078	0.80	1.13	0.0023

Table 6 Signal significance for VLL single-production at the LHC calculated with the neural network trained in the 700 GeV data, with the network details shown in Table 9

Mass of VLL (GeV)	3000 fb ⁻¹		
	$s/\sqrt{s+b}$	$\mathcal{Z}(< 1\%)$	\mathcal{Z}_A
100	238.17	258.55	26.13
200	91.62	94.61	9.81
300	25.35	25.58	2.68
400	8.73	10.19	0.92
500	3.32	11.14	0.35
600	1.37	11.45	0.14
700	0.61	11.62	0.063
800	0.38	11.58	0.029
900	0.21	11.54	0.014
1000	0.26	11.14	0.0069

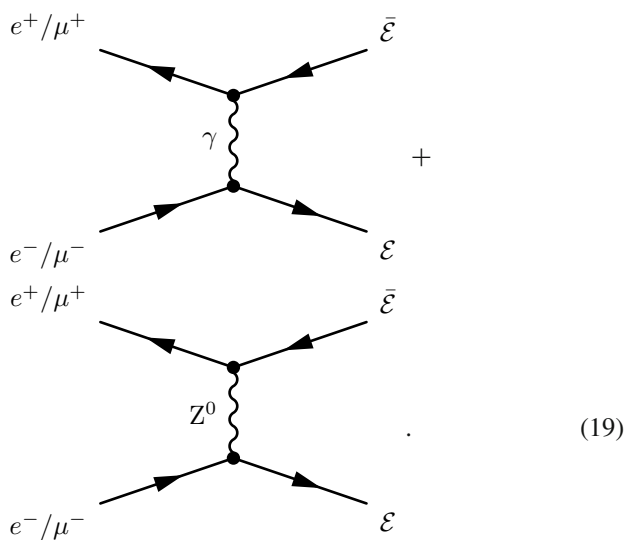
suggest that, significances above 5σ can be achievable even for larger VLL mass values at the LHC. This is particularly relevant in the doublet model while for singlet VLLs the significance quickly drops if we go beyond 200 GeV. If only the Asimov significance, \mathcal{Z}_A , which is the most conservative one, is considered we can still probe doublet VLLs up to about 200 GeV already at the LHC Run III. For the singlet scenario, considering the HL-LHC program, we can not exclude/discover VLLs, with the highest significance being 4.22σ for the $\mathcal{Z}(< 1\%)$ metric. While bellow the discovery threshold, it can still be regarded as a potential anomaly, whose significance can be increased with the combination of additional channels.

Although LEP constraints have already excluded VLLs up to 100.6 GeV, our results for 100 GeV are merely indicative of how large can the significance become for small doublet masses, if our analysis technique is employed.

As one can note, there is an overall trend of the significance dropping as the mass of the VLL decreases. However, we note that for masses above 600 GeV, we have consistently obtained significances close to 10σ for the $\mathcal{Z}(< 1\%)$ measure. Here, there are two main factors at play. First, considering larger masses, kinematic distributions tend to peak at higher energies, which is particularly relevant for mass distributions, enhancing the neural networks discriminating power. Additionally, we employ the evolutionary algorithm to every single point, meaning that each network is optimized to a specific phase space region.⁴ In a realistic search scenario we would not know the mass of the VLL and therefore it would not be reasonable to set a stronger preference in one of the various networks optimized towards different VLL masses. However, the networks that we use are engineered to be generic enough and can in principle be applied to distinct masses up to a certain discrimination power. To illustrate this, we apply the 700 GeV network to all scanned masses and present the results in Table 6. As one can see, the numerical values of the significance changed for the more lenient metrics, $s\sqrt{s+b}$ and $\mathcal{Z}(< 1\%)$, whereas for the most conservative one the significance remained the same. However, an increase was also experienced, in particular, for a VLL mass of 800 GeV, where the lenient Asimov metric grew from 7.84σ to 11.58σ . These results indicate that, despite the networks' training in distinct phase space regions, they are versatile enough to offer a good discrimination power, which also indicates absence of over-fitting. In essence, the hypothetical observation of a statistical excess for a given mass could motivate employing an optimized network to potentially enhance such a signal/excess.

⁴ The neural networks found to be optimized for each point are shown in Appendix C.

The low cross-sections of the singlet model call for a different approach on how to probe them. It is in this context that the near future electron and muon colliders can offer new opportunities. As it was mentioned in Sect. 3.2, production of particles via electroweak processes is favoured in these colliders. As such, it is instructive to understand how can a lepton collider enhance the cross-section for the case of pair-produced VLLs via an s -channel process. Note that the analysis that follows is independent of the chosen collider, as the s -channel cross-section is independent of the mass of the initial colliding particles. Therefore, all results shown here are valid both for the electron and muon machines. At LO, the main diagrams involved are shown in Fig. 5. Note that, since we are assuming non-zero couplings between the muon/electron and the VLL, there are also t -channel contributions of the form



However, such contributions are sub-leading when compared to production via the s -channel process, since these depend on mixing structures of the VLL with the muon, whereas in the s -channel the interaction vertices feature two leptons of the same flavour coupling directly to vectors via gauge interactions. Furthermore, in the particular case of our models, the approximate family symmetry that only allows couplings between muons and VLLs results in vanishing t -channel contributions at electron colliders, but not at muon machines. Note that this channel can be seen as a direct probe to the flavour structure in the leptonic sector if both muon and electron colliders become operational. It is then safe to neglect the t -channel contribution in the remainder of this analysis, where we use the same mixing structure as defined in Eq. (15). We then consider the $\mathcal{E} \rightarrow W\nu_\ell$ channel, which, depending on how the W bosons decay, can lead to the $2\nu_\ell + 4j$ or $2\nu_\ell + 2j + \ell\nu_\ell$ or $2\nu_\ell + \ell\nu_\ell + \ell\nu_\ell$ final states. For the topologies that involve jets as final states, the main backgrounds include diboson production (WW , WZ^0 and Z^0Z^0) with associated jets as well as tau pair production. For the

purely leptonic final state $2\nu_\ell + \ell\nu_\ell + \ell\nu_\ell$, diboson production is relevant. Indeed, a lepton collider is a much cleaner environment when compared to a hadronic machines in such a way that backgrounds involving jets can be safely discarded. VLLs can also decay into Z^0 bosons as $\mathcal{E}(\mathcal{E}) \rightarrow \mu Z^0$. Such a channel would be an important test for a hypothetical VLL discovery provided that the final states can contain at least 6 charged leptons. Besides being a very clean process, 6 lepton topologies are also expected to be small in the context of a SM background that typically results from triboson production processes. A detailed phenomenological analysis at lepton colliders using analogous deep learning methods to those discussed in the context of the LHC is beyond the scope of this article and is left for future work.

