

School of Science
Department of Physics and Astronomy
Master Degree in Physics

A Machine Learning approach for Heavy Neutral
Leptons search from D_s meson decays
in the CMS experiment

Supervisor:
Prof.ssa Alessandra Fanfani

Submitted by:
Marco Cruciani

Co-supervisors:
Dr. Tommaso Diotalevi
Dr. Leonardo Lunerti

Abstract

This thesis presents an application of Machine Learning (ML) techniques to search for Heavy Neutral Leptons (HNL) in D_s meson decays in the CMS experiment. The specific decay under study is $D_s^+ \rightarrow N(\rightarrow \mu^\pm \pi^\mp) \mu^+$, where N is the HNL and final state muons can have the same charge, allowing for lepton number violation. The signal signature comes from the displaced $N \rightarrow \mu\pi$ vertex while the background comes from accidental combination of muons and a track into a common vertex.

This work is carried out in the context of an ongoing analysis, that relies on a cut-based event selection, and its aim is to explore different ML approaches to improve the event selection. The training of the ML models relies on Monte Carlo generated samples for background and different signal mass hypotheses. These samples reproduce the data-taking condition of the CMS experiment during 2018 Run 2 with $\sqrt{s} = 13$ TeV. Several ML models have been trained with three different algorithms: boosted decision trees (BDT), gradient boosted decision trees (XGB) and artificial neural networks (ANN) and their performances have been evaluated. The best performance is obtained by the XGB algorithm.

The improvement in the event selection achieved with this ML approach translates into higher significances, in a range from 14 to 20%, over the cut-based approach. The results of this work show that the developed ML-based strategy is an effective contribution to the ongoing CMS analysis.

Contents

Introduction	7
1 Standard Model neutrinos and Heavy Neutral Leptons	9
1.1 Leptons of the Standard Model	9
1.1.1 Electroweak theory	10
1.1.2 Neutrinos	13
1.2 Majorana neutrinos	15
1.2.1 Majorana Lagrangian	16
1.2.2 Sterile neutrinos	17
1.3 Heavy Neutral Leptons in hadronic machines	19
1.3.1 Production from hadron decays	19
1.3.2 Other production channels	21
1.3.3 Sterile neutrino decays	22
1.4 Search for Heavy Neutral Leptons	23
1.4.1 Peak searches	23
1.4.2 Beam dump experiments	23
1.4.3 Electroweak based searches	25
1.4.4 Collider direct searches	25
1.5 Theoretical inputs for the HNL search from D_s decays	25
2 The Large Hadron Collider and CMS experiment	27
2.1 The Large Hadron Collider at CERN	27
2.2 The CMS detector	28
2.2.1 The tracking system	30
2.2.2 The electromagnetic calorimeter	31
2.2.3 The hadron calorimeter	32
2.2.4 The Muon system	33
2.3 The CMS trigger system and B-parking dataset	34
2.3.1 The CMS trigger system	34
2.3.2 The Data parking strategy	36
2.4 CMS Event reconstruction	38
2.4.1 Particle-flow elements	38
2.4.2 Particle-flow reconstruction	40
2.5 Data and Monte Carlo processing workflow	40
3 Search for Heavy Neutral Leptons from D_s meson decays	43
3.1 Analysis strategy	43
3.1.1 Signal yield	44
3.2 Data and Monte Carlo samples	45
3.2.1 B-parking dataset	45
3.2.2 Monte-Carlo simulated sample	45
3.3 Event Selection	46
3.3.1 Pre-selection	46
3.3.2 Correction factors	47
3.3.3 Signal selection	47
3.4 Limits extraction	50
3.4.1 Lifetime reweighting	50
3.4.2 Normalization channel yield	50

CONTENTS

3.4.3	Signal and background fit	50
3.4.4	Systematic uncertainties	51
3.4.5	Expected limit extraction procedure	53
4	Machine Learning algorithms for the search for Heavy Neutral Leptons	55
4.1	Brief introduction to Machine Learning	55
4.1.1	Boosted Decision Trees	56
4.1.2	Artificial Neural Networks	57
4.1.3	Evaluating the performance of the models	58
4.2	Signal topology and variable selection	59
4.2.1	Discriminating variables	60
4.2.2	Categorization of events	64
4.3	Training the models	65
4.3.1	Data preparation	65
4.3.2	Configuration of the ML models	65
4.3.3	Monitoring training progress	66
4.4	Results	69
4.4.1	Selecting the best algorithm	69
4.4.2	Most discriminating variables	73
4.4.3	Model scores and cut optimization	76
4.4.4	Results and comparison with the cut-based analysis	80
4.5	Summary and outlook	82
	Conclusions	85
	Bibliography	87
	A Monte Carlo	93

Introduction

The Standard Model's assumptions on massless neutrinos have been challenged by the observation of neutrino oscillations [1, 2], that show that active neutrinos are massive. The question of how active neutrinos obtain their mass is still open [3].

One of the extensions of the Standard Model that tries to explain neutrino masses consists in adding right-handed neutrino fields to the Lagrangian [4]. These are also called sterile neutrinos or Heavy Neutral Leptons (HNLs), because they are singlets under Standard Model gauge symmetries. The only possible way to observe such neutrinos is through their mixing with the active neutrinos.

A wide range of different experiments have looked for signature of HNL with masses that range from eV to TeV energy scales. No conclusive observation has been made yet, but each experiment has set upper limits on the mixing parameter with active neutrinos [5].

This thesis work focuses on the search for Heavy Neutral Leptons from decays of the D_s meson in the CMS experiment. The decay channel we target is $D_s^+ \rightarrow N (\rightarrow \mu^\pm \pi^\mp) \mu^+$, where N is a HNL. The final state muons can have same sign, allowing for lepton number violation. Among all charmed meson decays into Heavy Neutral Leptons, this has the largest branching ratio, and it allows probing the existence of HNLs with a mass below the D_s mass ($\lesssim 2$ GeV). Currently, there are no published results on HNL searches from charmed meson decays at LHC. This represents a clean experimental signature in the CMS experiment for HNL decaying within the tracker system (< 1 m). The whole decay chain is fully reconstructed, and it can profit from excellent track and muon reconstruction performances.

The selected final state consists of one muon emerging from the D_s meson decay vertex, and a reconstructed displaced vertex formed by the N decay products (muon and pion). The main background comes from accidental combinations of muons and a track into a common vertex. This analysis is possible thanks to a large dataset that was collected by triggering on single low- p_T non-prompt muons: the B-Parking dataset [6]. The B-Parking dataset was recorded at the CMS experiment in proton-proton collisions with $\sqrt{s} = 13$ TeV during the 2018 data-taking period.

This work has been carried out in the context of an ongoing analysis, where the event selection is based on a cut-based approach [7]. The optimization process of this event selection relies on Monte Carlo samples.

The purpose of this thesis work is to explore the usage of Machine Learning (ML) techniques in the event selection. This includes the implementation of a common framework to be able to make meaningful comparisons between different ML algorithms. The training of the ML models uses the same MC datasets used in the cut-based approach, and the signal datasets include different HNL mass hypotheses. These models are trained to discriminate the different HNL signatures against the background. After the training process, we assess the performance of the different ML algorithms and compare the one providing the best performance with the cut-based approach. This allows evaluating the impact of the implementation of these techniques in the analysis.

The thesis is organized as follows:

in chapter 1 we introduce the theoretical framework of the Standard Model and the Heavy Neutral Leptons. In chapter 2 we describe the CMS detector, CMS event reconstruction and the B-Parking dataset. In chapter 3 we present the analysis strategy and its cut-based selection. In chapter 4 we present our new selection implementation based on ML models and their results.

Chapter 1

Standard Model neutrinos and Heavy Neutral Leptons

Sterile neutrinos or Heavy neutral leptons (HNLs) are a class of neutrinos that do not interact via any of the Standard Model (SM) forces. They are part of many extensions of the SM because they have the potential to explain some fundamental issues that cannot be dealt with in the SM such as neutrino masses, the baryon asymmetry of the universe and the nature of dark matter [4, 8].

Several experiments have attempted to search for HNLs, however, as of now, there is no accepted evidence for their existence. These experiments have set exclusion limits on the parameter space of HNL masses and mixing angles [5].

This chapter is structured as follows: in section 1.1 we begin by providing an overview of the properties of Dirac-type fermions and neutrinos of the SM; then in section 1.2 we introduce the concept of Majorana neutrinos and a minimal extension of the SM that adds right-handed neutrino fields to the Lagrangian; in section 1.3 we overview the main channels of interest for HNL production with a particular focus on meson decays; section 1.4 analyzes the main exclusion regions in the parameter space of HNL masses and mixing angles, and the kinds of experiments that can probe them. Finally, section 1.5 contains the key theoretical parameters and formulas that are required for the computations of HNL production from D_s decays and decay rates, which are key to the analysis presented in this thesis.

1.1 Leptons of the Standard Model

Fermions are spin 1/2 particles that can be described through a Dirac spinor field ψ . The Lagrangian describing a free fermion is known as the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi , \quad (1.1)$$

and the corresponding equation of motion is the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 . \quad (1.2)$$

γ^μ are the Dirac matrices, m is the mass of the fermion and $\bar{\psi}$ is defined as

$$\bar{\psi} = \psi^\dagger \gamma^0 , \quad (1.3)$$

where ψ^\dagger is the Hermitian conjugate of ψ . Explicit forms of γ^μ depend on a choice of basis. In the Weyl basis or chiral basis the Dirac matrices are

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix} , \quad \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} , \quad (1.4)$$

where $\sigma^\mu = (\mathbb{1}, \sigma^i)$ and $\bar{\sigma}^\mu = (\mathbb{1}, -\sigma^i)$, with σ^i the Pauli matrices. In this basis the spinor can be written as:

$$\psi = \begin{pmatrix} \chi_L \\ \chi_R \end{pmatrix} , \quad (1.5)$$

where χ_L and χ_R are respectively a left-handed and a right-handed Weyl spinor [9].

We also use the notation

$$\psi_L = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix}, \quad \psi_R = \begin{pmatrix} 0 \\ \chi_R \end{pmatrix}, \quad (1.6)$$

that lets us write

$$\psi = \psi_L + \psi_R. \quad (1.7)$$

Projector operators can be defined to extract the left-handed and right-handed components of the spinor

$$P_L = \frac{1}{2}(\mathbb{1} - \gamma^5), \quad P_R = \frac{1}{2}(\mathbb{1} + \gamma^5), \quad (1.8)$$

so that

$$P_L \psi = \psi_L, \quad P_R \psi = \psi_R. \quad (1.9)$$

ψ_L and ψ_R are eigenstates of the chirality operator γ^5

$$\gamma^5 \psi_L = -\psi_L, \quad \gamma^5 \psi_R = \psi_R. \quad (1.10)$$

The handedness of a spinor is also referred to as chirality.

The mass term of the Dirac Lagrangian

$$\mathcal{L}_{\text{mass}} = -m\bar{\psi}\psi = m(\bar{\chi}_L \chi_R + \bar{\chi}_R \chi_L), \quad (1.11)$$

couples the left and right-handed components of the spinor. We can also see this more explicitly by looking at the Dirac equation

$$\begin{aligned} i\sigma^\mu \partial_\mu \chi_R - m\chi_L &= 0 \\ i\bar{\sigma}^\mu \partial_\mu \chi_L - m\chi_R &= 0, \end{aligned} \quad (1.12)$$

or equivalently

$$\begin{aligned} i\gamma^\mu \partial_\mu \psi_R - m\psi_L &= 0 \\ i\gamma^\mu \partial_\mu \psi_L - m\psi_R &= 0, \end{aligned} \quad (1.13)$$

which implies that even if we start with a purely left or right-handed spinor, the two components will mix with time evolution. Only a massless fermion can be purely left-handed or right-handed without mixing.

1.1.1 Electroweak theory

The electroweak theory is an $SU(2)_L \otimes U(1)_Y$ gauge theory. We will only focus the leptonic sector. We have the left-handed weak-isospin doublets:

$$L_\alpha = \begin{pmatrix} \nu_\alpha \\ \ell_\alpha \end{pmatrix}_L = \left\{ \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \right\}, \quad (1.14)$$

and the right-handed weak-isospin singlets:

$$R_\alpha = \ell_{\alpha R} = \{e_R, \mu_R, \tau_R\}. \quad (1.15)$$

Neutrinos are assumed to be massless and have no right-handed component¹. Weak hypercharge Y is defined according to the Gell-Mann-Nishijima relation:

$$Q = I_3 + \frac{1}{2}Y, \quad (1.16)$$

where Q is the usual electric charge and I_3 is the third component of weak isospin, which is $+1/2$ for the upper component of the doublet, $-1/2$ for the lower component and 0 for the right-handed singlets.

¹We talk about the reasoning behind this assumption in section 1.1.2.

To keep the notation less cluttered we will avoid the index α from now on, and it should be understood that the following Lagrangian terms are summed over the three lepton generations. The Lagrangian terms that describes leptons and their interactions is [10]:

$$\mathcal{L}_{\text{leptons}} = \bar{R}i\gamma^\mu \left(\partial_\mu + \frac{ig'}{2} a_\mu Y \right) R + \bar{L}i\gamma^\mu \left(\partial_\mu + \frac{ig'}{2} a_\mu Y + \frac{ig}{2} \vec{\sigma} \cdot \vec{b}_\mu \right) L , \quad (1.17)$$

with a_μ being the massless boson field for $U(1)_Y$, $\vec{b}_\mu = (b_\mu^1, b_\mu^2, b_\mu^3)$ the massless boson fields for $SU(2)_L$ and $\vec{\sigma} = \{\sigma^1, \sigma^2, \sigma^3\}$ the Pauli matrices.

A mass term for the charged leptons would look like:

$$\mathcal{L}_{\text{mass}} = -m_\ell (\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L) , \quad (1.18)$$

but it cannot be added directly to the Lagrangian because it mixes left-handed and right-handed components and is thus not gauge invariant. The SM way of introducing mass terms for the charged leptons is through their Yukawa coupling to the Higgs field.

Higgs and spontaneous symmetry breaking

To give mass to the gauge bosons and to the massive fermions in the SM, we need to introduce the Higgs field². The Higgs field is a complex scalar field that transforms as a weak-isospin doublet under $SU(2)_L$ and has weak hypercharge $Y = 1$

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} . \quad (1.19)$$

The related Lagrangian term is

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi) , \quad (1.20)$$

with D_μ the covariant derivative and V the Higgs potential:

$$D_\mu = \left(\partial_\mu + \frac{ig'}{2} a_\mu Y + \frac{ig}{2} \vec{\sigma} \cdot \vec{b}_\mu \right) , \quad V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 . \quad (1.21)$$

We also add the so-called Yukawa Lagrangian term that couples the Higgs field to the leptons

$$\mathcal{L}_{\text{Yukawa}} = -y_\ell (\bar{L} \phi R + \bar{R} \phi^\dagger L) , \quad (1.22)$$

where y_ℓ are the Yukawa couplings.

What we are interested in is the fact that the Higgs field acquires a vacuum expectation value

$$\langle \phi \rangle_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} , \quad (1.23)$$

and when we consider the physics around this vacuum state, we have a spontaneous symmetry breaking (SSB) of the $SU(2)_L \otimes U(1)_Y$ symmetry, but we preserve the $U(1)_{\text{EM}}$ symmetry. The bosons acquire mass:

$$\begin{array}{ccc} \text{massless } a_\mu, \vec{b}_\mu & & \text{massless } A_\mu, \text{ massive } W_\mu^\pm, Z_\mu \\ \xrightarrow{\text{after SSB}} & & \\ \text{massless } \phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} & & \text{massive } H^0 \end{array}$$

Lepton interactions

It can be shown that after SSB the Lagrangian terms for the leptons (eq. (1.17)) becomes:

$$\mathcal{L}_{\text{leptons}} = \mathcal{L}_W + \mathcal{L}_0 , \quad (1.24)$$

²See refs. [9, 10] for a more complete discussion on the Higgs mechanism.

where \mathcal{L}_W is the Lagrangian term for the charged interactions mediated by the W boson

$$\begin{aligned}\mathcal{L}_W &= -\frac{g}{\sqrt{2}} [\bar{\nu}_L \gamma^\mu \ell_L W_\mu^+ + \bar{\ell}_L \gamma^\mu \nu_L W_\mu^-] \\ &= -\frac{g}{2\sqrt{2}} [\bar{\nu} \gamma^\mu (1 - \gamma^5) \ell W_\mu^+ + \bar{\ell} \gamma^\mu (1 - \gamma^5) \nu W_\mu^-]\end{aligned}\quad (1.25)$$

and \mathcal{L}_0 is the Lagrangian term for the neutral interactions mediated by the neutral bosons Z and A , usually expressed in terms of the Weinberg angle $\theta_W = \arctan(g'/g)$ and the electromagnetic coupling constant $e = g \sin \theta_W = g' \cos \theta_W$:

$$\begin{aligned}\mathcal{L}_0 &= e \bar{\ell} \gamma^\mu \ell A_\mu - \frac{g}{2 \cos \theta_W} \nu_L \gamma^\mu \nu_L Z_\mu \\ &\quad - \frac{g}{2 \cos \theta_W} [2 \sin^2 \theta_W \bar{\ell}_R \gamma^\mu \ell_R + (2 \sin^2 \theta_W - 1) \bar{\ell}_L \gamma^\mu \ell_L] Z_\mu \\ &= e \bar{\ell} \gamma^\mu \ell A_\mu - \frac{g}{4 \cos \theta_W} \bar{\ell} \gamma^\mu (1 - \gamma^5) \ell Z_\mu \\ &\quad - \frac{g}{4 \cos \theta_W} [2 \sin^2 \theta_W \bar{\ell} \gamma^\mu (1 + \gamma^5) \ell + (2 \sin^2 \theta_W - 1) \bar{\ell} \gamma^\mu (1 - \gamma^5) \ell] Z_\mu\end{aligned}\quad (1.26)$$

These kinds of interactions can be described by the Feynman diagrams in fig. 1.1.

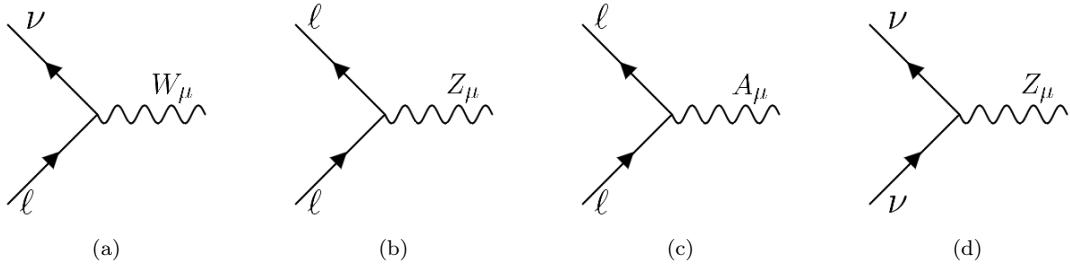


Figure 1.1: Feynman diagrams for the interactions of leptons. (a) is the charged current interaction mediated by the W boson. (b), (c), (d) are the neutral current interactions mediated by the Z boson.

Charged lepton mass

It can be shown that after SSB the Yukawa Lagrangian term for the leptons (eq. (1.22)) becomes:

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} &= -y_\ell \frac{v + H}{\sqrt{2}} (\bar{\ell}_R \ell_L + \bar{\ell}_L \ell_R) \\ &= -\frac{y_\ell}{\sqrt{2}} H \bar{\ell} \ell - \frac{y_\ell v}{\sqrt{2}} \bar{\ell} \ell\end{aligned}\quad (1.27)$$

The first term gives rise to the interaction between the Higgs boson and the leptons. The second term is the mass term for the charged leptons

$$m_\ell = \frac{y_\ell v}{\sqrt{2}}. \quad (1.28)$$

Lepton number

Lepton number is a quantum number that is conserved in the SM. It is actually an accidental symmetry of the SM, since it is not imposed by the gauge symmetry.

This number comes from the fact that the Lagrangian is invariant under the transformation

$$\begin{aligned}\ell_L^\alpha &\rightarrow e^{i\varphi_\alpha} \ell_L^\alpha & \ell_R^\alpha &\rightarrow e^{i\varphi_\alpha} \ell_R^\alpha & \nu_L^\alpha &\rightarrow e^{i\varphi_\alpha} \nu_L^\alpha \\ \bar{\ell}_L^\alpha &\rightarrow e^{-i\varphi_\alpha} \bar{\ell}_L^\alpha & \bar{\ell}_R^\alpha &\rightarrow e^{-i\varphi_\alpha} \bar{\ell}_R^\alpha & \bar{\nu}_L^\alpha &\rightarrow e^{-i\varphi_\alpha} \bar{\nu}_L^\alpha\end{aligned}\quad (1.29)$$

This holds for each generation $\alpha = e, \mu, \tau$, with φ_α arbitrary phases. Because of Noether's theorem, this symmetry implies the conservation of the lepton number $L_\alpha = \{L_e, L_\mu, L_\tau\}$ and consequently, also of the total lepton number $L = L_e + L_\mu + L_\tau$. We assign the lepton number to be $+1$ for leptons and -1 for antileptons.

1.1.2 Neutrinos

Neutrinos are neutral leptons that only interact through the weak force. They are quite unique in the SM because they are the only fermions that are assumed to be massless and to have no right-handed component. With the evidence of neutrino oscillations, we know that these assumptions are wrong, and we need to find a way to include massive neutrinos in the SM.

Helicity and chirality

Helicity of a particle can be defined as the normalized component of its spin along its direction of motion:

$$h = \frac{\vec{S} \cdot \vec{p}}{|\vec{S}| |\vec{p}|}. \quad (1.30)$$

For a spin 1/2 particle we can write the spin operator as

$$\vec{S} = \frac{1}{2} \vec{\Sigma}, \quad (1.31)$$

where $\vec{\Sigma}$ is simply a four-component version of the Pauli matrices:

$$\vec{\Sigma} = \begin{pmatrix} \vec{\sigma} & 0 \\ 0 & \vec{\sigma} \end{pmatrix}. \quad (1.32)$$

Thus,

$$h = \frac{\vec{\Sigma} \cdot \vec{p}}{|\vec{p}|}. \quad (1.33)$$

If we look at the Dirac equation (eq. (1.13)) in Fourier space for a chiral spinor $\psi = \begin{pmatrix} \chi_L \\ \chi_R \end{pmatrix}$ we get:

$$\begin{aligned} i\sigma^\mu \partial_\mu \chi_R - m\chi_L &= 0 \\ i\bar{\sigma}^\mu \partial_\mu \chi_L - m\chi_R &= 0 \end{aligned} \implies \begin{aligned} (E - \vec{\sigma} \cdot \vec{p}) \chi_R &= m\chi_L \\ (E + \vec{\sigma} \cdot \vec{p}) \chi_L &= m\chi_R \end{aligned}, \quad (1.34)$$

which in the massless case becomes

$$\begin{aligned} (\vec{\sigma} \cdot \vec{p}) \chi_R &= E\chi_R \\ (\vec{\sigma} \cdot \vec{p}) \chi_L &= -E\chi_L \end{aligned}. \quad (1.35)$$

If we want to express this same equations for the four-component Dirac spinors $\psi_L = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix}$ and $\psi_R = \begin{pmatrix} 0 \\ \chi_R \end{pmatrix}$ we can use the $\vec{\Sigma}$ matrix and write

$$\frac{\vec{\Sigma} \cdot \vec{p}}{E} \psi_L = -\psi_L, \quad \frac{\vec{\Sigma} \cdot \vec{p}}{E} \psi_R = \psi_R, \quad (1.36)$$

which is exactly the same as eq. (1.10): chirality and helicity are only the same for massless particles [11]. In the massive case helicity is not Lorentz invariant: it is always possible to boost past the particle and 'change' its direction of motion.

In the case of neutrinos helicity and chirality are approximately equivalent because in all practical cases neutrinos are ultrarelativistic and thus $|\vec{p}| \gg m$. According to experimental evidence, the helicity of neutrinos is always negative i.e. they are always left-handed. The first experimental measurement of the helicity of neutrinos was performed by Goldhaber et al. in 1958 [12].

Neutrinos assumptions in the standard model

In the SM, neutrinos are massless. This is because experimental evidence at the time of the formulation of the SM suggested that neutrinos mass was very small and compatible with zero. Also, it is assumed that they have no right-handed component. This is because there is no experimental evidence of neutrinos with right-handed helicity; also, such a particle would not be charged under the weak force and be practically impossible to detect. So a massless left-handed neutrino was the simplest solution that made the least amount of extra assumptions and could explain all experimental evidence.

Neutrino oscillations

However, in the last decades experimental evidence has emerged that indicates that neutrinos are not massless: neutrino oscillations. The key ideas behind neutrino oscillations are [11]:

- neutrinos are produced in weak interactions as flavor eigenstates

$$\nu_\alpha = \{\nu_e, \nu_\mu, \nu_\tau\} , \quad (1.37)$$

but these are not the same as the mass eigenstates ν_i ;

- for instance we produce a neutrino state ν_β which can be written as a linear combination of the mass eigenstates:

$$|\nu_\beta\rangle = \sum_i k_i |\nu_i\rangle , \quad (1.38)$$

- since they have a different mass, the time evolution of the mass eigenstates is different:

$$|\nu_i(t)\rangle = \exp[-i(E_i t - \vec{p}_i \cdot \vec{x}_i)] |\nu_i(0)\rangle , \quad (1.39)$$

where E_i is the energy of the mass eigenstate ν_i , \vec{p}_i is its momentum and \vec{x}_i is the travelled distance;

- because of this, if we observe the neutrino after a time t and a distance L from the production point, there is a probability of observing a flavor eigenstate $\nu_\alpha \neq \nu_\beta$

This behavior has been observed in several experiments. The first two experiments that won the Nobel Prize for this observation are the Super-Kamiokande [1] experiment and the Subdury Neutrino Observatory experiment [2].

If we call the mass eigenvalues m_1, m_2 and m_3 , the current experimental measurements have been able to measure the mass differences [13]:

$$\begin{aligned} \Delta_{21}m^2 &= m_2^2 - m_1^2 \equiv \Delta m_{sol}^2 \approx 7.5 \times 10^{-5} \text{eV}^2 \\ \Delta_{31}m^2 &= m_3^2 - m_1^2 \equiv \Delta m_{atm}^2 \approx 2.5 \times 10^{-3} \text{eV}^2 , \end{aligned} \quad (1.40)$$

where the naming solar (sol) and atmospheric (atm) comes from the fact that the first one can be measured in solar neutrino experiments and the second one in atmospheric neutrino experiments.

Dirac neutrinos

The simplest way to introduce massive neutrinos in the SM is to assume they are Dirac fermions like all other fundamental fermions of the SM, and add right-handed neutrino fields to the Lagrangian. Then we can write a Dirac mass term for neutrinos through the Higgs mechanism just like we did for the charged leptons (eq. (1.27)). We use Greek letters (e.g. ν_α) to denote flavor eigenstates, Latin letters (e.g. ν_i) to denote mass eigenstates. Summation over repeated indices is implied.

In the ‘flavor basis’ the charged current interaction Lagrangian is

$$\mathcal{L}_W = -\frac{g}{\sqrt{2}} \ell_L^\alpha \gamma^\mu \nu_L^\alpha W_\mu^- + \text{h.c.} , \quad (1.41)$$

but the mass term will be:

$$\mathcal{L}_m^\nu = -\frac{1}{2} \bar{\nu}_L^\alpha M_{\alpha\beta} \nu_R^\beta + \text{h.c.} , \quad (1.42)$$

and the mass matrix $M_{\alpha\beta}$ is not diagonal. This spoils the accidental symmetry of lepton number that we had in the massless case (see section 1.1.1) because now it can only hold for the total lepton number $L = L_e + L_\mu + L_\tau$, with a transformation using a single phase for all fields:

$$\begin{aligned} \ell_L^\alpha &\rightarrow e^{i\varphi} \ell_L^\alpha & \ell_R^\alpha &\rightarrow e^{i\varphi} \ell_R^\alpha & \nu_L^\alpha &\rightarrow e^{i\varphi} \nu_L^\alpha & \nu_R^\alpha &\rightarrow e^{i\varphi} \nu_R^\alpha & \text{for all } \alpha = e, \mu, \tau . \\ \bar{\ell}_L^\alpha &\rightarrow e^{-i\varphi} \bar{\ell}_L^\alpha & \bar{\ell}_R^\alpha &\rightarrow e^{-i\varphi} \bar{\ell}_R^\alpha & \bar{\nu}_L^\alpha &\rightarrow e^{-i\varphi} \bar{\nu}_L^\alpha & \bar{\nu}_R^\alpha &\rightarrow e^{-i\varphi} \bar{\nu}_R^\alpha \end{aligned} \quad (1.43)$$

The flavor and mass eigenstates are related by a unitary matrix V known as the *Pontecorvo-Maki-Nakagawa-Sakata* (PMNS) matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = V \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} V_{e1} & V_{e2} & V_{e3} \\ V_{\mu 1} & V_{\mu 2} & V_{\mu 3} \\ V_{\tau 1} & V_{\tau 2} & V_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (1.44)$$

Thus, in the ‘mass basis’ the charged current interaction Lagrangian is

$$\mathcal{L}_W = -\frac{g}{\sqrt{2}} \ell_L^\alpha \gamma^\mu V_{\alpha k} \nu_L^k W_\mu^- + \text{h.c.}, \quad (1.45)$$

and the mass term will be:

$$\mathcal{L}_m^\nu = -\frac{1}{2} \bar{\nu}_L^k m_k \nu_R^k + \text{h.c.}, \quad (1.46)$$

where m_k are the masses of the mass eigenstates ν_k . The mass terms will come from the Yukawa couplings of the Higgs just like for the other fermions $m_k = y_k^\nu v / \sqrt{2}$. The very small masses of neutrinos imply that the Yukawa couplings y_k^ν are extremely small. The SM does not provide any prediction for the Yukawa couplings of fermions, but this result is thought to be ‘unnatural’ and a hint to new physics beyond the SM [11].

1.2 Majorana neutrinos

A possible alternative to Dirac neutrinos is the Majorana neutrino.