For a comprehensive understanding on how the cross section depends on a lepton collider center-of-mass energy, E_{CM} , we use `FeynCalc` to obtain the following expression

$$\sigma = \frac{\sqrt{\frac{E_{CM}^2}{4} - M_{\mathcal{E}}^2}}{256\pi E_{CM}^7 (M_{Z^0}^2 - E_{CM}^2)^2} \times (\alpha_1 E_{CM}^8 + \alpha_2 E_{CM}^6 M_{\mathcal{E}}^2 + \alpha_3 E_{CM}^6 M_{Z^0}^2 + \alpha_4 E_{CM}^4 M_{Z^0}^2 M_{\mathcal{E}}^2 + \alpha_5 E_{CM}^4 M_{Z^0}^4 + \alpha_6 E_{CM}^2 M_{Z^0}^4 M_{\mathcal{E}}^2), \tag{20}$$

where α_i for $i = 1, \dots, 6$ are dimensionless constants, proportional to the product of various couplings, i.e. $\alpha_j = \alpha_j(U_L^e, U_R^e, g, g', \theta_W)$, M_{Z^0} is the mass of the Z^0 boson and $M_{\mathcal{E}}$ the mass of the singlet VLL. With the numerical values in Eq. (15) these constants read as

$$\begin{aligned} \alpha_1 &= 0.189325, & \alpha_2 &= -0.213435, & \alpha_3 &= -0.328608, \\ \alpha_4 &= 0.403093, & \alpha_5 &= 0.152126, & \alpha_6 &= -0.209655. \end{aligned} \tag{21}$$

In Fig. 9 we plot the corresponding cross-section as a function of the centre-of-mass energy, for VLL masses in the range between 500 GeV and 6 TeV. Do note that we are studying the singlet scenario, where the decay width always remain below the mass of the VLL. It is interesting to note that the cross sections are rather large, above 13.23 fb for a VLL mass of 500 GeV and $E_{CM} = 1.5$ TeV, dropping to 0.095 fb for a VLL mass of 6TeV and $E_{CM} = 14$ TeV. This increase of the cross section allows to probe higher mass ranges than the ones at the reach of the LHC, in a much cleaner environment. Immediately noticeable is the fact that the cross section hits a maximum shortly after $E_{CM} \sim 2M_{VLL}$, with a subsequent drop as E_{CM} increases. This implies that when the collider beam energy is two times the mass of the VLL, the double production of VLL is enhanced and the discovery potential is maximized. For lower masses, the drop is more pronounced when compared to the high-mass regime, essentially because we are taking E_{CM} in the range of 3 – 14 TeV for low masses,

Cross-section vs Centre of mass energy

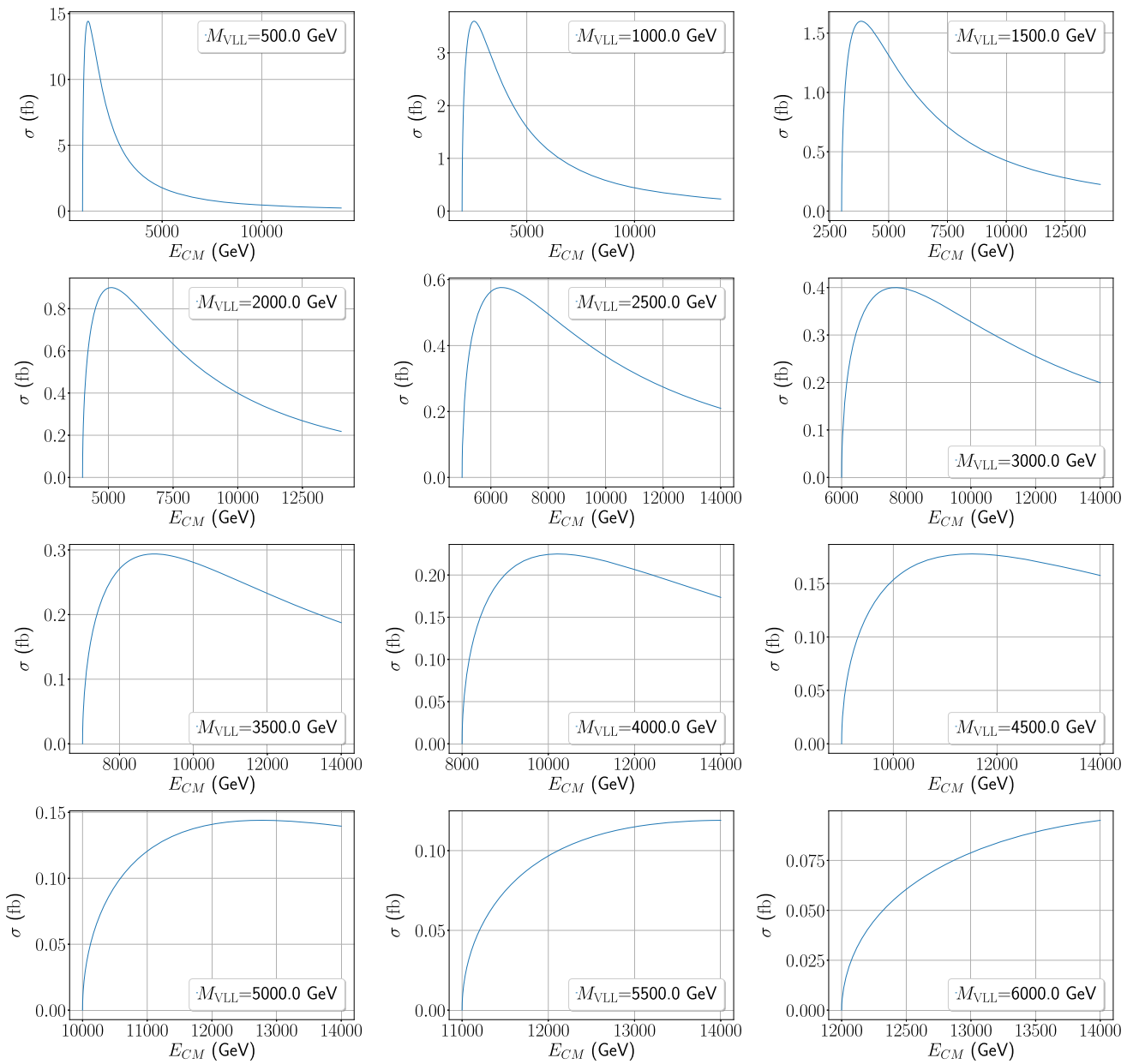


Fig. 9 VLL pair production cross section, in fb, as a function of centre-of-mass energy for a lepton collider, in GeV. Each individual box is representative of a different mass of the singlet VLL, which is also given in GeV

while for high masses it occurs for much higher centre-of-mass energies (beyond $E_{CM} = 14$ TeV), and therefore not as relevant for the proposed lepton colliders. In particular, for $M_{VLL} = 500$ GeV and at $E_{CM} = 14$ TeV we have $\sigma = 0.23$ fb, while for the same mass, at $E_{CM} = 3$ TeV, we have $\sigma = 4.63$ fb.

Fixing the centre-of-mass energy and looking at various mass points, as shown in Table 7, we notice that for the proposed high-energy colliders, the variation of the mass does not cause significant deviations in the cross-section. For example, taking $E_{CM} = 14$ TeV, we observe that the cross-section remains nearly constant for the displayed masses, ranging from 0.23 fb for a 200 GeV VLL to 0.18 fb for a 3.7 TeV one.

We end this section with a comment about the cross-sections at the high energy end of the colliders. As the energy grows the s -channel cross sections decrease. However, there is an alternative process that grows with $\ln^2(s/m_f^2)$, where f stands for the incoming fermion. This is $e^+e^- \rightarrow e^+e^-E(\mathcal{E})\bar{E}(\bar{\mathcal{E}})$ for an electron-positron collider and $\mu^+\mu^- \rightarrow \mu^+\mu^-E(\mathcal{E})\bar{E}(\bar{\mathcal{E}})$ for a muon collider. The photon fusion processes have cross-sections that grow with the centre-of-mass energy [56,57] and although they are not competitive for the lower energies they become dominant at high energies. We present in Table 7 the values of these cross-sections for an energy of 14 TeV which shows that they can play an important role for very high energy lepton colliders.

5 Conclusions

In this article we have studied two simple SM extensions featuring, each of them, a new vector-like lepton and a sterile neutrino. We have confronted the cases of a doublet and a singlet VLL and discussed their collider phenomenology, both at the LHC and future leptonic machines. For the former we have employed Deep Learning techniques to compute

the statistical significance of a hypothetical discovery. In the selection of benchmark scenarios we have required that the coupling of VLL to muons is consistent with flavour constraints and the branching fractions of Higgs and Z bosons to muons. We have also shown that, decay width of both VLL doublet components becomes increasingly larger with growing exotic lepton masses. In our analysis we have only considered scenarios where the width is smaller than the mass.