First let us introduce the charge-conjugate field ψ^c of a spinor field. If we have a charged spinor field ψ that follows the Dirac equation

$$(i\gamma^\mu \partial_\mu - q\gamma^\mu A_\mu - m)\psi = 0, \quad (1.47)$$

we want the charge conjugate field to follow the same equation but with opposite charge:

$$(i\gamma^\mu \partial_\mu + q\gamma^\mu A_\mu - m)\psi^c = 0. \quad (1.48)$$

It can be shown [11] that this requirement is fulfilled by the spinor field:

$$\psi^c = \mathcal{C} \bar{\psi}^T, \quad (1.49)$$

with \mathcal{C} being a unitary matrix that must satisfy

$$C\gamma^\mu C^{-1} = -\gamma^{\mu T}. \quad (1.50)$$

We know that Weyl spinors are massless and have two components. It is reasonable to ask if there is a way to have a two-component spinor which is also massive. A spinor with this property is called a Majorana spinor. Let us look again at the Dirac equation for a fermion $\psi = \psi_L + \psi_R$ (eq. (1.13)):

$$\begin{aligned} i\gamma^\mu \partial_\mu \psi_R &= m\psi_L \\ i\gamma^\mu \partial_\mu \psi_L &= m\psi_R. \end{aligned} \quad (1.51)$$

The two equations are the same if

$$\psi_R = \mathcal{C} \bar{\psi}_L^T = \psi_L^c. \quad (1.52)$$

Equation (1.52) is called the Majorana condition. It’s easy to see that since we can write

$$\psi = \psi_L + \mathcal{C} \bar{\psi}_L^T, \quad (1.53)$$

ψ is the same as its charge conjugate ψ^c :

$$\psi^c = \mathcal{C} \bar{\psi}_L^T + \psi_L = \psi, \quad (1.54)$$

which is the defining property of Majorana fermions: they are their own antiparticles.

It is also interesting to note that this condition only makes sense for completely neutral particles. For instance, eqs. (1.55) and (1.56):

$$(i\gamma^\mu \partial_\mu - q\gamma^\mu A_\mu - m)\psi = 0, \quad (1.55)$$

$$(i\gamma^\mu \partial_\mu + q\gamma^\mu A_\mu - m)\psi = 0, \quad (1.56)$$

can both hold for the same field only if $q = 0$.

So neutrinos are unique because they are the only known fundamental fermions that could be Majorana fermions instead of Dirac.

1.2.1 Majorana Lagrangian

Consider a massive neutrino $\nu = \nu_L + \nu_R$. It should follow the Dirac Lagrangian and have a mass term:

$$\mathcal{L}_{\text{mass}} = -m\bar{\nu}\nu = -m(\bar{\nu}_L\nu_R + \bar{\nu}_R\nu_L) , \quad (1.57)$$

If we impose the Majorana condition (eq. (1.52)) we get

$$\mathcal{L}_{\text{mass}} = -\frac{m}{2}(\bar{\nu}_L\nu_L^C + \bar{\nu}_L^C\nu_L) , \quad (1.58)$$

which is a mass term for a Majorana neutrino. The $1/2$ avoids double-counting because ν_L^C and $\bar{\nu}_L$ are not independent.

We can also put this in a more convenient form

$$\bar{\nu}_L^C = (\mathcal{C}\bar{\nu}_L^T)^\dagger \gamma^0 = \nu_L^T(\gamma^0)^T \mathcal{C}^\dagger \gamma^0 = -\nu_L^T \mathcal{C}^\dagger , \quad (1.59)$$

thus the full Lagrangian for a Majorana neutrino is [11]:

$$\mathcal{L}_{\text{Majorana}} = \bar{\nu}_L i\gamma^\mu \partial_\mu \nu_L - \frac{m}{2}(-\nu_L^T \mathcal{C}^\dagger \nu_L + \bar{\nu}_L C \bar{\nu}_L^T) . \quad (1.60)$$

Lepton number for Majorana neutrinos

If we test the mass term for a Majorana neutrino under a global phase transformation

$$\nu_L \rightarrow e^{i\varphi} \nu_L \quad \bar{\nu}_L \rightarrow e^{-i\varphi} \bar{\nu}_L \quad (1.61)$$

we get:

$$\mathcal{L}_{\text{mass}} = -\frac{m}{2}(-\nu_L^T C^\dagger \nu_L + \bar{\nu}_L C \bar{\nu}_L^T) \rightarrow -\frac{m}{2}(-e^{i2\varphi} \nu_L^T C^\dagger \nu_L + e^{-i2\varphi} \bar{\nu}_L C \bar{\nu}_L^T) , \quad (1.62)$$

which clearly shows that in this case there is no lepton number conservation. If we still use the conventional SM definition of lepton number, a Majorana neutrino can violate lepton number by 2 units. The most famous example of this is neutrinoless double beta decay, which is a hypothetical process that violates lepton number by 2 units and is only possible if neutrinos are Majorana particles [14].

Intuitively, if we compare the Feynman diagrams of fig. 1.2 we can see that a Majorana neutrino can be exchanged in the internal line of the neutrinoless double beta decay diagram, while a Dirac neutrino cannot.

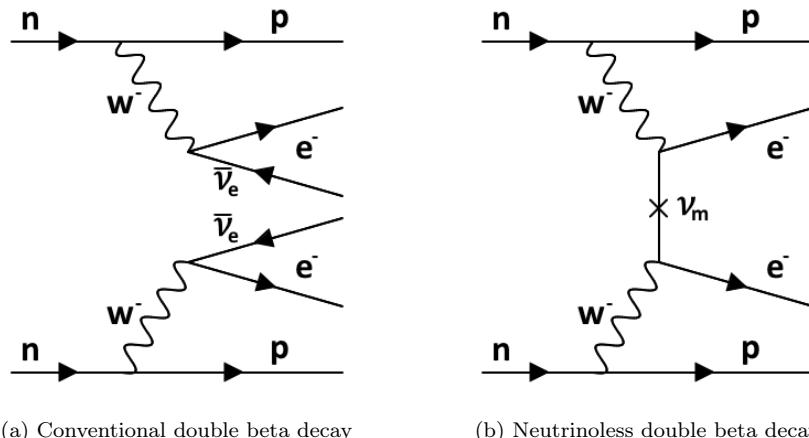


Figure 1.2: Feynman diagrams for SM double beta decay and neutrinoless double beta decay. ν_m is a Majorana neutrino.

Majorana mass for SM neutrinos

If we want to give SM neutrinos a Majorana mass term, without adding any new fields, we need to use higher-dimensional operators. Consider the Dirac mass term of charged leptons:

$$m\bar{\ell}\ell = m\bar{\ell}_L\ell_R + \text{h.c.} , \quad (1.63)$$

$\bar{\ell}_L\ell_R$ has weak-isospin $I = 1/2$ and hypercharge $Y = -1$, so we are able to introduce a term like this through its interaction with the Higgs doublet that has $I = 1/2$ and $Y = +1$ in the Yukawa term (eq. (1.22)). However, a Majorana mass for SM neutrinos would have the form:

$$\mathcal{L}_{\text{mass}} = -\frac{m}{2}\bar{\nu}_L\nu_L^c + \text{h.c.} , \quad (1.64)$$

$\bar{\nu}_L\nu_L^c$ are both left-handed, making them have a total weak-isospin $I = 1$ and hypercharge $Y = 2$: there is no Higgs-like scalar field in the SM that has these quantum numbers. If we want to build a Majorana mass term for left-handed neutrinos we need to use a higher-dimensional operator [11]:

$$\mathcal{L}_5 = \frac{g}{\mathcal{M}} (L_L^T \sigma_2 \phi) C^\dagger (\phi^T \sigma_2 L_L) + \text{h.c.} , \quad (1.65)$$

where \mathcal{M} is an energy scale required to make the total energy dimension 4. This is appropriately called \mathcal{L}_5 because it is a dimension 5 operator (recall that fields are operators in a quantum field theory, with energy dimension $\psi \sim [E]^{3/2}$ and $\phi \sim [E]$). But we know that operators with dimension larger than 4 are not renormalizable [15]. This would make the SM an effective field theory that only approximates a more fundamental theory at low energies below the scale \mathcal{M} .

An interesting thing to remark however is that a Majorana mass term is allowed for a ‘sterile’ right-handed neutrino ν_R which also has no weak hypercharge. This is because such neutrino is a singlet under the SM gauge group and thus a Majorana mass term is naturally gauge invariant.

1.2.2 Sterile neutrinos

With the term ‘sterile’ neutrinos we refer to neutrinos that are not charged under any of the SM interactions. The interest in sterile neutrinos comes from the fact that they could be the explanation for some open questions of physics that are not explained by the SM (e.g. [4, 8]). For this reason they have been extensively studied in the last decades and a wide variety of experiments and theoretical models have been developed to constrain their existence and properties.

Minimal sterile neutrino model

A minimal and generic way to introduce sterile neutrinos in the SM [4], is to modify the Lagrangian by adding \mathcal{N} right-handed neutrinos N_I with $I = 1, \dots, \mathcal{N}$:

$$\delta\mathcal{L} = \overline{N}_I i\gamma^\mu \partial_\mu N_I - Y_{\alpha I}^\nu \overline{L}_\alpha \tilde{\phi} N_I - \frac{M_I}{2} \overline{N}_I^c N_I + \text{h.c.} , \quad (1.66)$$

where ϕ is the Higgs doublet, L_α are the left-handed lepton doublets ($\alpha = e, \mu, \tau$), $Y_{\alpha I}^\nu$ are the Yukawa couplings that generate a Dirac mass $M_{\alpha I}^D = Y_{\alpha I}^\nu \langle \phi \rangle_0$, $\tilde{\phi}$ is the conjugate of the Higgs doublet $\tilde{\phi} = i\sigma_2 \phi^*$ and M_I are the Majorana masses of the right-handed neutrinos. These terms are all allowed within the SM, because the right-handed neutrinos are $SU(2)_L \otimes U(1)_Y$ singlets.

One generation extended model

To understand the implications of the minimally extended model, we go more in depth with a simplified version of it³. Our simplification is assuming that there is only one conventional left-handed active neutrino ν_L in the doublet $L = \begin{pmatrix} \nu_L \\ \ell_L \end{pmatrix}$ and one right-handed sterile neutrino N . The Lagrangian terms that we can add to account for the sterile neutrino are

$$\delta\mathcal{L} = \bar{N} i\gamma^\mu \partial_\mu N - m_D \bar{\nu}_L N - \frac{m_R}{2} \bar{N}^c N + \text{h.c.} . \quad (1.67)$$

³A full derivation can be found in ref. [11]

Now we introduce the vector of neutrino fields:

$$\chi_L = \begin{pmatrix} \nu_L \\ N^c \end{pmatrix}. \quad (1.68)$$

The previous Lagrangian terms for mass can now be written as

$$\mathcal{L}_{\text{mass}} = -\frac{1}{2}\overline{\chi_L^c} M \chi_L + \text{h.c.}, \quad (1.69)$$

where M is the mass matrix

$$M = \begin{pmatrix} 0 & m_D \\ m_D & m_R \end{pmatrix}. \quad (1.70)$$

Now we can diagonalize the mass matrix M with a unitary matrix V :

$$V^\dagger M V = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad (1.71)$$

where $m_1 < m_2$ are the masses of the two neutrino mass eigenstates. The neutrino mass eigenstates $n = \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$ will be related to the flavor eigenstates $\chi_L = \begin{pmatrix} \nu_L \\ N^c \end{pmatrix}$ through the unitary matrix V (also called mixing matrix):

$$\chi_L = V n_L \implies n_L = V^\dagger \chi_L. \quad (1.72)$$

For simplicity, we assume that V is real⁴, and thus we can write it as:

$$V = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (1.73)$$

If we assume N to be ‘heavy’, i.e. $m_R \gg m_D$ we can write the mass eigenstates as

$$n_1 \approx \nu_L - \frac{m_D}{m_R} N^c, \quad n_2 \approx N^c + \frac{m_D}{m_R} \nu_L, \quad (1.74)$$

and the mixing angle as

$$\tan \theta \approx \frac{m_D}{m_R}. \quad (1.75)$$

This means that in the end we get:

1. a light neutrino n_1 with mass m_1 which is almost the same as the active neutrino ν_L ;
2. a heavy neutrino n_2 with mass m_2 which is almost the same as the sterile neutrino N .

Thus, this simplified model that adds a heavy neutrino N lets us give a small mass to the active neutrino ν_L .

If we substitute $\nu_L \simeq n_1 + \frac{m_D}{m_R} N^c$ in the Lagrangian of interaction for leptons (eq. (1.25) and eq. (1.26)) we can see how N interacts:

$$\mathcal{L}_{\text{int}} = \frac{g}{2\sqrt{2}} \overline{N^c} \frac{m_D}{m_R} \gamma^\mu (1 - \gamma^5) \ell W_\mu^+ + \frac{g}{4 \cos \theta_W} \overline{N^c} \frac{m_D}{m_R} \gamma^\mu (1 - \gamma^5) \nu_L Z_\mu + \text{h.c.} \quad (1.76)$$

It's clear that the interactions of the heavy neutrino N are strongly suppressed because of the ratio $m_D/m_R \ll 1$.

Heavy neutral leptons interactions with the SM

If we go through analogous steps of the previous section for the full model of eq. (1.66), assuming that the eigenvalues of M_I are much larger than the elements of M_D , we reach analogous conclusions:

1. we have three light neutrinos which are almost the same as the active neutrinos;
2. we have \mathcal{N} heavy neutrinos which are almost the same as the sterile neutrinos.

⁴The full matrix should also have some complex phases, but for our heuristic discussion they are not important.

Note that from now on for simplicity we will refer to the heavy neutrino states as HNLs. The interactions of these HNLs are similar to the ones of active neutrinos (eqs. (1.25) and (1.26)):

$$\mathcal{L}_{\text{int}} = \frac{g}{2\sqrt{2}} \overline{N_I^c} \sum_{\alpha} V_{\alpha}^* \gamma^{\mu} (1 - \gamma^5) \ell_{\alpha} W_{\mu}^+ + \frac{g}{4 \cos \theta_W} \overline{N_I^c} \sum_{\alpha} V_{\alpha}^* \gamma^{\mu} (1 - \gamma^5) \nu_{\alpha} Z_{\mu} , \quad (1.77)$$

but with a suppression factor $V_{\alpha} = M_{\alpha I}^D M_I^{-1} \ll 1$.

It can be shown that when HNLs are not degenerate in mass, they are produced and decay independently and do not oscillate between themselves. So to describe the phenomenology of HNLs, it is sufficient to use 1 HNL, which only requires 4 parameters: its mass, which we refer to as m_N , and its 3 couplings, which we refer to as $V_{\alpha} (\alpha = e, \mu, \tau)$ [16].

1.3 Heavy Neutral Leptons in hadronic machines

In hadronic machines, where we typically have a proton-nucleus collision or a proton-proton collision, there are three main processes that can produce HNLs:

1. HNL from decays of hadrons;
2. HNL from deep inelastic scattering of protons on nucleons;
3. HNL from coherent proton-nucleus scattering;

Ref. [16] reviews how significant each of these processes is for HNL mass $0.5 \text{ GeV} \lesssim m_N \lesssim 5 \text{ GeV}$. Here we will briefly summarize some main results.

1.3.1 Production from hadron decays

The first thing that we must consider is that the main channels of production must be the ones from the lightest hadrons of each flavor. This is because:

- they decay only through the weak interaction;
- they have a small decay width, that makes sterile production more efficient; other hadrons will be dominated by strong interactions.

Baryon vs meson production

Production from baryons is usually less relevant than from mesons because:

- because of baryon number conservation, a baryon decay must always include in the final state a proton or neutron (or heavier baryon) which will inevitably reduce the kinematical window for a HNL by at least $\sim 1 \text{ GeV}$;
- since baryons are fermions, they must decay to three or more particles to be able to produce a HNL (while mesons can have purely leptonic 2-body decays).

For instance, if we consider the lightest baryon of a given flavor decaying into a final state with a HNL, the HNL with the highest mass that can be produced is lighter than the one produced by the lightest meson of the same flavor [17].

HNLs can be produced in meson decays via a 2-body leptonic decays (fig. 1.3 left) or a semileptonic decay (fig. 1.3 right).

Light unflavored and strange mesons

The relevant mesons are:

$$\pi^+(u\bar{d} \text{ 140 MeV}) , \quad K^+(u\bar{s} \text{ 494 MeV}) , \quad K_S^0(d\bar{s} \text{ 498 MeV}) , \quad K_L^0(d\bar{s} \text{ 498 MeV}) .$$

(the neutral pion decays almost exclusively to photons). The pion is the lightest meson and thus can only decay to a HNL through a 2-body leptonic decay $\pi^+ \rightarrow \ell^+ N$. The kaon can also decay through a 3-body decay $K^+ \rightarrow \pi^0 \ell^+ N$ and $K_{L/S}^0 \rightarrow \pi^- \ell^+ N$. The resulting branching fractions are shown in fig. 1.4.

These channels are relevant if the HNL mass is $m_N \lesssim 400 \text{ MeV}$. An interesting thing to note is that the branching fractions in fig. 1.4 can reach values higher than 1. This is because with a massive neutrino, there is an enhancement effect compared to a massless neutrino (it's 'easier' to produce a particle with helicity opposite to its chirality if it is very massive).

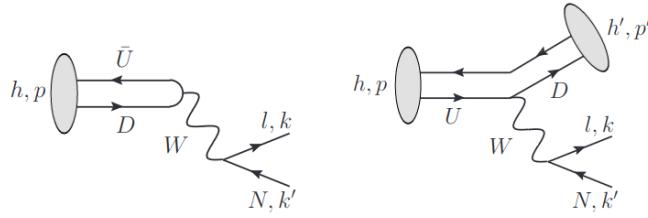


Figure 1.3: Meson decays that can produce HNLs. Left: 2-body leptonic decay. Right: semileptonic decay.

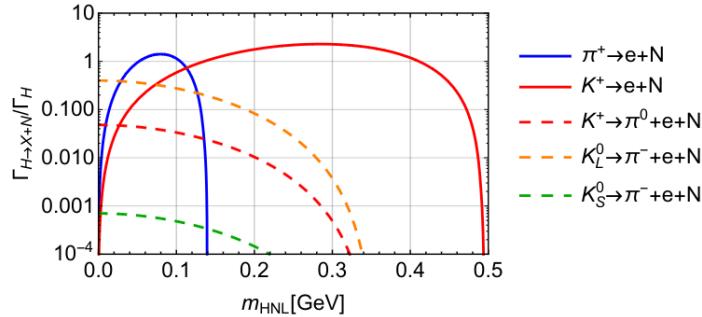


Figure 1.4: Decay width of light mesons into HNL related to the total measured decay width. This is built by taking $V_e = 1$ and $V_\mu = V_\tau = 0$. Taken from Ref. [16].

Charmed mesons

The relevant mesons are:

$$D^0(c\bar{u} \text{ 1865 MeV}), \quad D^+(c\bar{d} \text{ 1870 MeV}), \quad D_s(c\bar{s} \text{ 1968 MeV}).$$

D^0 is neutral and cannot decay with a 2-body purely leptonic decay. Because of the CKM matrix, the most favored decay is to kaons ($|V_{cd}| \simeq 0.22|V_{cs}|$). This means that for these channels the sterile mass is limited to $m_N < m_D - m_K \approx 1.4$ GeV. On the other hand, the D^+ and D_s mesons can decay through a 2-body leptonic decay and thus can produce a sterile neutrino with mass almost equal to their own. The resulting branching fractions are shown in fig. 1.5.

The plot shows that the branching fractions for the D_s 2-body decay are the highest: this makes it a promising channel for the search of HNLs. The decay channel $D_s^+ \rightarrow \mu^+ N$ is precisely the decay channel that is studied in this thesis work (see chapter 3).

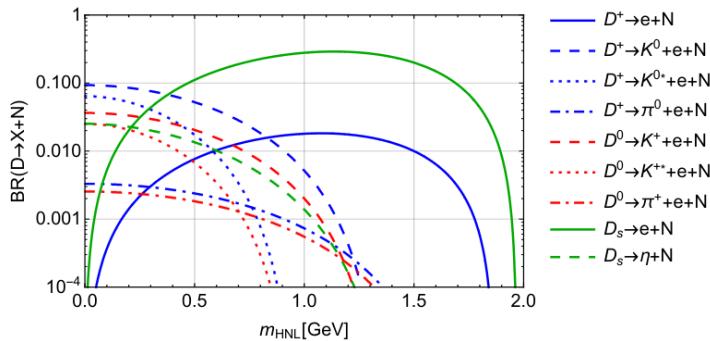


Figure 1.5: Branching ratio of charmed mesons into HNL. This is built by taking $V_e = 1$ and $V_\mu = V_\tau = 0$. Taken from Ref. [16].

Bottom mesons

The relevant mesons are:

$$B^- (b\bar{u} \text{ 5279 MeV}), \quad B^0 (b\bar{d} \text{ 5280 MeV}), \quad B_s (b\bar{s} \text{ 5367 MeV}), \quad B_c (b\bar{c} \text{ 6276 MeV}).$$

Just like we have said for the D^0 , B^0 and B_s decay with a 3-body decay with a final state meson. The decay favored by the CKM matrix is the one to D mesons ($|V_{ub}| \simeq 0.1|V_{cb}|$). This means that for these channels the sterile mass is limited to $m_N < m_B - m_D \approx 3.4$ GeV. The charged mesons B^- and B_c can decay through a 2-body leptonic decay and thus can produce a HNL with mass almost equal to their own. The resulting branching fractions are shown in fig. 1.6.

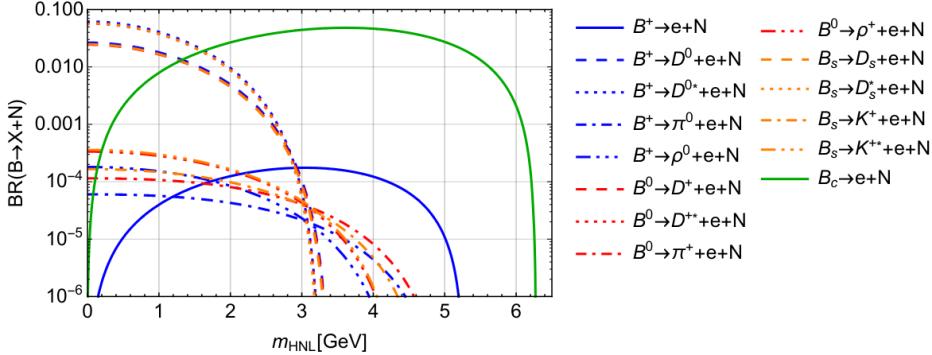


Figure 1.6: Branching ratio of bottom mesons into HNL. This is built by taking $V_e = 1$ and $V_\mu = V_\tau = 0$. Taken from Ref. [16].

1.3.2 Other production channels

Deep inelastic scattering

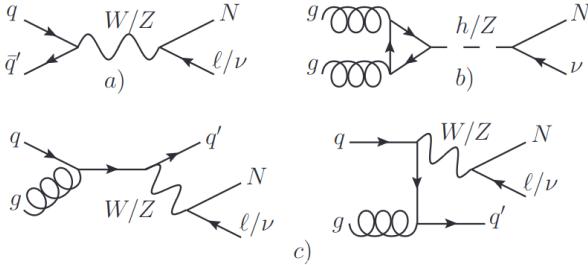


Figure 1.7: Production channels for HNLs. a) Drell-Yan process, b) gluon fusion, c) quark-gluon fusion.

The possible Feynman diagrams involved in this process are shown in fig. 1.7. These channels turn out to be strongly suppressed compared to production from meson decay [16]. To understand this intuitively, consider the center of mass energy in the parton-parton frame:

$$s_{\bar{q}q'} = sx_1x_2 , \quad (1.78)$$

where x_1 and x_2 are the momentum fractions of the partons and s is the center of mass energy of the proton-proton collision. The parton distribution function (PDF) is defined as the probability of finding a parton with momentum fraction x in a proton. At high energies, the PDF is strongly peaked at small x and decreases rapidly for larger x . This means that the center of mass energy in the parton-parton frame is small and thus the production of a HNL is suppressed.

Coherent proton-nucleus scattering

These processes may be enhanced by using heavy nuclei with large Z and by the fact that there is a large center of mass energy. However, they turn out to be subdominant compared to production from meson decay [16].

Tau lepton decays

A promising channel is the τ decay of $\tau \rightarrow NX$ and if X hadronizes into charged pions or kaons, mass and energy reconstruction can be very precise [5]. It is very important because the constraints on $|V_\tau|$ are much weaker than the ones on $|V_e|$ and $|V_\mu|$ (see section 1.4).

1.3.3 Sterile neutrino decays

As we can understand by looking at the interaction terms of the HNL (eq. (1.77)), the Feynman diagrams involved in the HNL decay are those in fig. 1.8. This means that we can have purely leptonic decay or semileptonic decays. In the case of quarks in the final state their hadronization process must be taken into account.

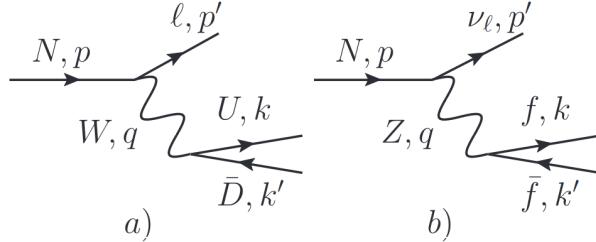


Figure 1.8: Feynman diagrams for the decay of a HNL; a) is charged-current decay, b) is neutral-current decay. U and D refers to any up-like or down-like fermion (quark or lepton), while f refers to any fermion.

Figure 1.9 shows the branching ratios of HNL decays into neutrinos, leptons and hadrons. An important thing we can notice is that semileptonic decays quickly become prevalent when $m_N \gtrsim m_\pi$.

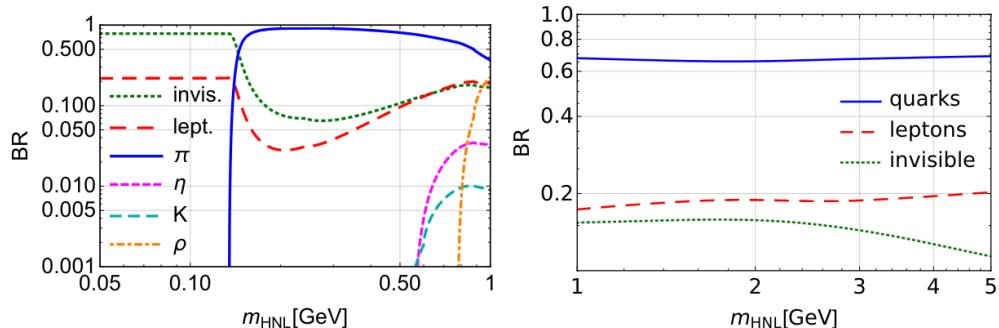


Figure 1.9: Branching ratios of HNL decays, assuming $V_e : V_\mu : V_\tau = 1 : 1 : 1$. Left is $m_N < 1 \text{ GeV}$ and right is $1 \text{ GeV} < m_N < 5 \text{ GeV}$ [16].

Another aspect to be considered is that quarks could hadronize into a single meson or multi-hadron states. Figure 1.10 shows that multi-hadron states are more than 50% when $m_N \gtrsim 2 \text{ GeV}$.

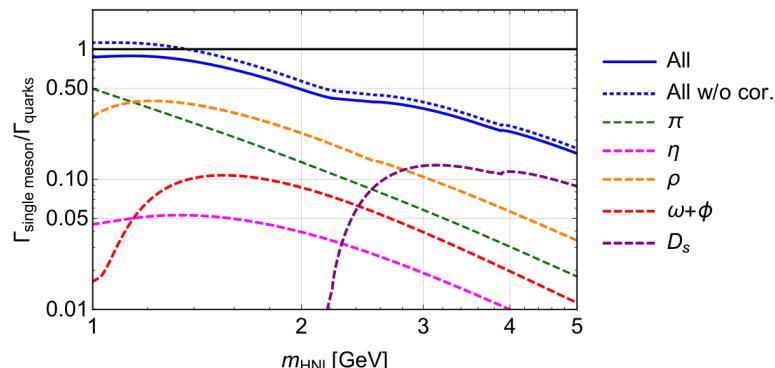


Figure 1.10: HNL decay widths into single meson channels, divided by the total decay width into quarks with QCD corrections [16].

1.4 Search for Heavy Neutral Leptons

Experimental searches have been performed in several experiments, but no evidence of HNLs has been found so far. Ref. [5] contains a review of the experimental searches and limits on HNLs. Different kinds of experiments can deal with different ranges of masses and mixing parameters. We mainly focus on the mass range $0.1 \text{ GeV} \lesssim m_N \lesssim 500 \text{ GeV}$ and briefly mention other searches. Figure 1.11 shows the current constraints on the mixing angle $|V_e|$ as a function of the HNL mass m_N from different experiments, fig. 1.12 shows the same for $|V_\mu|$ and fig. 1.13 for $|V_\tau|$.

Neutrino oscillation experiments

Neutrino oscillation experiments, both on solar or atmospheric neutrinos and on neutrinos from reactors and beams, combined with cosmological data, virtually exclude the existence of sterile neutrinos with masses below $\sim 1 \text{ eV}$ (see for instance refs. [18, 19]).

Neutrino-less double beta decay

Strong constraints on the $m_N - |V_e|$ plane can be obtained from the current limits on lifetime of neutrino-less double beta decay of experiments like KamLAND-Zen and GERDA (see for instance ref. [20, 21]).

1.4.1 Peak searches

Peak searches in the decays of mesons are very powerful probes. Figures 1.11 and 1.12 show that very stringent constraint can be obtained in the mass range $0.1 \text{ GeV} \lesssim m_N \lesssim 0.5 \text{ GeV}$. These constraints come from the 2-body decays of pions and kaons mentioned in section 1.3.1. For a decay $X \rightarrow \ell N$, in the rest frame of the decaying particle X , we have, due to energy-momentum conservation:

$$E_\ell = \frac{M_X^2 + m_\ell^2 - m_N^2}{2M_X}, \quad (1.79)$$

so there should be a peak in the energy spectrum of ℓ , other than the one already present for the known active neutrino. These limits are shown in the pictures with labels $X \rightarrow \ell\nu$.

Less stringent limits that are based on peak searches are the ones of the Belle experiment. These are based on the decay $B \rightarrow X\ell N$ where the B s are produced by the $\Upsilon(4S)$ resonance [22]. Figure 1.13 also shows projected limits based on using large datasets of τ decays from B-factories [23].

All these bounds have minimal assumptions and are based almost exclusively on kinematical constraints, which makes them very robust [5].