In the context of the LHC studies, we have performed Monte-Carlo simulations to generate data for signal and background topologies. The signal is characterized by the presence of a single isolated lepton and a substantial amount of MET for the singlet case. For the doublet case, the signal involves 4 light jets, a charged lepton and a neutrino as final state particles. To separate the signal from the background we constructed neural networks which follow from an implementation of an evolution algorithm that maximises the Asimov significance. We have shown that, for doublet VLLs, we can exclude masses up to 1 TeV with more than five standard deviations. In particular, for masses of 1 TeV, we obtain a significance of $\mathcal{Z}(< 1\%) = 11.14\sigma$ for the high luminosity phase of the LHC, with an integrated luminosity of $\mathcal{L} = 3000$ fb⁻¹. We have also verified that one can already test the doublet VLL scenario at the Run III of the LHC, which will deliver $\mathcal{L} = 300$ fb⁻¹ of data. For such luminosity, one can test VLL masses up to about 300 GeV obtaining $s/\sqrt{s+b} = 8.02\sigma$, $\mathcal{Z}(< 1\%) = 8.09\sigma$ and $\mathcal{Z}_A = 2.56\sigma$. For the singlet scenario, production cross-sections are substantially smaller and the expected number of events is usually zero, with an exception for VLL masses of 100 GeV and 200 GeV. While for the former one finds $s/\sqrt{s+b} = 2.98\sigma$, $\mathcal{Z}(< 1\%) = 4.22\sigma$ and $\mathcal{Z}_A = 0.0023\sigma$ at $\mathcal{L} = 3000$ fb⁻¹, the run III estimation gives $s/\sqrt{s+b} = 0.94\sigma$, $\mathcal{Z}(< 1\%) = 1.33\sigma$ and $\mathcal{Z}_A = 0.00094\sigma$ at $\mathcal{L} = 300$ fb⁻¹, experiencing a great drop for a mass of 200 GeV. In particular, the latter scenario can not be excluded, with a statistical significance bellow the discovery threshold.

Table 7 Singlet VLL double production cross-section at lepton colliders. As indicated in the top row of the table, all considered mass values correspond to the bold benchmark point in Eq. (15). We show the cross-section for a centre-of-mass energy $E_{CM} = 1.5$ and 3 TeV, which are

	200 GeV	700 GeV	1200 GeV	1700 GeV	2200 GeV	2700 GeV	3200 GeV	3700 GeV
$E_{CM} = 1.5$ TeV	19.18 fb	5.51 fb	–	–	–	–	–	–
$E_{CM} = 3$ TeV	5.00 fb	4.21 fb	2.49 fb	–	–	–	–	–
$E_{CM} = 10$ TeV	0.46 fb	0.45 fb	0.44 fb	0.41 fb	0.39 fb	0.35 fb	0.31 fb	0.26 fb
$E_{CM} = 14$ TeV	0.23 fb	0.23 fb	0.23 fb	0.22 fb	0.21 fb	0.21 fb	0.19 fb	0.18 fb
$E_{CM} = 14$ TeV ($e^+e^- \gamma\gamma$)	17.95 fb	0.54 fb	0.095 fb	0.027 fb	0.0095 fb	0.0037 fb	0.001585 fb	0.00069 fb
$E_{CM} = 14$ TeV ($\mu^+\mu^- \gamma\gamma$)	9.10 fb	0.26 fb	0.045 fb	0.012 fb	0.0043 fb	0.0017 fb	0.00070 fb	0.00030 fb

the target energies for the CLIC collider, whereas $E_{CM} = 3, 10$ and 14 TeV correspond to the current proposal for the future $\mu^+\mu^-$ collider. Points marked with “–” indicate that there is not enough energy to pair-produce the particles at that mass

Owing to the low production cross-sections of the singlet scenario, a supplementary analysis was made within the context of lepton colliders. We have performed numerical computations for the expected VLL pair-production cross-section in the s -channel. We find that, in general, larger cross-sections are obtained when compared to the LHC analysis, allowing for a much wider range of masses to be probed with relevance for singlet VLLs. In particular, we note that for luminosities of the order of ab^{-1} the, even for singlet VLLs with mass 3.7 TeV one can expect 260 events at $E_{\text{CM}} = 10$ TeV and 180 events if $E_{\text{CM}} = 14$ TeV. Furthermore, even for a VLL as heavy as 6 TeV, a 14 TeV lepton machine delivering a luminosity of 1 ab^{-1} is expected to produce almost 100 events. With this in mind we conclude that the study of VLL particles at future electron or muon colliders is a rather relevant physics case scenario to be explored, allowing to significantly extend the current reach of the LHC.

Acknowledgements The authors would like to thank Celso Nishi for valuable comments made to the initial version of this manuscript. J.G., F.F.F., and A.P.M. 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. A.P.M., F.F.F., J.G. and R.S. are supported by the project PTDC/FIS-PAR/31000/2017. A.P.M., F.F.F., J.G. are also supported by the projects CERN/FIS-PAR/0021/2021. Additionally, A.P.M. and J.G. are supported by CERN/FIS-PAR/0019/2021. J.G. is also directly funded by FCT through the doctoral program grant with the reference 2021.04527.BD. A.P.M. is also supported 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 grant, contract number 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). R.S. is supported by CFTC-UL under FCT contracts UIDB/00618/2020, UIDP/00618/2020, and by the projects CERN/FISPAR /0002/2017, CERN/FIS-PAR/0014/2019 and by the HARMONIA project of the National Science Centre, Poland, under contract UMO- 2015/18/M/ST2/00518. A.O. is supported by the FCT project CERN/FIS-PAR/0029/2019.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: Due to the big memory size of the root files, they are not provided. Instead, csv data files used in the numerics are provided and can be found in one of the author’s GitHub page https://github.com/Mrazi09/VLL_collider.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-

right holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

Appendix A: Feynman Rules

In this appendix, the list of Feynman rules relevant for the numerical analysis is presented. All rules are displayed in the mass basis. To simplify the notation, we define the chirality projection operators as $P_L = (1 - \gamma_5)/2$ and $P_R = (1 + \gamma_5)/2$. Notation wise, we define the following: θ_W is the Weinberg mixing angle, g the $SU(2)_L$ gauge coupling and g' the $U(1)_Y$ gauge coupling. The Latin indices $i, j = (1, 2, 3, 4)$ denote $(e, \mu, \tau, E(\mathcal{E}))$ for leptons, while for neutrinos $i, j = (1, 2, 3, 4, 5, 6) = (v_1, v_2, v_3, v_4, v_5, v_6)$, with v_5 and v_6 arising from the doublet representation of the VLL and v_4 is the sterile neutrino. For the singlet model, there is no extra left-handed neutrinos, hence the latin indices only run from 1 to 4.