1.4.2 Beam dump experiments

Figures 1.11 to 1.13 show limits coming from beam dump experiments: PS191, CHARM, IHEP-JINR, BEBC, FMMF, NuTeV and NOMAD with the corresponding labels. These searches are based on searching for HNLs produced by mesons that subsequently decay into some kind of visible final state. The products are searched by placing the detector at relatively large distances away from the target, to allow the HNL to decay before reaching the detector [5]. The limits labelled as ‘LBNE’ refer to projected limits calculated for the future DUNE experiment [24].

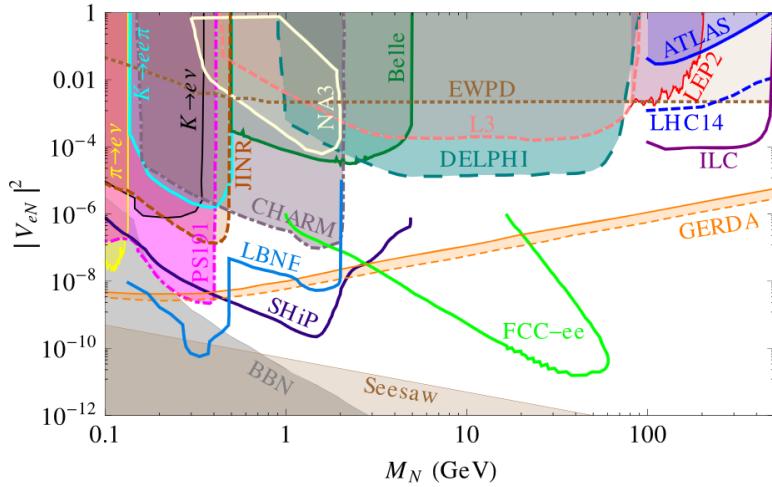


Figure 1.11: Constraints on the mixing angle $|V_e|$ as a function of the HNL mass m_N from different experiments. Taken from Ref. [5].

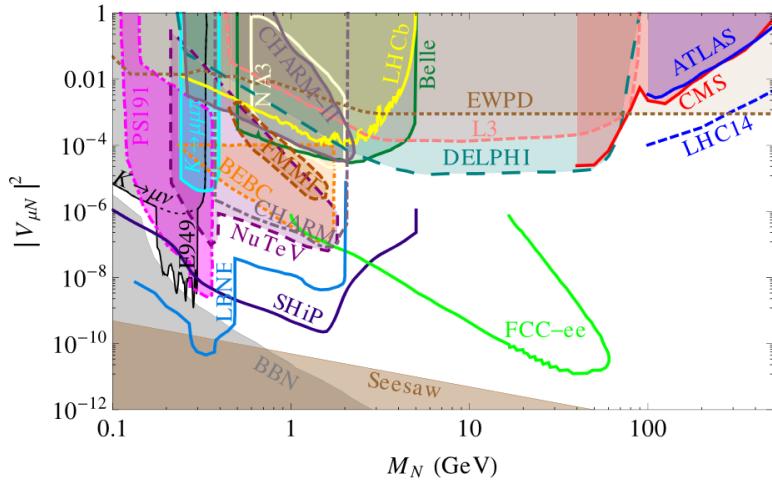


Figure 1.12: Constraints on the mixing angle $|V_\mu|$ as a function of the HNL mass m_N from different experiments. Taken from Ref. [5].

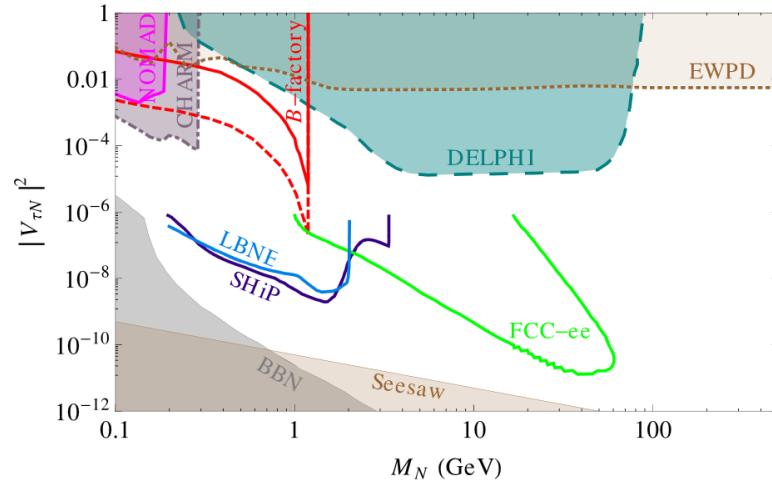


Figure 1.13: Constraints on the mixing angle $|V_\tau|$ as a function of the HNL mass m_N from different experiments. Taken from Ref. [5].

1.4.3 Electroweak based searches

Direct search in Z decays

Direct search for sterile neutrinos in Z decays consists in looking for the decay $Z \rightarrow \nu N$ where N is a HNL (this is only possible if $m_N < m_Z$). N will then decay either in a neutral current or in a charged current process:

$$\begin{aligned} N \rightarrow Z^* \nu &\quad Z^* \rightarrow \ell\ell, \nu\nu, qq \quad (\text{NC}) \\ N \rightarrow W^* \ell &\quad W^* \rightarrow \ell\nu, qq' \quad (\text{CC}) \end{aligned} \quad (1.80)$$

Figures 1.11 to 1.13 show the limits based on this process that have been calculated for the LEP experiments DELPHI and L3 [25, 26] with labels ‘DELPHI’ and ‘L3’. The future experiment FCC-ee [27] could provide new very strong constraints, that are shown in the figures with the label ‘FCC-ee’.

Electroweak precision tests

Figures 1.11 to 1.13 under the label ‘EWPD’ show limits based on precision measurements on various EW observables, such as Z invisible decay width, unitarity of leptonic mixing matrix, lepton universality [5].

1.4.4 Collider direct searches

Direct searches at LHC experiments ATLAS and CMS are shown in Figures 1.11 and 1.12 under the label ‘ATLAS’ and ‘CMS’. The limits shown are based on a set of data from a $\sqrt{s} = 8$ TeV run [28, 29]. The final states under consideration are dilepton plus jets, that can come from:

$$pp \rightarrow W^* \rightarrow N\ell^\pm \rightarrow \ell^\pm \ell^\pm jj. \quad (1.81)$$

This channel is very significant because of the possibility of same sign leptons in the final state which would be a strong evidence of the existence of HNLs, because of the violation of lepton number conservation [5]. A compilation of experimental limits on the mixing angles $|V_\alpha|$ is provided by the Particle Data Group (PDG) [30]. Several more recent results by the CMS are also available (see for instance [31, 32, 33, 34, 35]) that are not present in the aforementioned review.

1.5 Theoretical inputs for the HNL search from D_s^+ decays

This thesis work is focused on the search for HNLs from the decay of D_s^+ mesons. In particular, the decay channel under consideration is $D_s^+ \rightarrow \mu^+ N$ followed by $N \rightarrow \mu\pi$, where N is a HNL. As discussed in section 1.3.1, the $D_s^+ \rightarrow \mu^+ N$ has the highest branching ratios among the charmed mesons. The theoretical predictions on the branching ratios of the decay channel under consideration is necessary for the signal yield estimation, discussed in section 3.1.1. In this study, we make no assumption about the Majorana or Dirac nature of such neutrino and look for both same sign and opposite sign signatures.

The branching fraction of the $D_s^+ \rightarrow N\mu^+$ decay can be obtained by dividing the width of this channel by the total decay width of the D_s^+ meson. The width of this channel can be computed using the following equation for a generic meson h [16]:

$$\Gamma(h \rightarrow \ell_\alpha N) = \frac{G_F^2 f_h^2 m_h^3}{8\pi} |\mathcal{V}_{UD}|^2 |V_\alpha|^2 [y_N^2 + y_\ell^2 - (y_N^2 - y_\ell^2)] \sqrt{\lambda(1, y_N^2, y_\ell^2)}, \quad (1.82)$$

where G_F is the Fermi constant, f_h is the meson form factor, \mathcal{V}_{UD} is the Cabibbo-Kobayashi-Maskawa matrix element associated to the valence quarks of the meson, V_α is the mixing between the HNL and the active neutrino, $y_\ell = m_\ell/m_h$, $y_N = m_N/m_h$ are defined from the masses of the lepton (m_ℓ), the meson (m_h), and the HNL (m_N). λ is the Källén function defined as:

$$\lambda(a, b, c) = a^2 + b^2 + c^2 - 2ab - 2ac - 2bc. \quad (1.83)$$

The total decay width of the D_s^+ is obtained from its lifetime as

$$\Gamma(D_s^+) = \frac{\hbar}{\tau(D_s^+)}. \quad (1.84)$$

The branching fraction of the $N \rightarrow \mu\pi$ decay can be obtained by dividing the width of this channel by the total decay width of the HNL. The width of this channel can be computed using the following equation [16]:

$$\Gamma(N \rightarrow \ell_\alpha h) = \frac{G_F^2 f_h^2 |\mathcal{V}_{UD}|^2 |V_\alpha|^2 m_N^3}{16\pi} [1 - x_\ell^2 - x_h^2 (1 + x_\ell^2)] \sqrt{\lambda(1, x_\ell^2, x_h^2)}, \quad (1.85)$$

where f_h is the final state meson form factor, $x_\ell = m_\ell/m_N$, $x_h = m_h/m_N$.

The total width of the HNL is computed from the equation [15]:

$$\Gamma_N = \frac{G_F^2 m_N^5}{96\pi^3} |V_\mu|^2 \times (10.95). \quad (1.86)$$

In all computations above, the required constants are taken from their values reported on the PDG review of particle physics [13].

Chapter 2

The Large Hadron Collider and CMS experiment

2.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC) is the largest and highest-energy particle accelerator in the world. It is installed in an underground tunnel of 27 km circumference, located at the CERN laboratory in Geneva, Switzerland. Two beam pipes run through the tunnel in opposite directions, carrying two beams of protons or heavy ions. The particles inside the beam pipes are accelerated by radio-frequency cavities and guided by superconducting magnets. The beams are made to collide at four points along the ring, where the main experiments are located:

- ALICE: *A Large Ion Collider Experiment* focused on the study of heavy-ion collisions;
- ATLAS: *A Toroidal LHC ApparatuS* a general-purpose detector;
- CMS: *Compact Muon Solenoid* a general-purpose detector;
- LHCb: *Large Hadron Collider beauty* focused on heavy flavor physics.

In order to accelerate the particles, the LHC uses a chain of smaller accelerators, which progressively increase the energy of the beams before injecting them into the main LHC ring. Figure 2.1 shows the CERN accelerator complex and some other experiments at CERN that use the same accelerator chain. The different stages for a beam of protons are:

- extraction from hydrogen gas;
- LINAC4 linear accelerator;
- Proton Synchrotron Booster (PSB);
- Proton Synchrotron (PS);
- Super Proton Synchrotron (SPS);
- injection into the LHC.

For a beam of lead ions,

- vaporization of lead;
- LINAC3 linear accelerator;
- Low Energy Ion Ring (LEIR);
- injection into the PS;

and then the same stages as the protons.

The most important figures for a particle accelerator are the luminosity and the center of mass energy. The instantaneous luminosity $\mathcal{L}_{\text{inst}}(t)$ is proportional to the rate of events dN/dt through the cross-section σ :

$$\frac{dN}{dt} = \sigma \mathcal{L}_{\text{inst}}(t) , \quad (2.1)$$

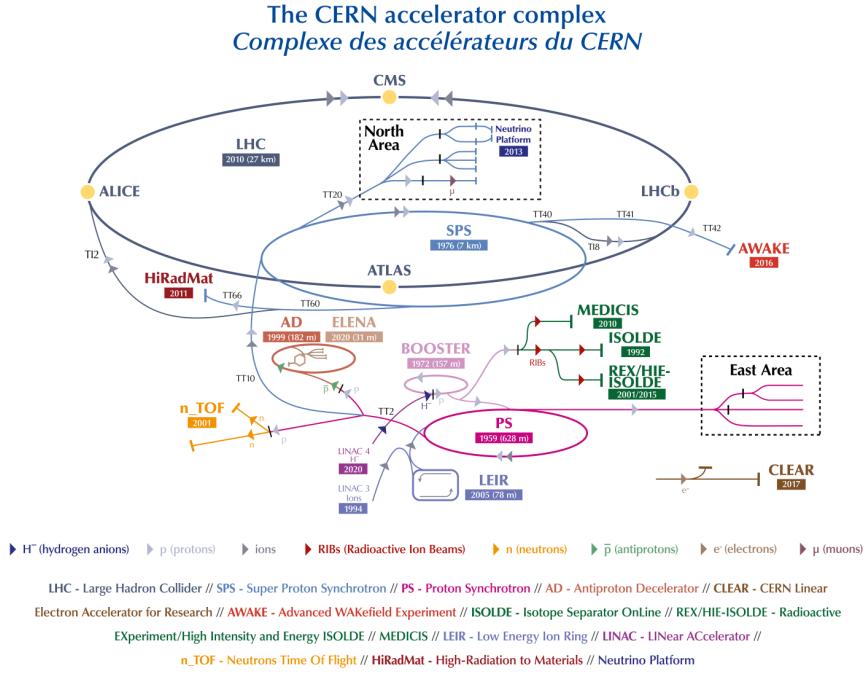


Figure 2.1: The CERN accelerator complex, from Ref. [36].

and the number of events N_{tot} is the integration of this relation over a specific time interval

$$N_{\text{tot}} = \sigma \int_{t_1}^{t_2} \mathcal{L}(t) dt = \sigma L_{\text{int}} \quad (2.2)$$

where L_{int} is called integrated luminosity.

In the assumption of a perfectly head-on collision, the center of mass energy \sqrt{s} is related to the energy of the colliding particles E_1 and E_2 by

$$\sqrt{s} = 2\sqrt{E_1 E_2}, \quad (2.3)$$

and it is a measure of the maximum energy available for the production of new particles in the collision. In its Run-1, from 2010 to 2013, the LHC has reached peak values of instant luminosity of $\mathcal{L}_{\text{inst}} = 7 \times 10^{33} \text{ cm}^2 \text{s}^{-1}$ with a center of mass energy of $\sqrt{s} = 8 \text{ TeV}$ [37]. In its Run-2, from 2015 to 2018, it has reached peak values of instant luminosity of $\mathcal{L}_{\text{inst}} = 2 \times 10^{34} \text{ cm}^2 \text{s}^{-1}$ with a center of mass energy of $\sqrt{s} = 13 \text{ TeV}$ [38]. The center of mass energy of the LHC is limited by the maximum magnetic field that can be produced by the superconducting magnets that guide the particles along their trajectory. The relation between the momentum p of a proton and the radius R of its trajectory is:

$$p[\text{GeV}/c] \simeq 0.3 \times B[\text{T}]R[\text{m}], \quad (2.4)$$

thus with a radius of $R = 2804 \text{ m}$ fixed by the geometry of the tunnel and a maximum magnetic field of $B = 8.33 \text{ T}$ the maximum momentum that can be theoretically achieved by a proton is $p \sim 7 \text{ TeV}/c$. Producing such a high magnetic field requires the use of superconducting NbTi dipole magnets cooled down to $\sim 2 \text{ K}$ [39].

2.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is a general-purpose detector located at one of the collision points of the LHC. Its goal is studying a wide range of physics processes, from the Higgs boson to the search for new physics beyond the SM. Its design is based on a cylindrical geometry, centered around the interaction point, with several layers of sub-detectors that allow to reconstruct the properties of the particles produced in the collisions [40]. Figure 2.2 shows a general view of the CMS detector and its sub-detectors. The main components, from innermost to outermost, are:

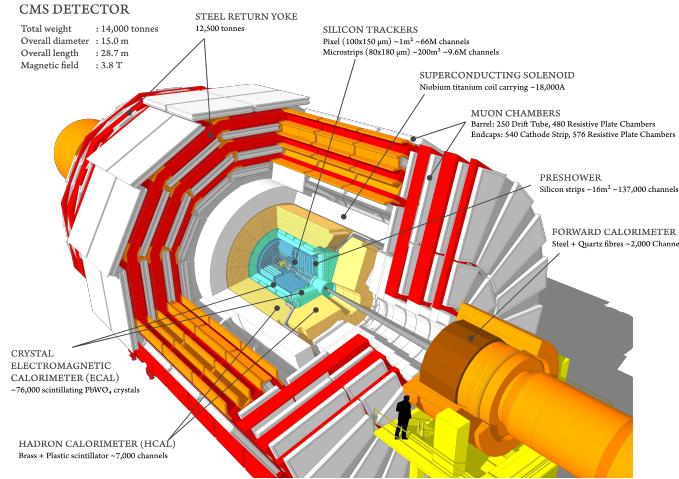


Figure 2.2: General view of the CMS detector and its sub-detectors, from [41].

- the tracking system, which allows to reconstruct the trajectories of charged particles;
- the electromagnetic calorimeter, which measures the energy of electrons and photons;
- the hadronic calorimeter, which measures the energy of hadrons;
- the superconducting solenoid;
- the muon system, which allows to reconstruct the trajectories of muons.

When dealing with cylindrical detectors, it is customary to distinguish between the barrel region, which is the central region of the detector, and the endcap region, which is the region at the two ends of the cylinder.

Coordinate system

The coordinate system used in CMS is a right-handed Cartesian system, with the origin at the nominal interaction point. The x axis points towards the center of the LHC ring, the y axis points upwards, and the z axis points along the beam direction. The azimuthal angle ϕ is measured in the $x - y$ plane and ranges from $-\pi$ to π . The polar angle θ is measured from the z axis and ranges from 0 to π . For practical reasons, the pseudorapidity η is often used instead of θ , defined as $\eta = -\ln \tan(\theta/2)$ and ranges from $-\infty$ to $+\infty$. Figure 2.3 shows a schematic view of the CMS coordinate system. When referring to the transverse plane, we mean the plane perpendicular to the beam direction, described by a set of two coordinates (x, y) or (r, ϕ) , where r is the distance from the beam axis. The transverse momentum p_T is the momentum of a particle projected onto the transverse plane, and the transverse energy E_T is defined as $E_T = \sqrt{m^2 + p_T^2}$ [40].

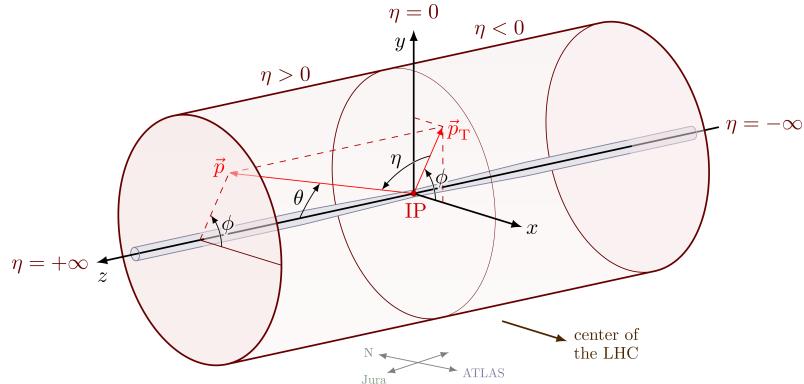


Figure 2.3: Schematic view of the CMS coordinate system.

2.2.1 The tracking system

The tracking system is the closest sub-detector to the interaction point, and it is designed to measure the trajectories of charged particles and reconstruct the vertices of the collisions. It has two main components: the pixel detector and the silicon strip tracker.

The pixel detector is the innermost component, and as such it requires a very high granularity to cope with the high particle flux. The current pixel detector has been installed in 2017 during the Phase-1 upgrade to replace the original one, which was installed in 2008. It is composed of four concentric layers in the barrel region and three disks in each endcap region, for a total coverage of $|\eta| < 2.5$. The concentric layers are located at a radial distance of 2.9, 6.8, 10.9, 16.0 cm from the beam axis while the disks on each end are located at 29.1, 39.6, 51.6 cm from the center of the detector [42]. The layout of the current pixel detector, as well as the original one, is shown in fig. 2.4.

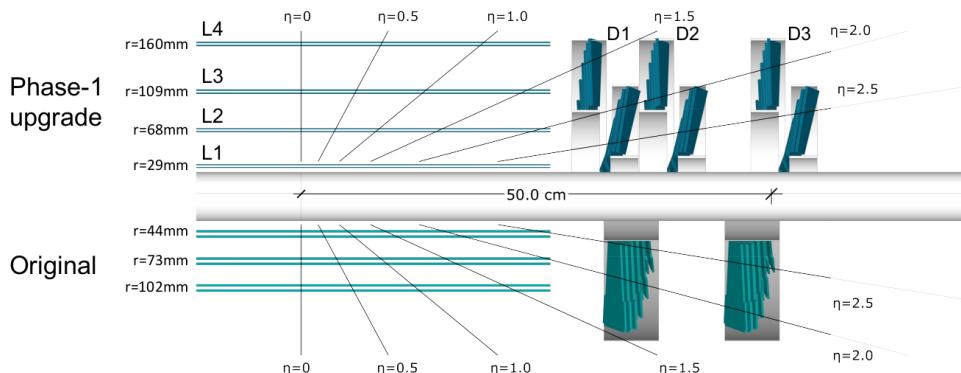


Figure 2.4: Layout of the CMS pixel detector, with a comparison between the old and the new detector.

The pixel detector is built from 1184 modules in the barrel region and 672 modules in the endcap region. Each module is composed of a silicon sensor and 16 readout chips. The pixel size is $100 \times 150 \mu\text{m}^2$. The pixel layers provide 3-D measurements of the hits on their sensors, where two coordinates are given by the position of the hit on the sensor plane and the third is provided by the position of the sensor plane itself. The resolution on individual hits is $\sim 10 \mu\text{m}$ in the transversal plane, and $\sim 20 \mu\text{m}$ in the z direction [42].

The silicon strip tracker is outside the pixel detector, and it is composed of four subsystems: the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker Outer Barrel (TOB), and the Tracker EndCaps (TEC). The TIB and TID are composed of four barrel layers and three disks respectively, positioned in a radial distance range of $20 \text{ cm} < r < 55 \text{ cm}$. The TOB is positioned outside the TIB/TID system, and it is composed of six barrel layers, extending the coverage up to a radial distance of 116 cm. The TIB, TID, and TOB are in a z range of $|z| < 118 \text{ cm}$. Outside this range, we have the TEC. The TEC is composed of nine disks, and it covers the z range of $124 \text{ cm} < |z| < 282 \text{ cm}$. The total coverage of the silicon strip tracker in η is $|\eta| < 2.5$ [40]. The layout of the full pixel and strip tracker is shown in fig. 2.5.

The silicon strip detector is built from a total of 15 148 modules divided among the four subsystems. Each module carries one thin ($320 \mu\text{m}$) or two thick ($500 \mu\text{m}$) silicon sensors. The sensors on the modules have different geometries depending on the subsystem, with typical dimensions of about $6 \times 12 \text{ cm}^2$ in the inner barrel and $10 \times 9 \text{ cm}^2$ in the outer barrel. The strip pitch is different for each subsystem, and ranges from $\sim 100 \mu\text{m}$ to $\sim 180 \mu\text{m}$. The strip layers provide 2-D measurements of the hits on their sensors but some layers (shown in blue in fig. 2.5) have back-to-back modules with a shift in the stereo angle, which allows 3-D measurements. The resolution on individual hits on the $r - \phi$ plane for the barrel part is $\sim 20 - 30 \mu\text{m}$ for the TIB/TID and $\sim 30 - 50 \mu\text{m}$ for the TOB [40].

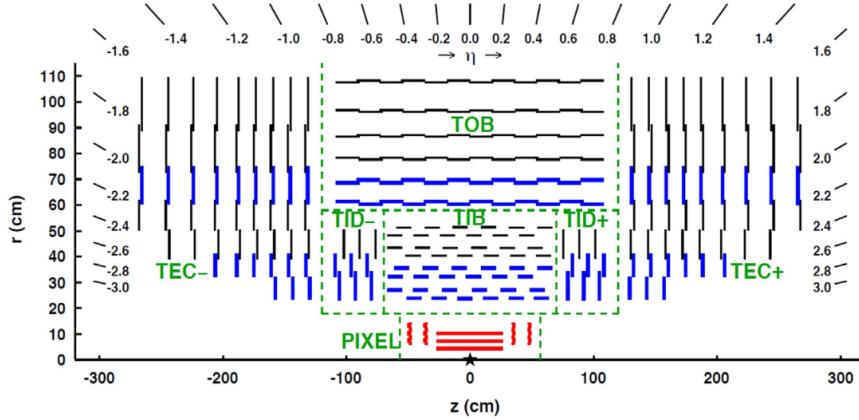


Figure 2.5: Layout of the full tracker system of the CMS experiment. The pixel detector layers shown in the picture are not updated to the Phase-1 upgrade. The collision point is indicated by a star. Green dashed lines separate the different subsystems. Black lines indicate layers that provide $2 - D$ hits, while the blue lines layers that provide $3 - D$ hits. The picture is taken from Ref. [43].

2.2.2 The electromagnetic calorimeter

The electromagnetic calorimeter of CMS (ECAL) is a hermetic and homogeneous calorimeter. It consists of 61 200 lead tungstate (PbWO_4) crystals, which cover the barrel region with $|\eta| < 1.48$ and 7 324 crystals in each of the endcap regions with $1.48 < |\eta| < 3.00$. The length of the crystals in the barrel region corresponds to 25.8 radiation lengths (X_0) and $24.7 X_0$ in the endcap region. Preshower detectors are located in front of each endcap crystals. The preshower detectors consist of two planes of silicon sensors interleaved with lead, and its thickness corresponds to $3 X_0$. These preshower detectors are important for distinguishing between single photons and neutral pions, which decay into two photons. The main purpose of the ECAL is to measure the energy of electrons and photons. The energy deposited in the ECAL crystals is detected in the form of scintillation light by avalanche photodiodes (APDs) in the barrel and by vacuum phototriodes (VPTs) in the endcaps [40].

The crystals of the barrel region are contained in thin-walled structures, called submodules, which are of different shapes and sizes depending on the position in the detector. The submodules are arranged into modules, with a total of 400 or 500 crystals each, and the modules are arranged into groups of 4 called supermodules. On the other hand, crystals in the endcap region are grouped into 5×5 called supercrystals. Each endcap is divided into two halves, called *Dees*, which are built from 138 supercrystals (and 18 partial supercrystals) each. All modular units are placed in quasi-projective geometry, to minimize the possibility of particles passing through the gaps between the crystals: the axis of the crystals forms a small angle (3°) with respect to the vector pointing to the nominal interaction point [40]. The layout of the ECAL and its modular units are shown in fig. 2.6.

The barrel region crystals have a front face of $22 \times 22 \text{ mm}^2$, a rear face of $26 \times 26 \text{ mm}^2$, and a length of 230 mm. The endcap region crystals have a front face of $28.62 \times 28.62 \text{ mm}^2$, a rear face of $30 \times 30 \text{ mm}^2$, and a length of 220 mm. PbWO_4 has a density of 8.28 g/cm^3 , short X_0 of 0.89 cm, and a small Molière radius of 2.2 cm. This allows having a compact calorimeter with fine granularity.

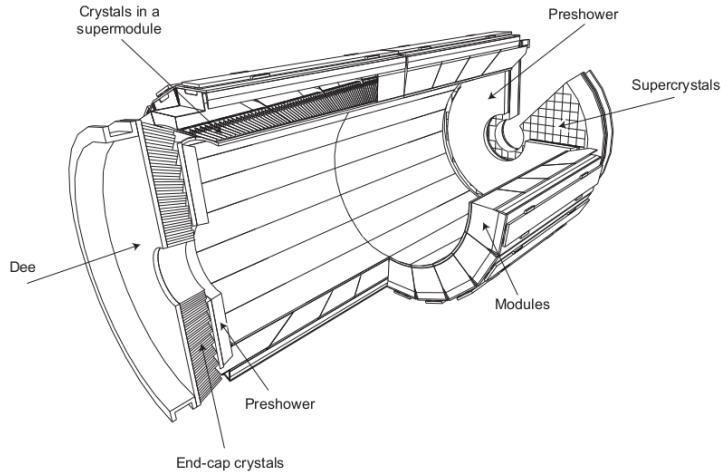


Figure 2.6: Layout of the CMS ECAL, with each modular unit that composes it.

2.2.3 The hadron calorimeter

The hadron calorimeter system of the CMS detector is designed to measure the energy of hadrons jets and is important for its contribution to the measurement of the missing transverse energy. It is composed of four subsystems: the Hadron Barrel (HB), the Hadron Endcap (HE), the Hadron Outer (HO), and the Hadron Forward (HF) calorimeters [44, 40].

The HB is a sampling calorimeter that covers the range $|\eta| < 1.4$. It is composed of 36 wedges, made of passive absorber layers interleaved with active plastic scintillator layers. The first and last layer of the absorber are made of steel plates for structural support, and the rest of the absorber is made of brass. In terms of interaction lengths (λ_I), the HB has a depth of $5.82 \lambda_I$ at $\eta = 0$ and $10.6 \lambda_I$ at $\eta = 1.3$. The active part of the calorimeter uses scintillator tiles and wavelength shifting fibers to collect the light. The HB is divided into two half-barrels along the z axis, and each half-barrel is divided into 18 azimuthal wedges, thus each wedge covers a $\Delta\phi$ of 20° . The HB has $\sim 40\,000$ scintillator tiles, that are grouped together into *megatiles* to limit the number of individual elements. Each megatile covers a $\Delta\phi$ of 5° and is divided into 16 $|\eta|$ sectors. Clear fibers are used to collect the light and transport it hybrid photodiodes (HPDs) for readout.

The layout of the HB and other hadron calorimeters is shown in fig. 2.7.

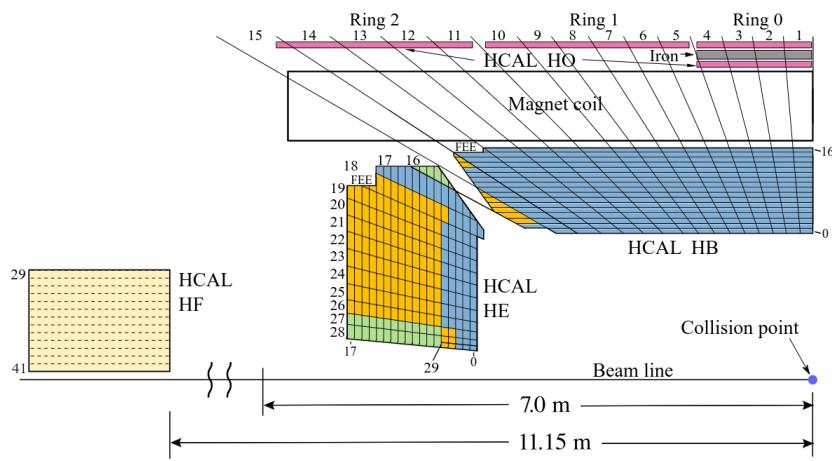


Figure 2.7: Layout of the CMS hadron calorimeter subsystems.