For the doublet case, the Feynman rules read as

$$= i \left(g_L^Z \gamma_\mu P_L + g_R^Z \gamma_\mu P_R \right),$$

$$g_L^Z = \frac{1}{2} (g \cos \theta_W - g' \sin \theta_W)$$

$$\left[(U_L^e)^*_{j4} (U_L^e)_{i4} + \sum_{a=1}^3 (U_L^e)^*_{ja} (U_L^e)_{ia} \right],$$

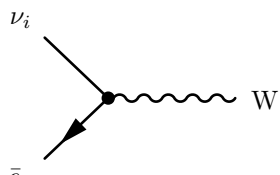
$$g_R^Z = -g' \sin(\theta_W) \sum_{a=1}^3 (U_R^e)^*_{ia} (U_R^e)_{ja}$$

$$+ \frac{1}{2} (U_R^e)^*_{i4} (U_R^e)_{j4} (g \cos \theta_W - g' \sin \theta_W), \quad (A1)$$

$$= i \left(g_L^{W^+} \gamma_\mu P_L + g_R^{W^+} \gamma_\mu P_R \right),$$

$$g_L^{W^+} = -\frac{g}{\sqrt{2}} \left[(U_L^e)^*_{j4} (U_\nu)_{i5} + \sum_{a=1}^3 (U_L^e)^*_{ja} (U_\nu)_{ia} \right],$$

$$g_R^{W^+} = -\frac{g}{\sqrt{2}} (U_\nu)_{i6}^* (U_R^e)_{j4}, \quad (A2)$$

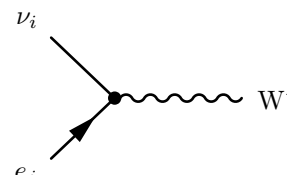


$$= i \left(g_L^{W-} \gamma_\mu P_L + g_R^{W-} \gamma_\mu P_R \right),$$

$$g_L^{W-} = -\frac{g}{\sqrt{2}} \left[(U_\nu)_{i5}^* (U_L^e)_{j4} + \sum_{a=1}^3 (U_\nu)_{ia}^* (U_L^e)_{ja} \right],$$

$$g_R^{W-} = -\frac{g}{\sqrt{2}} (U_R^e)_{j4}^* (U_\nu)_{i6},$$

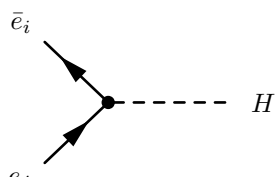
(A3)



$$= i g_L^{W+} \gamma_\mu P_L,$$

$$g_L^{W+} = -\frac{g}{\sqrt{2}} \sum_{a=1}^3 (U_L^e)_{ja}^* (U_\nu)_{ia},$$

(A6)

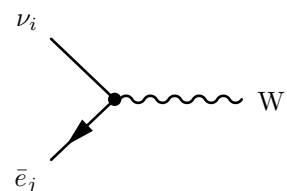


$$= i \left(g_L^H P_L + g_R^H P_R \right),$$

$$g_L^H = \frac{1}{\sqrt{2}} \left[\sum_{a=1}^3 (U_L^e)_{j4}^* (U_R^e)_{ia}^* \Theta_a + \sum_{a=1}^3 \sum_{b=1}^3 (U_R^e)_{ib}^* (U_L^e)_{ja}^* \Pi_{ab} \right],$$

$$g_R^H = \frac{1}{\sqrt{2}} \left[\sum_{a=1}^3 (U_R^e)_{ja} (U_L^e)_{i4} \Theta_a^* + \sum_{a=1}^3 \sum_{b=1}^3 (U_L^e)_{ia} (U_R^e)_{jb} \Pi_{ab}^* \right],$$

(A4)

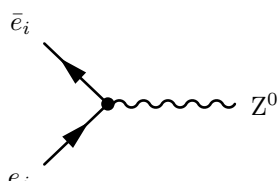


$$= i g_L^{W-} \gamma_\mu P_L,$$

$$g_L^{W-} = -\frac{g}{\sqrt{2}} \sum_{a=1}^3 (U_\nu)_{ia}^* (U_L^e)_{ja},$$

(A7)

Similarly, for the singlet case we have,

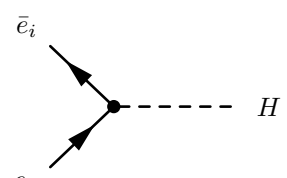


$$= i \left(g_L^Z \gamma_\mu P_L + g_R^Z \gamma_\mu P_R \right),$$

$$g_L^Z = -g' (U_L^e)_{j4}^* (U_L^e)_{i4} \sin(\theta_W) + \frac{1}{2} (g \cos \theta_W - g' \sin \theta_W) \sum_{a=1}^3 (U_L^e)_{ja}^* (U_L^e)_{ia},$$

$$g_R^Z = -g' \sin(\theta_W) \left[(U_R^e)_{i4}^* (U_R^e)_{j4} + \sum_{a=1}^3 (U_R^e)_{ia}^* (U_R^e)_{ja} \right],$$

(A5)



$$= i \left(g_L^H P_L + g_R^H P_R \right),$$

$$g_L^H = \frac{1}{\sqrt{2}} \left[\sum_{a=1}^3 (U_R^e)_{i4}^* (U_L^e)_{ja}^* \theta_a + \sum_{a=1}^3 \sum_{b=1}^3 (U_R^e)_{ib}^* (U_L^e)_{ja}^* \Pi_{ab} \right],$$

$$g_R^H = \frac{1}{\sqrt{2}} \left[\sum_{a=1}^3 (U_L^e)_{ia} (U_R^e)_{j4} \theta_a^* + \sum_{a=1}^3 \sum_{b=1}^3 (U_L^e)_{ia} (U_R^e)_{jb} \Pi_{ab}^* \right].$$

(A8)