The HE is an endcap continuation of the HB, and it covers the range $1.3 < |\eta| < 3.0$. It is a calorimeter with the same sampling structure of the HB i.e. passive brass absorber interleaved with active scintillator tiles. The total length of the calorimeter, including the electromagnetic calorimeter in front of it, is about $10 \lambda_I$. It is divided into 18 sectors in ϕ , each covering 20° . The scintillator plates on either side are divided into 14 $|\eta|$ sectors. Like the HB the light is

collected by clear fibers and transported to HPDs for readout.

The HO is located in the central barrel region outside the solenoid magnet, and it is designed to measure the energy of the hadrons that are not contained by the stopping power of the EB and HB. The HO is physically constrained because it is placed between the solenoid and the magnet return yoke. The iron of the magnet return yoke and the solenoid itself serve as the absorber material, and the active part is composed of scintillator tiles. The HO is divided into 5 rings along z , and each ring is divided into 12 sectors in ϕ . It covers the range $|\eta| < 1.26$, and it extends the length of the calorimeter system to a minimum of $11.8 \lambda_I$ in the central region. The light is carried to silicon photomultipliers (SiPMs) for readout.

The HF is located in the forward region on either side, ~ 11 m away from the interaction point. It is designed to measure the energy of hadrons in the very harsh radiation environment of the forward region. Each HF module is composed of 18 wedges of steel with quartz fibers embedded in them. Unlike the other subsystems, the HF detects the Cherenkov light produced by the particles in the quartz fibers. Half of the fibers run over the full length of the module, while the other half only begin after the first 22 cm. The different fibers are read out separately and this arrangement makes it possible to distinguish between electromagnetic, which deposit a significant fraction of their energy in the first 22 cm, and hadronic showers, which penetrate deeper into the absorber. The HF covers the range $2.85 < |\eta| < 5.19$, the full absorber length along z is 165 cm which is $\sim 10 \lambda_I$.

2.2.4 The Muon system

Other than identifying muons and measuring their momentum, the muon system is also one of the most important components for the trigger system of CMS. It is composed of three different types of gaseous detectors: Drift Tubes Chambers (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). The DTs cover the region $|\eta| < 1.2$, the CSCs cover the region $0.9 < |\eta| < 2.4$, and the RPCs cover the region $|\eta| < 1.8$. The layout of the muon system is shown in fig. 2.8.

The DTs are organized in 4 stations of concentric cylinders divided into 5 ‘wheels’ along the z axis. The three inner stations are composed of 60 chambers, while the outermost station is composed of 70 chambers. There are 12 sectors in ϕ (following the geometry of the yoke) for each station. The main unit of the DT is a superlayer (SL): the outermost chambers in the 4th station have 2 SLs, while the chambers in the inner stations have 3 SLs. The SLs are composed of 4 layers rectangular drift cells, staggered by a half-cell. For DT chambers with 3 SLs, the two outer SLs have a wire parallel to the beam axis, which allows measurements on the transverse plane, while the inner SL has a wire perpendicular to the beam axis, thus providing full 3-D measurements of the muon position. The wire chambers are filled with a gas mixture of 85% Ar, 15% CO₂.

The CSCs system is meant to be the CMS endcap muon system. It is composed of 540 trapezoidal chambers, arranged in four stations that have approximately the same z position. In each station the chambers are placed to form rings at different radial distances from the beam axis, and each chamber covers 10° or 20° in ϕ . The CSCs are multiwire proportional chambers, composed of 6 layers of anode wires interleaved with 7 cathode strips panels. The cathode strips run radially, at a constant $\delta\phi$ width, and the anode wires run along the ϕ direction. Since the strip and wire directions are orthogonal, the CSCs provide 2-D measurements of the muon position. The CSCs are filled with a gas mixture of 40% Ar, 50% CO₂, and 10% CF₄.

The RPC system is used to provide a fast trigger signal for the muon system. The RPCs of the barrel part are arranged in 4 concentric stations, that are divided into 12 sectors in ϕ . For the two stations closest to the interaction point, two layers of RPCs are placed with the DTs in between, while for the two outermost stations, two layers of RPCs are placed side by side on the innermost side of the DTs. The RPCs of the endcap part are arranged in 3 disks. The endcap RPCs have a trapezoidal shape, to match the geometry of the CSCs, and they cover 10° or 20° in ϕ . The CMS RPC modules are double-gap RPCs: each module is composed of two gas gaps operated in avalanche mode, with a common pick-up strip in between. The gas mixture used in the RPCs is 96.2% R134a (C₂H₂F₄), 3.5% iC₄H₁₀, and 0.3% SF₆.

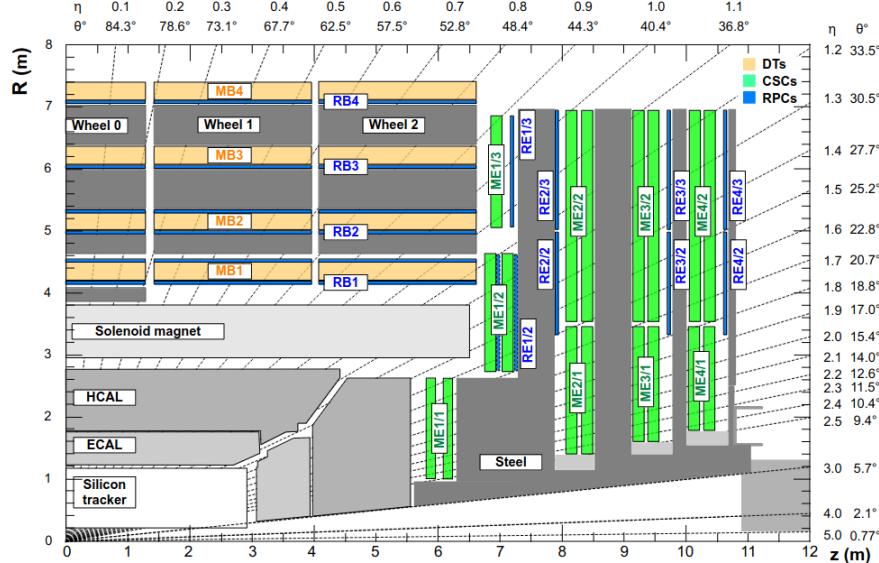


Figure 2.8: Layout of the CMS muon system, from Ref. [45].

2.3 The CMS trigger system and B-parking dataset

The CMS physics program ranges over a wide variety of physics processes. Since the discovery of the Higgs boson [46, 47], measurements of its properties have become one of the main goals, together with precision measurements of the SM properties in the sectors of electroweak interactions, quantum chromodynamics, top quarks, and the search for new physics beyond the SM.

The LHC collides bunches of particles with a maximum rate of ~ 40 MHz, meaning bunches collide every 25 ns. At the design luminosity of $10^{34} \text{ cm}^{-2} \text{s}^{-1}$, the proton-proton interaction frequency is more than 1 GHz. Of all collisions that take place, only a small fraction contains interesting physics processes for the CMS experiment, and for practical reasons, only a small fraction of these interesting events can be stored for further analysis. The purpose of the trigger system is to select the interesting events and bring the rate of events down to a level that can be stored and analyzed.

2.3.1 The CMS trigger system

The CMS trigger system has two levels [48]: the Level-1 (L1) trigger system and the High-Level Trigger system (HLT). The L1 trigger selects up to around 100 kHz of the most interesting events with a latency of 4 μs . The HLT performs a more detailed analysis of the events selected by the L1 trigger, reducing the rate to a few kHz with a latency of ~ 100 ms.

The L1 trigger

The L1 trigger is hardware-based, and it provides coarse information about the events by using the signals from the calorimeters and the muon system. The selections are based on a set of algorithms called ‘seeds’, that check that a given criterion is satisfied. A collection of seeds is called ‘menu’, and the menu is updated regularly to adapt to the changing conditions of the detector and the different physics goals of the experiment.

When an event satisfies the conditions of at least one seed in a menu, it is passed in the trigger chain for further analysis. This step initiates a readout of the full event from the detector’s data acquisition system, and the information of event is then sent to the HLT.

The trigger algorithms are usually based on criteria applied to one or more physics objects: muons, jets, tau leptons, photons, electrons, sum of transverse energy, or also a combination of different objects. The criteria are typically constraints on the objects p_T , E_T or η or some combination of them. These criteria are meant to be a first selection of the most interesting events: a typical interesting event contains some relatively high- p_T objects in the barrel region,

while most of the events have low- p_T objects in the forward regions. As an illustrative example, fig. 2.9 shows how the L1 trigger rate was allocated to various object seeds in a Run-2 menu.

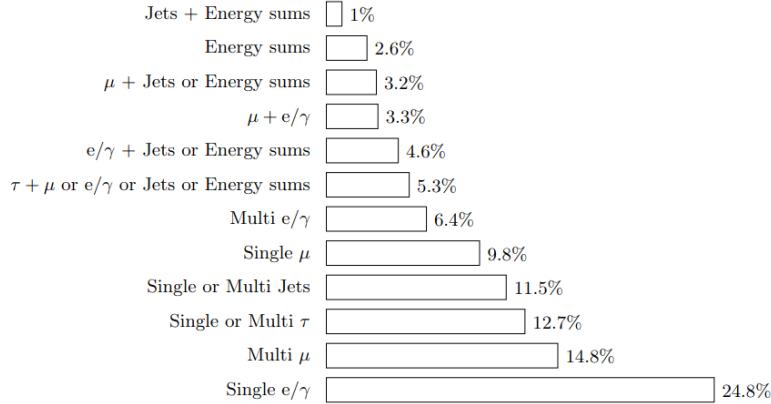


Figure 2.9: Fractions of the 100 kHz L1 trigger rate allocated to various single and multi-object triggers and cross triggers in a typical Run-2 menu.

An important concept in the trigger system is the *prescale*. The prescale is a factor that is used to reduce the rate of a trigger. It is indicated by a positive integer N , and it means that only one out of every N events that satisfy the seed condition is actually accepted. A prescale column is a set of prescales applied to each seed of a trigger menu. The prescale columns are updated depending on the instantaneous luminosity, in a way that maximizes the efficiency of the trigger system while also keeping the rate under the maximum manageable value of 100 kHz.

The following paragraphs briefly summarize the workflow of the L1 trigger [49].

The calorimeter trigger has two levels: Layer-1 receives the energy deposits from the calorimeters calibrates and sorts them, and then sends the information to Layer-2; Layer-2 uses the information to reconstruct physics objects such as electrons, photons, jets, and energy sums.

Electrons and photons are indistinguishable at the L1 trigger level, because no tracking information is available yet. The electron/photon candidates are built by looking for energy clusters in the ECAL+HCAL energy deposits around a ‘seed’ with energy threshold of $E_T > 2$ GeV. This same clustering technique is used for tau lepton candidates.

For jets, the L1 trigger uses a ‘sliding window’ technique, which spans the full $\eta - \phi$ space in 3×3 calorimeter regions. If the central region of the 3×3 matrix has E_T higher than the eight neighbors, and above a certain threshold, the region is flagged as a jet candidate.

The muon trigger system has three muon track finders (MTF): the barrel muons (BMTF), the DT-CSC overlap muons (OMTF) and the endcap muons (EMTF). The trigger primitives from the muon system are used to build muon track candidates. Each MTF uses a different algorithm that is optimized for the specific region of the muon system it covers. Each MTF sends up to 36 muon candidates to the Global Muon Trigger (GMT), which then selects the 8 best candidates.

The final L1 trigger decision is made by the Global Trigger (GT), which receives the muon and calorimeter objects and executes the selection based on the trigger menu. A diagram of the L1 trigger workflow is shown in fig. 2.10.

The HLT

The HLT consists of a farm of processors running high-level physics algorithms, which are based on a simplified version of the ones used for the offline reconstruction, with an output rate around 1kHz. The HLT menu is divided in different *paths*, and each path corresponds to a sequence of reconstruction and filtering modules that can rely on information from all the sub-detectors, including the tracker system. The modules within a path are ordered in increasing complexity, so that if the first module in a path fails, the rest of the path is not executed. The HLT can also use the ‘regionality’ of the detector to reduce the computing time, by considering only regions around the previously selected L1 objects. The final trigger decision is the logical OR of the decisions of the different paths. The trigger rate of the HLT can also be reduced with prescales, like the L1 trigger. When an event passes the HLT selection, it is stored and further processed at CERN Tier-0 center, to produce the reconstructed objects (described in section 2.4).

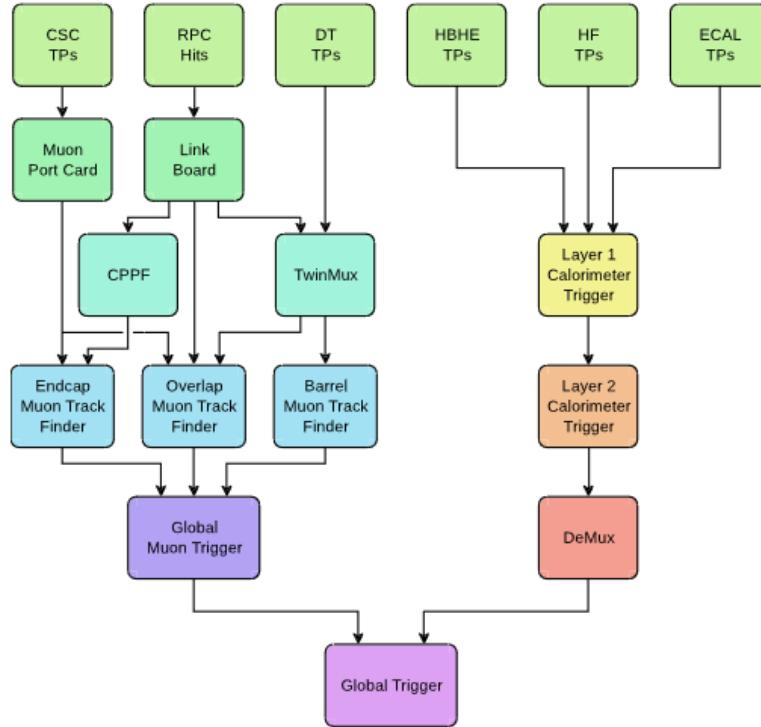


Figure 2.10: Diagram of the CMS L1 trigger system during Run-2.

2.3.2 The Data parking strategy

For the usual core physics program of CMS, data streams are promptly reconstructed at the CERN Tier-0 data center, and available for analysis within 48 hours. However, a new data-taking strategy was devised by the CMS experiment called ‘data parking’: this strategy allows storing events that pass the HLT selection directly to tape storage, without prompt reconstruction. Removing the need for prompt reconstruction also allows recording more than the usual 1 kHz of physics events. The rate is only constrained by the bandwidth of the CMS data acquisition system and the amount of tape storage space [50].

In particular, a dedicated stream called ‘B-Parking’, allowed to record a large unbiased sample of b hadron decays in the 2018 data-taking period at $\sqrt{s} = 13$ TeV [6].