Appendix B: Kinematic distributions for VLL single-production

Single production of the doublet VLL with mass $M = 700$ GeV

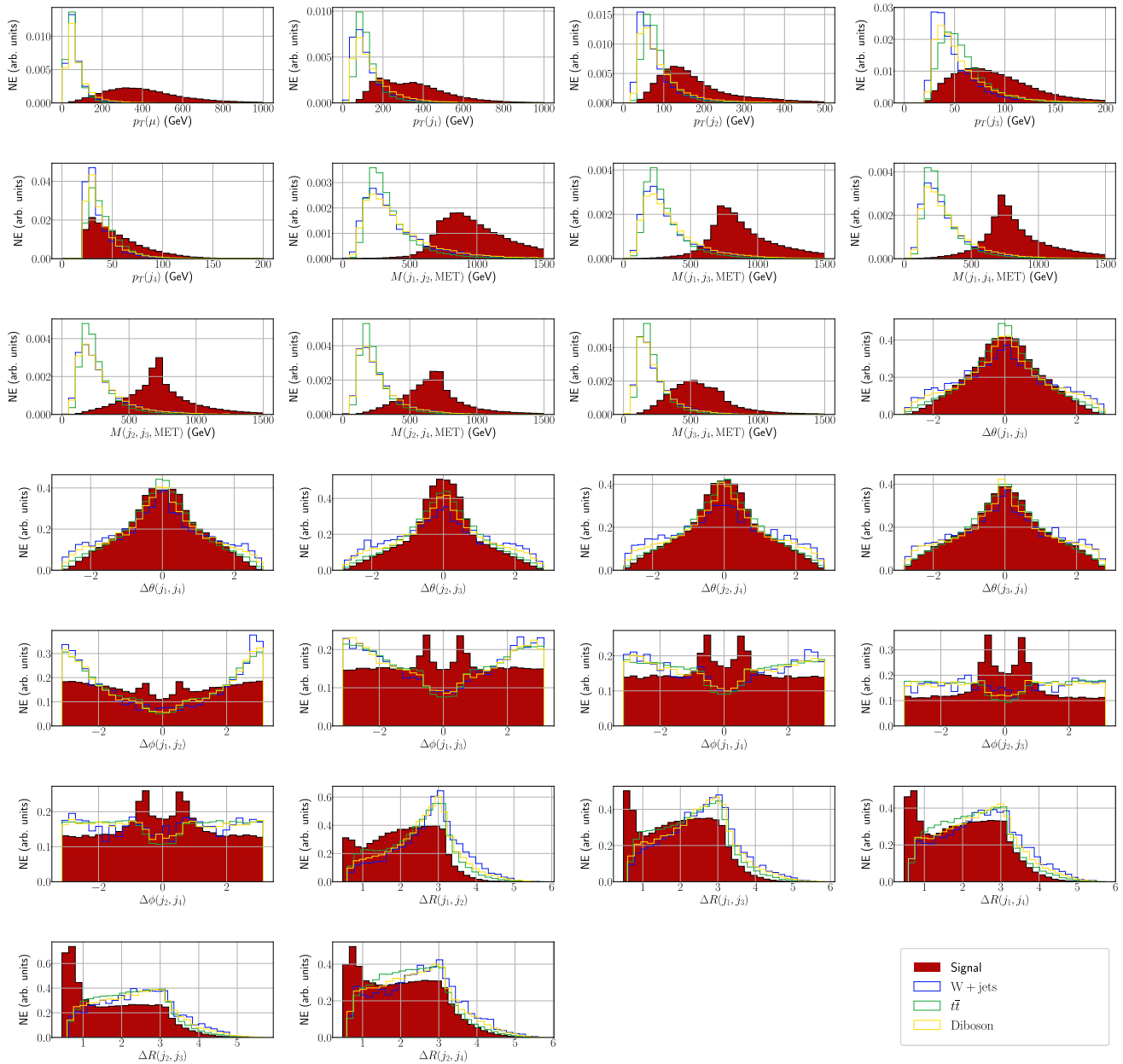


Fig. 10 A sample of the kinematic variables that the neural network uses for classification. Data is simulated for a doublet VLL with mass of 700 GeV and in the ATLAS detector. From left to right and top to bottom, the variables are p_T of the muon, p_T of jet 1, 2, 3 and 4, mass distributions for the combinations (j_1, j_2, MET) , (j_1, j_3, MET) , (j_1, j_4, MET) , (j_2, j_3, MET) , (j_2, j_4, MET) , (j_3, j_4, MET) , $\Delta\theta$ dis-

tributions for the combinations (j_1, j_3) , (j_1, j_4) , (j_2, j_3) , (j_2, j_4) and (j_3, j_4) , $\Delta\phi$ distributions for the combinations (j_1, j_2) , (j_1, j_3) , (j_1, j_4) , (j_2, j_3) , (j_2, j_4) and (j_3, j_4) and ΔR distributions for the combinations (j_1, j_2) , (j_1, j_3) , (j_1, j_4) , (j_2, j_3) and (j_2, j_4) . In the y-axis, it is indicated that events are normalized (NE). We consider 30 bins in all background and signal histograms

Single production of singlet VLL with mass $M = 700$ GeV

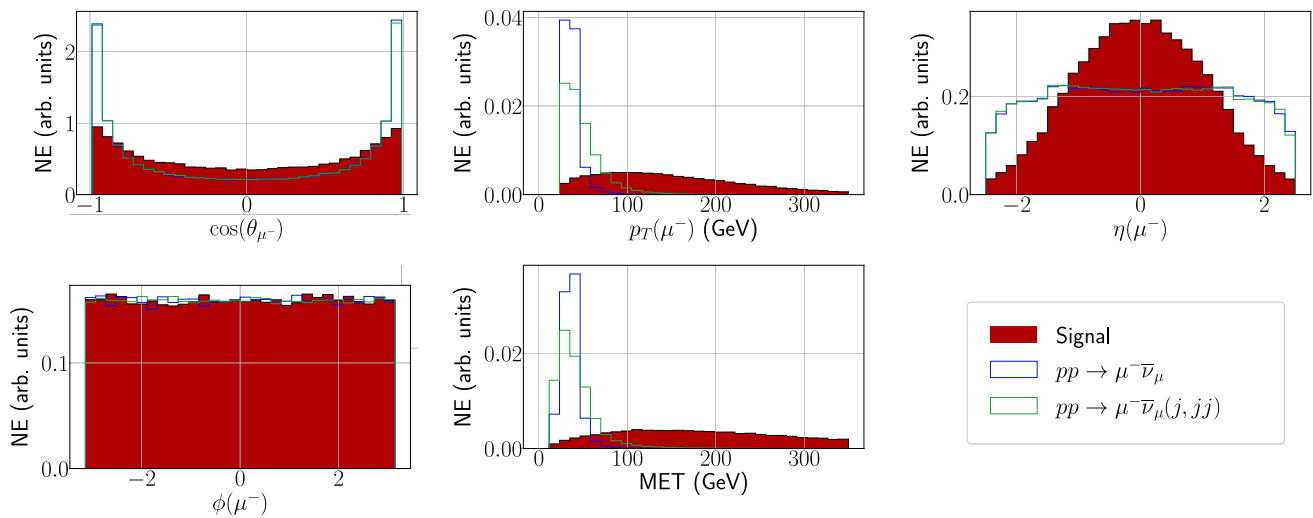


Fig. 11 Kinematic variables that the neural network uses for classification. Data is simulated for a doublet VLL with mass of 700 GeV and in the ATLAS detector. From left to right and top to bottom, the variables are $\cos(\theta_{\mu^-})$, transverse momentum of the

muon, pseudo-rapidity of the muon, azimuthal angle of the muon and MET. In the y-axis, it is indicated that events are normalized (NE). We consider 30 bins for all background and signal histograms

Appendix C: Neural networks found via the genetic algorithm

Table 8 Architectures for the neural networks that the genetic algorithm finds for the scanned masses (100–500 GeV) of the doublet VLL scenario

Mass	Doublet
100 GeV	Layers: 1 input + 1 hidden + 1 output. Input and hidden layer with 1024 neurons, output layer with 4 neurons Regularizer: L1L2 (layer 1 and 2) and none (layer 3) Activation function: RELU (for layers 1 and 2) and Sigmoid (layer 3) Initializer: VarianceScaling, with uniform distribution (layer 1 and 2) in fan_in mode and GlorotUniform (layer 3) with null seed
200 GeV	Layers: 1 input + 3 hidden + 1 output. Hidden and input layers with 1024 neurons each, output layer with 4 neurons Regularizer: L1L2 (layers 1–4) and none (for layer 5) Activation function: RELU (layers 1–4) and Sigmoid (layer 5) Initializer: VarianceScaling, with uniform distribution (layers 1–4) in fan_in mode and GlorotUniform (layer 5) with null seed
300 GeV	Layers: 1 input + 4 hidden + 1 output Input and hidden layers with 512 neurons, output layer with 4 neurons Regularizer: L1L2 (for layers 1–5) and none (for layer 6) Activation function: Sigmoid (for all layers) Initializer: VarianceScaling, with truncated normal distribution (for layers 1–5) in fan_in mode and GlorotUniform distribution (layer 6) with null seed
400 GeV	Layers: 1 input + 2 hidden + 1 output Hidden and input layers with 256 neurons each, output layer with 4 neurons Regularizer: L1L2 (for layers 1–3) and none (for layer 4) Activation function: RELU (layers 1–3) and Sigmoid (layer 4) Initializer: RandomNormal, with 0 mean and 0.05 standard deviation (for layers 1–3) and GlorotUniform distribution (layer 4) with null seed
500 GeV	Layers: 1 input + 1 output Input layer with 1024 neurons, output layer with 4 neurons Regularizer: L1L2 (layer 1) and none (layer 2) Activation function: Sigmoid (all layers) Initializer: VarianceScaling, with truncated normal distribution (layer 1) in fan_in mode and GlorotUniform distribution (layer 2) with null seed

Table 9 Architectures for the neural networks that the genetic algorithm finds for the scanned masses (600–1000 GeV) of the doublet VLL scenario

Mass	Doublet
600 GeV	<p>Layers: 1 input + 4 hidden + 1 output</p> <p>Hidden and input layers with 512 neurons each, output layer with 4 neurons</p> <p>Regularizer: L1L2 (for layers 1–5) and none (for layer 6)</p> <p>Activation function: Sigmoid (for all layers)</p> <p>Initializer: RandomNormal, with mean 0 and 0.05 standard deviation (for layers 1–5) and GlorotUniform distribution (layer 6) with null seed</p>
700 GeV	<p>Layers: 1 input + 3 hidden + 1 output</p> <p>Hidden and input layers with 512 neurons each, output layer with 4 neurons</p> <p>Regularizer: L1L2 (for layers 1–4) and none (for layer 5)</p> <p>Activation function: Sigmoid (for all layers)</p> <p>Initializer: VarianceScaling, with uniform distribution (for layers 1–4) in fan_in mode and GlorotUniform distribution (layer 5) with null seed</p>
800 GeV	<p>Layers: 1 input + 3 hidden + 1 output</p> <p>Hidden and input layers with 256 neurons each, output layer with 4 neurons</p> <p>Regularizer: L1L2 (for layers 1 and 4) and none (for layer 5)</p> <p>Activation function: ELU (for layers 1–4) and Sigmoid (layer 5)</p> <p>Initializer: VarianceScaling, with truncated normal distribution (for layers 1–4) in fan_in mode and GlorotUniform distribution (layer 5) with null seed</p>
900 GeV	<p>Layers: 1 input + 1 hidden + 1 output</p> <p>Hidden and input layers with 256 neurons each, output layer with 4 neurons</p> <p>Regularizer: L1L2 (for layers 1 and 2) and none (for layer 3)</p> <p>Activation function: Tanh (for layers 1 and 2) and Sigmoid (layer 3)</p> <p>Initializer: RandomNormal, with mean 0 and 0.05 standard deviation (for layers 1 and 2) and GlorotUniform distribution (layer 3) with null seed</p>
1000 GeV	<p>Layers: 1 input + 2 hidden + 1 output</p> <p>Hidden and input layers with 2048 neurons each, output layer with 4 neurons</p> <p>Regularizer: L1L2 (for layers 1–3) and none (for layer 4)</p> <p>Activation function: ELU (for layers 1–3) and Sigmoid (layer 4)</p> <p>Initializer: VarianceScaling, with uniform distribution (for layers 1–3) in fan_in mode and GlorotUniform distribution (layer 4) with null seed</p>

Table 10 Architectures for the neural networks that the genetic algorithm finds for the scanned masses (100 and 200 GeV) of the singlet VLL scenario

Mass	Singlet
100 GeV	<p>Layers: 1 input + 1 hidden + 1 output</p> <p>Hidden and input layers with 512 neurons each, output layer with 3 neurons</p> <p>Regularizer: L1L2 (for layers 1–4) and none (for layer 5)</p> <p>Activation function: sigmoid (for all layers)</p> <p>Initializer: RandomNormal, with 0 mean and 0.05 standard deviation (for layers 1 and 2) and VarianceScaling, with uniform distribution in fan_in mode (last layer)</p>
200 GeV	<p>Layers: 1 input + 1 hidden + 1 output</p> <p>Hidden and input layers with 2048 neurons each, output layer with 3 neurons</p> <p>Regularizer: L1L2 (for layers 1 and 2) and none (for layer 3)</p> <p>Activation function: sigmoid (for all layers)</p> <p>Initializer: VarianceScaling, with uniform distribution (layer 1 and 2) in fan_in mode and fan_avg mode (last layer)</p>

Appendix D: ROC plots for Doublet VLLs

Fig. 12 Receiver operator characteristic (ROC) plots for the neural networks of the doublet VLLs from the masses of 100 GeV (top) to 500 GeV (bottom)

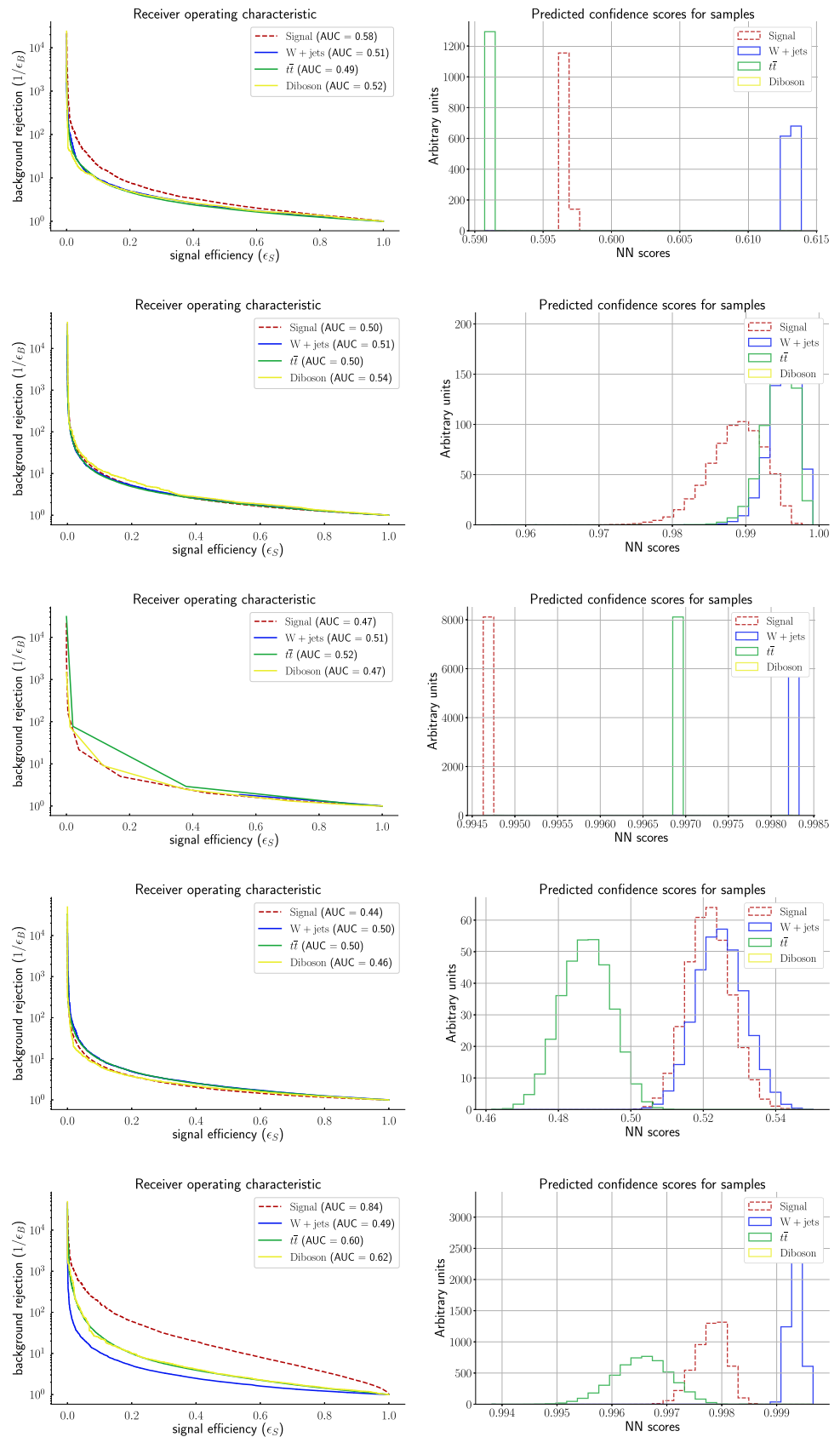
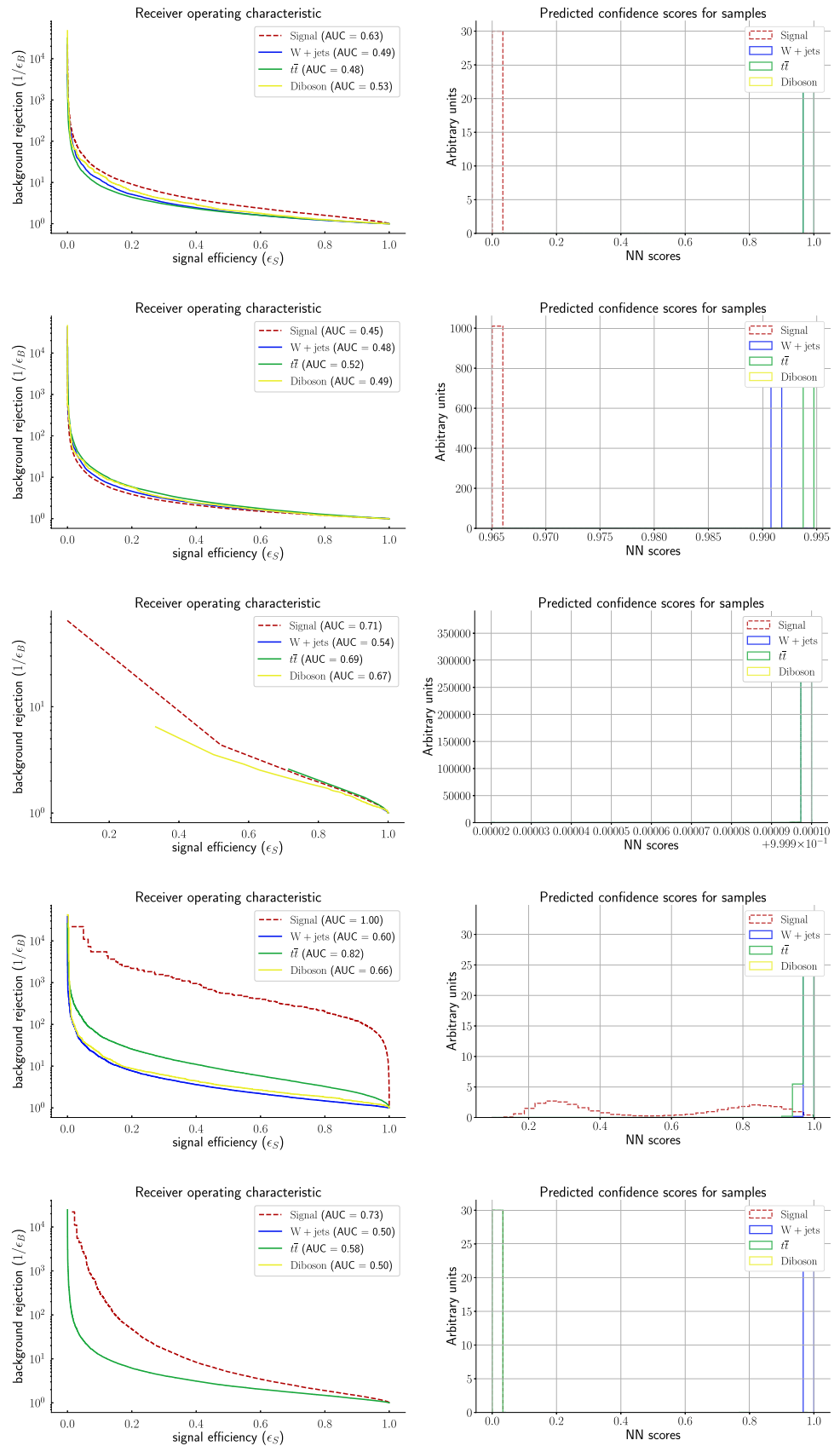


Fig. 13 Receiver operator characteristic (ROC) plots for the neural networks of the doublet VLLs from the masses of 600 GeV (top) to 1000 GeV (bottom)



References

1. Y. Fukuda et al., (Super-Kamiokande), Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.* **81**, 1562–1567 (1998). [arXiv:hep-ex/9807003](#)
2. G. Bertone, D. Hooper, J. Silk, Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.* **405**, 279–390 (2005). [arXiv:hep-ph/0404175](#)
3. B. Abi et al., (Muon g-2), Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.* **126**, 141801 (2021). [arXiv:2104.03281](#) [hep-ex]
4. R. Aaij et al. (LHCb), Test of lepton universality in beauty-quark decays (2021). [arXiv:2103.11769](#) [hep-ex]
5. A.P. Morais, R. Pasechnik, W. Porod, Prospects for new physics from gauge left-right-colour-family grand unification hypothesis. *Eur. Phys. J. C* **80**, 1162 (2020). [arXiv:2001.06383](#) [hep-ph]
6. F. J. de Anda, A. Aranda, A. P. Morais, R. Pasechnik, Gauge couplings evolution from the Standard Model, through Pati-Salam theory, into E8 unification of families and forces (2020). [arXiv:2011.13902](#) [hep-ph]
7. A. Aranda, F. J. de Anda, A. P. Morais, R. Pasechnik, A Different Take on E8 Unification (2021). [arXiv:2107.05421](#) [hep-ph]
8. A. Aranda, F. J. de Anda, A. P. Morais, R. Pasechnik, Sculpting the standard model from low-scale Gauge-Higgs-Matter E8 grand unification in ten dimensions (2021). [arXiv:2107.05495](#) [hep-ph]
9. F.F. Freitas, J. Gonçalves, A.P. Morais, R. Pasechnik, Phenomenology of vector-like leptons with Deep Learning at the Large Hadron Collider. *JHEP* **01**, 076 (2021). [arXiv:2010.01307](#) [hep-ph]
10. W. Yin, M. Yamaguchi, Muon g - 2 at multi-TeV muon collider (2020). [arXiv:2012.03928](#) [hep-ph]
11. A. Crivellin, F. Kirk, C.A. Manzari, M. Montull, Global electroweak fit and vector-like leptons in light of the Cabibbo angle anomaly. *JHEP* **12**, 166 (2020). [arXiv:2008.01113](#) [hep-ph]
12. S. Bißmann, G. Hiller, C. Hormigos-Feliu, D.F. Litim, Multi-lepton signatures of vector-like leptons with flavor. *Eur. Phys. J. C* **81**, 101 (2021). [arXiv:2011.12964](#) [hep-ph]
13. G. Hiller, C. Hormigos-Feliu, D.F. Litim, T. Stuedtner, Anomalous magnetic moments from asymptotic safety. *Phys. Rev. D* **102**, 071901 (2020). [arXiv:1910.14062](#) [hep-ph]
14. G. Hiller, C. Hormigos-Feliu, D.F. Litim, T. Stuedtner, Model building from asymptotic safety with Higgs and flavor portals. *Phys. Rev. D* **102**, 095023 (2020). [arXiv:2008.08606](#) [hep-ph]
15. G. Guedes, J. Santiago, New leptons with exotic decays: collider limits and dark matter complementarity (2021). [arXiv:2107.03429](#) [hep-ph]
16. O. Eberhardt, G. Herbert, H. Lacker, A. Lenz, A. Menzel, U. Nierste, M. Wiebusch, Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations. *Phys. Rev. Lett.* **109**, 241802 (2012). [arXiv:1209.1101](#) [hep-ph]
17. P. A. Zyla et al. (Particle Data Group), *Rev. Part. Phys. PTEP* **2020**, 083C01 (2020)
18. A. Strumia, F. Vissani, Neutrino masses and mixings and... (2006). [arXiv:hep-ph/0606054](#)
19. P. Achard et al., (L3), Search for heavy neutral and charged leptons in e^+e^- annihilation at LEP. *Phys. Lett. B* **517**, 75–85 (2001). [arXiv:hep-ex/0107015](#)
20. A. M. Sirunyan et al. (CMS), Search for vector-like leptons in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D* **100**, 052003 (2019). [arXiv:1905.10853](#) [hep-ex]
21. G. Aad et al., (ATLAS), Search for heavy lepton resonances decaying to a Z boson and a lepton in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP* **09**, 108 (2015). [arXiv:1506.01291](#) [hep-ex]
22. S. Chatrchyan et al., (CMS), Search for Heavy Lepton Partners of Neutrinos in Proton-Proton Collisions in the Context of the Type III Seesaw Mechanism. *Phys. Lett. B* **718**, 348–368 (2012). [arXiv:1210.1797](#) [hep-ex]
23. I. Esteban, M.C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, A. Zhou, The fate of hints: updated global analysis of three-flavor neutrino oscillations. *JHEP* **09**, 178 (2020). [arXiv:2007.14792](#) [hep-ph]
24. F. del Aguila, J. de Blas, M. Perez-Victoria, Effects of new leptons in Electroweak Precision Data. *Phys. Rev. D* **78**, 013010 (2008). [arXiv:0803.4008](#) [hep-ph]
25. V. Kuksa, N. Volchanskiy, Factorization in the model of unstable particles with smeared mass. *Central Eur. J. Phys.* **11**, 182–194 (2013). [arXiv:1109.1541](#) [hep-ph]
26. V. I. Kuksa, Convolution formula for decay rate, *Phys. Lett. B* **633**, 545–549 (2006), [Erratum: *Phys.Lett.B* **664**, 315 (2008)], [arXiv:hep-ph/0508164](#)
27. V.I. Kuksa, Factorization method in the model of unstable particles with a smeared mass. *Phys. Atom. Nucl.* **72**, 1063–1073 (2009). [arXiv:0902.4892](#) [hep-ph]
28. S.D. Thomas, J.D. Wells, Phenomenology of Massive Vectorlike Doublet Leptons. *Phys. Rev. Lett.* **81**, 34–37 (1998). [arXiv:hep-ph/9804359](#)
29. W. Porod, F. Staub, SPheno 3.1: Extensions including flavour, CP-phases and models beyond the MSSM. *Comput. Phys. Commun.* **183**, 2458–2469 (2012). [arXiv:1104.1573](#) [hep-ph]
30. D. M. Straub, flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond (2018). [arXiv:1810.08132](#) [hep-ph]
31. A. Freitas, J. Lykken, S. Kell, S. Westhoff, Testing the Muon g-2 Anomaly at the LHC. *JHEP* **05**, 145 (2014) [Erratum: *JHEP* **09**, 155 (2014)], [arXiv:1402.7065](#) [hep-ph]
32. R. Dermisek, K. Hermanek, N. McGinnis, Muon g - 2 in two Higgs doublet models with vectorlike leptons (2021). [arXiv:2103.05645](#) [hep-ph]
33. K. Kannike, M. Raidal, D. M. Straub, A. Strumia, Anthropic solution to the magnetic muon anomaly: the charged see-saw. *JHEP* **02**, 106 (2012). [Erratum: *JHEP* **10**, 136 (2012)], [arXiv:1111.2551](#) [hep-ph]
34. F. Jegerlehner, A. Nyffeler, The Muon g-2. *Phys. Rept.* **477**, 1–110 (2009). [arXiv:0902.3360](#) [hep-ph]
35. R. Dermisek, A. Raval, Explanation of the Muon g-2 anomaly with vectorlike leptons and its implications for higgs decays. *Phys. Rev. D* **88**, 013017 (2013). [arXiv:1305.3522](#) [hep-ph]
36. A. Crivellin, M. Hoferichter, P. Schmidt-Wellenburg, Combined explanations of $(g-2)_\mu$, e and implications for a large muon EDM. *Phys. Rev. D* **98**, 113002 (2018). [arXiv:1807.11484](#) [hep-ph]
37. C. Bonilla, A. E. Cárcamo Hernández, J. Gonçalves, F. F. Freitas, A. P. Morais, R. Pasechnik, Collider signatures of vector-like fermions from a flavor symmetric 2HDM (2021). [arXiv:2107.14165](#) [hep-ph]
38. F. Staub, SARAH 4: A tool for (not only SUSY) model builders. *Comput. Phys. Commun.* **185**, 1773–1790 (2014). [arXiv:1309.7223](#) [hep-ph]
39. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07**, 079 (2014). [arXiv:1405.0301](#) [hep-ph]
40. T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, P. Z. Skands, An introduction to PYTHIA 8.2, *Comput. Phys. Commun.* **191**, 159–177 (2015). [arXiv:1410.3012](#) [hep-ph]
41. R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo (NNPDF), Parton distributions with QED corrections. *Nucl. Phys. B* **877**, 290–320 (2013). [arXiv:1308.0598](#) [hep-ph]

42. J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3), DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP* **02**, 057 (2014). [arXiv:1307.6346](https://arxiv.org/abs/1307.6346) [hep-ex]
43. R. Brun, F. Rademakers, ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A* **389**, 81–86 (1997)
44. François Chollet et al., Keras, <https://keras.io> (2015)
45. A. Elwood, D. Krücker, Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders (2018). [arXiv:1806.00322](https://arxiv.org/abs/1806.00322) [hep-ex]
46. G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71**, 1554 (2011) [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)], [arXiv:1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an]
47. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, arXiv e-prints (2011). [arXiv:1106.1813](https://arxiv.org/abs/1106.1813) [cs.AI]
48. T. K. Charles et al. (CLICdp, CLIC), The Compact Linear Collider (CLIC)—2018 Summary Report, 2/2018 (2018). <https://doi.org/10.23731/CYRM-2018-002>. [arXiv:1812.06018](https://arxiv.org/abs/1812.06018) [physics.acc-ph]
49. M J Boland et al. (CLIC, CLICdp), Updated baseline for a staged Compact Linear Collider (2016). <https://doi.org/10.5170/CERN-2016-004>. [arXiv:1608.07537](https://arxiv.org/abs/1608.07537) [physics.acc-ph]
50. The International Linear Collider Technical Design Report - Volume 2: Physics, (2013). [arXiv:1306.6352](https://arxiv.org/abs/1306.6352) [hep-ph]
51. V. Shiltsev, Muon colliders -opening new horizons for particle physics. https://indico.cern.ch/event/1022802/contributions/4293870/attachments/2218112/3755618/MuonCollider_UChi_03292021_v1.pdf (2021)
52. Muon collider forum kickoff meeting, <https://indico.fnal.gov/event/47038/> (2021)
53. V. Shtabovenko, R. Mertig, F. Orellana, FeynCalc 9.3: New features and improvements. *Comput. Phys. Commun.* **256**, 107478 (2020). [arXiv:2001.04407](https://arxiv.org/abs/2001.04407) [hep-ph]
54. Perspectives on the determination of systematic uncertainties at hl-lhc, <https://cds.cern.ch/record/2642427/files/ATL-PHYS-SLIDE-2018-893.pdf> (2018)
55. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC, (2022), [arXiv:2212.09379](https://arxiv.org/abs/2212.09379) [hep-ex]
56. V.M. Budnev, I.F. Ginzburg, G.V. Meledin, V.G. Serbo, The Two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation. *Phys. Rept.* **15**, 181–281 (1975)
57. S. Frixione, M.L. Mangano, P. Nason, G. Ridolfi, Improving the Weizsacker-Williams approximation in electron - proton collisions. *Phys. Lett. B* **319**, 339–345 (1993). [arXiv:hep-ph/9310350](https://arxiv.org/abs/hep-ph/9310350)