The 2018 B-parking dataset

The trigger strategy of the 2018 B-Parking dataset is designed to record a large unbiased sample of b hadron decays, exploiting the fact that in proton-proton collisions at the LHC, b hadrons are often produced in pairs. The trigger targets the muon coming from semi-leptonic decays of b hadrons, that are typically produced as low- p_T non-prompt muons, since b hadrons usually have a significant travel distance of $\sim 0.1 - 0.5$ mm before decaying. The b hadron that produces the trigger muon is called ‘tag’, and the other b hadron (which is usually produced in the same event) is called ‘signal’, and is unbiased by the trigger logic.

The key achievement of the B-Parking dataset was collecting a large sample of data while leaving the core physics program unaffected. This is done by exploiting the natural decrease of the instantaneous luminosity at the end of an LHC fill. This makes it possible to use part of the bandwidth of the trigger system to record the B-Parking data, while still keeping the L1 trigger rate below the maximum manageable value of around 100 kHz. As illustrative example, fig. 2.11 shows the HLT rate of CMS physics and B-Parking in a typical fill: the CMS physics stream starts from a rate ~ 2 kHz which decreases over time, while the rate of the B-Parking stream starts from 0 and is increased in steps up to peaks of ~ 5 kHz.

The L1 B-Parking trigger requires the trigger muon to have $|\eta| < 1.5$ and a specific p_T threshold which evolves with time, and become less tight as the instantaneous luminosity decreases, or a trigger muon with $p_T > 22$ GeV without restrictions on η .

The HLT paths have the same p_T requirements of the L1 seeds and a further selection on

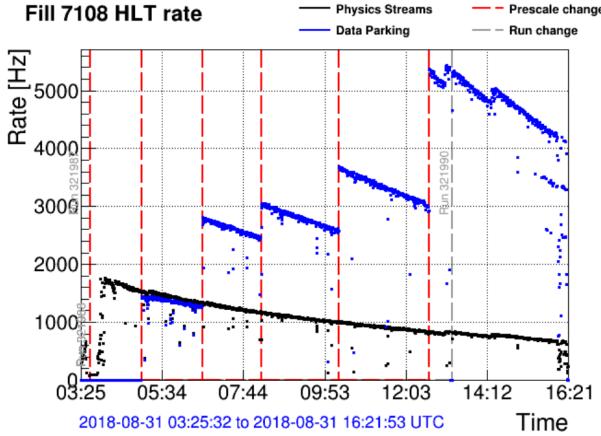


Figure 2.11: HLT rate of the CMS physics (black) and B-Parking (blue) streams over time in a typical fill. The red dashed lines indicate a change in the prescale values of the B-Parking trigger. The total time interval shown is approximately 13 hours.

the impact parameter significance of the muon (i.e. the ratio of the impact parameter to its uncertainty). The thresholds of the L1 seeds and HLT paths are relaxed as the luminosity decreases, which increases the trigger rate but also decreases the b hadron purity of the sample. Table 2.1 shows an example of the settings of the B-Parking trigger in a typical fill.

Settings	peak $\mathcal{L}_{\text{inst}}$ ($10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$)	L1 μp_T threshold (GeV)	HLT μp_T threshold (GeV)	HLT IPS threshold	trigger purity (%)	Peak rate kHz
1	1.7	12	12	6	92	1.5
2	1.5	10	10	6	87	2.8
3	1.3	9	9	5	86	3.0
4	1.1	8	8	5	83	3.7
5	0.9	7	7	4	59	5.4

Table 2.1: Summary of the settings of the B-Parking trigger, including the L1 and HLT thresholds. The b hadron purity column refers to the purity as computed from simulations.

The b hadron purity is defined as the fraction of events triggered by muons that actually come from a b meson decay:

$$P = \frac{N(B \rightarrow X\mu)}{N(\mu)}. \quad (2.5)$$

This is evaluated with simulations, and has an average value of ~ 0.75 which has been validated with data. The trigger purity in the data is measured by considering the $B^0 \rightarrow D^{*+}\mu\nu$ decay, which has a large branching fraction. The subsequent decays that are targeted are $D^{*+} \rightarrow D^0\pi^+ \rightarrow K^-\pi^+\pi^+$. The difference of invariant mass between the $K^-\pi^+\pi^+$, which form the D^{*+} candidate, and the $K^-\pi^+$ candidate, shows a narrow peak around the mass of the pion, which is fitted to obtain the yield of D^{*+} and consequently of B^0 (shown in fig. 2.12). The number of expected events is corrected to account for reconstruction efficiency, acceptance effects and branching ratios of the channel. The full B-Parking dataset corresponds to an integrated luminosity of 41.6 fb^{-1} , and recorded a total of $\sim 10^{10}$ unbiased b hadron decays [6].

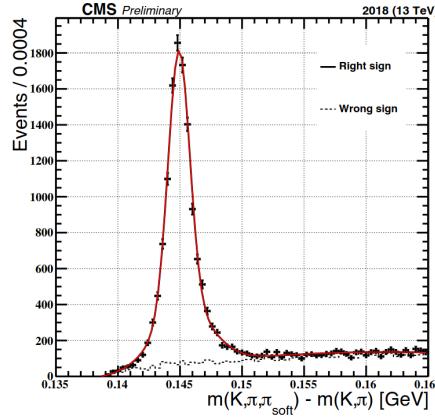


Figure 2.12: Invariant mass difference of the D^{*+} and D^0 candidates. Right sign data points refer to events where K and μ have same charge, while wrong sign data points refer to the events where they have opposite charge. The red solid line is the fit to the data.

2.4 CMS Event reconstruction

A particle travelling through the detector is expected to produce several signals in the various CMS subdetectors, depending on the nature of the particle.

- Charged particles give rise to signals (*hits*) in the sensitive layers of the tracker system, allowing reconstruction of their trajectory (*track*) and their origin (*vertex*);
- Electrons and photons are absorbed in the ECAL, where they produce *clusters* of energy, which allow measuring the particle's energy and direction;
- Charged and neutral hadrons may deposit a fraction of their energy in the ECAL, but should be primarily absorbed in the HCAL, which measures their energy;
- at the energies involved at the LHC, muons are practically minimum ionizing particles, and they lose only a small fraction of their energy when traversing the detector material. Muons are detected both in the tracker and in the muon system. Although they are not absorbed in the calorimeters, they might deposit there a small fraction of their energy;
- Neutrinos pass through the CMS detector with almost no interaction, their presence can only be inferred by reconstructing the transverse momentum imbalance in collision events.

The signals produced by the particles in the subdetectors are used to reconstruct various *physics objects* and measure their properties. The software used by CMS to reconstruct such objects is based on correlating the basic elements from all detector layers (tracks and clusters) to identify each final-state particle, and reconstructing the particle properties by combining the corresponding measurements on the basis of this identification. This approach is called *particle-flow* (PF) reconstruction [51].

2.4.1 Particle-flow elements

The reconstruction of the basic PF elements uses advanced algorithms specifically designed for each element. They are briefly described in the following sections, based on ref. [51].

Charged particle tracks

The goal of the track reconstruction is to find the parameters of the trajectory of a charged particle, given the hits in the tracker. The original track reconstruction algorithm used by CMS is the Combinatorial Track Finder (CTF), which is based on the Kalman filtering technique. The algorithm used by PF is a new iterative tracking method. The core idea is that in initial iterations the search is for tracks that are easiest to find (e.g. large p_T) and the hits associated with these tracks are removed for the subsequent iterations, reducing the combinatorial complexity and simplifying the search for the following steps.

In each iteration, the algorithm starts from seeds, which are 2 or 3 hits in the inner tracker that constrain the parameters of a track. The tracks provided by the seeds are then propagated to the following tracker layers, and the hits that are compatible with the track are added to the track. Multiple tracks with different combinations of hits are built in parallel, and the best combination is chosen by a quality estimator. For each track, the algorithm associates the hits and an estimate of the track parameters, and the track is refitted using the Kalman filter.

The tracks are categorized based on their expected purity according to: their hits on the pixel and tracker system, the track fit χ^2 and the longitudinal and transverse impact parameters (with associated uncertainties). Most analyses use *HighPurity* tracks, which are the tightest selection of tracks [52].

Primary vertex reconstruction

The goal of primary vertex reconstruction is to find the position of the primary vertices, i.e. the vertices of proton-proton interactions in each event, by using the information from the tracks.

First, a selection of tracks that are produced promptly after the collision is made, based on the number of hits in the pixel and tracker system, the χ^2 of the track fit, and the impact parameters. The selected tracks are clustered based on their point of closest approach to the center of the beam spot. The clusters with at least two tracks are then fitted with an adaptive vertex fitter, which computes the best estimate and associated uncertainties of x, y, z coordinates of the vertex.

Muon objects

The muon system of the CMS experiment allows muon identification with high efficiency over its acceptance range. The calorimeters before the muon system are helpful to absorb other particles and increase the purity of the muons that reach the muon system.

Three different high-level muon types are reconstructed:

- *standalone muon*: hits within the DTs or CSCs are clustered to form a track segment which is used as seed for a track reconstruction in the muon system only;
- *global muon*: standalone muons are matched with tracks in the inner tracker, if they can be propagated onto a common surface. All hits are combined and fitted to form a global muon track; for $p_T \gtrsim 200$ GeV the muon system hits improve the momentum resolution of these muons;
- *tracker muon*: inner tracks with $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV are matched with hits in the muon system. If at least one muon segment that matches the extrapolated track is found, and the difference between the x coordinate of the track and the segment below a certain threshold, the track is promoted to a tracker muon.

A set of selection criteria are applied to the reconstructed muon candidates. The cuts are based on variables coming from the fitting procedure, such as χ^2 , the number of hits and the quality of the matching between the muon system and the inner tracker. The cuts of the selection have different levels of purity and efficiency, that can be used by different analyses for their specific needs. These are referred to as muon identification (ID) criteria, and some commonly used ones are [53]:

- *Loose muon ID*: it efficiently selects muons prompt muons and muons coming from both light and heavy flavor hadrons;
- *Medium muon ID*: it is used to select prompt muons and muons from heavy flavor hadrons;
- *Tight muon ID*: it is a tighter selection designed to suppress signatures coming from decay in flight and hadronic punch-through;
- *Soft muon ID*: it is used to select muons with low p_T , especially for B-physics and quarkonia analyses;
- *High p_T muon ID*: it is used to efficiently select muons with $p_T > 200$ GeV.

In particular, the soft muon ID, which is a flag used in this work, has the following requirements:

- the muon is a tracker muon that satisfies the HighPurity flag, and has hits from at least six layers of the inner tracker, including at least one pixel hit;

- the pulls between the muon system segments and the track are required to be below 3 on the x and y axes;
- its transverse impact parameter is below 0.3 cm and its longitudinal impact parameter is below 20 cm.

Another important quantity associated to the reconstructed muon is the isolation of the muon, that gives a measure of the number of tracks surrounding the muon, allowing distinction between prompt muons and those from weak decays within jets. The isolation of a muon is computed as $\text{Iso}(\mu) = \sum p_T^{\text{track}}/p_T^\mu$ where the sum is on the tracks in a geometric cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ around the muon.

Calorimeter clusters

Calorimeter clusters have several purposes:

- detect and measure energy and direction of neutral particles such as photons or neutral hadrons;
- separate these neutral particles from deposits of charged particles;
- reconstruct and identify electrons and their showers;
- help in the energy measurement of charged hadrons determined by the tracker.

The clustering is performed separately in each subdetector: ECAL (barrel and endcap), HCAL (barrel and endcap) and preshower layers.

The clustering algorithm starts from cluster seeds, which are cells with energy above a certain threshold and with more energy than neighboring cells. The algorithm then forms *topological cluster* by aggregating cells that neighbor a cluster seed. The clusters within a topological cluster are reconstructed by fitting the energy deposits with Gaussian functions. This allows to measure the energy and position of the cluster.

For the energy measurement in the ECAL endcaps, the energy is a function of the measurement of the ECAL endcap itself and the two preshower layers. Similarly, the energy measurement in the HCAL is a function of the measurement of the HCAL and the ECAL.

2.4.2 Particle-flow reconstruction

The key feature of the PF algorithm is to put together information from multiple subdetectors to achieve the best possible reconstruction of the physics objects. Thus, a fundamental part of PF is the *link algorithm*, that connects the PF elements from different subsystems. This algorithm forms PF *blocks* which are groups of PF elements that have direct links or indirect links through other elements.

For each PF block, an identification and reconstruction sequence is performed. When a physics object is identified and reconstructed, the corresponding PF elements are removed from the block. The steps are the following:

- the muon candidates are identified and reconstructed;
- the electrons are identified and reconstructed, with a particular importance given to reconstructing their bremsstrahlung photons; in this same step isolated photon clusters are identified;
- the PF elements that remain at this step are subject to cross-identification of charged hadrons, neutral hadrons and photons.

When all blocks of an event have been processed, a post-processing step is performed that helps reduce misreconstruction and improve the performance of the algorithm.

2.5 Data and Monte Carlo processing workflow

To produce a simulated sample in CMS, the following steps are performed (further explained below the list):

1. generating the event (also known as ‘GEN’ step);

2. simulating the interaction of final state particles with the detector (also known as ‘SIM’ step);
3. simulating of the response of the detector’s electronics to the particles (also known as ‘DIGI’ step) and adding the contribution of pileup (also known as ‘PU’ mixing);
4. simulating the CMS trigger (L1 and HLT) based on the simulated response.

After reaching this step, the simulated sample has reached a format that is analogous to the ones produced by the real detector, and the following steps are the same for both data and simulation:

5. the CMS reconstruction software is run on the sample (also known as ‘RECO’ step), and the reconstructed events are saved in a format called Analysis Object Data (AOD);
6. the AOD format is refined and skimmed, to produce lighter formats that, while losing some information, are much more lightweight and easier to handle. This data formats are suitable for most CMS analyses.

Event generation

Monte Carlo event generators allow simulation of a wide range of physics processes at LHC [54]. The simulation of high energy physics events and their interactions with the detector is an essential part of the CMS physics program. Different Monte Carlo event generators are used to simulate different physics processes. For instance they are used to: simulate a signal of new physics over the SM background; be a guiding tool in the process of designing a new experiment; help in the choice of the best reconstruction techniques.

When simulating hadron-hadron collisions like the ones at the LHC, the basic steps of the physics process that are simulated are the following:

- the primary hard sub-process
- the parton showers associated with the particles involved in the primary hard sub-process;
- the hadronization of the partons into final state hadrons;
- the secondary interactions that can contribute to the underlying event;
- the unstable particles decays.

This processes results in a collection of final state particles, however a key additional step is simulation of pileup interactions. When a proton-proton collision occurs, the number of interactions in the same bunch crossing is not fixed, especially at the high luminosities of the LHC: in the 2018 run the average pileup per bunch crossing was ~ 30 [55].

Simulation of the detector response and trigger

After the generation of the final state particles, the next step is to simulate their interaction with the CMS detector e.g. the energy deposits in the calorimeters, the hits in the tracker, ... For the CMS detector, this is done with GEANT4 [56]. GEANT4 is a C++ object-oriented toolkit that uses information about the geometry, materials and other properties of CMS to simulate the passage of particles through the detector.

To simplify the computations, the other particles coming from pileup interactions are not simulated yet at this stage, but only added in the DIGI step. Other than the signal of interest produced on the detectors’ electronics, additional noise from the so-called ‘minimum bias’ events is added, that is meant to reproduce the real data-taking conditions caused by pileup.

At this point, the L1 and HLT triggers are simulated, and the reconstruction software is run on the sample of simulated events.

Chapter 3

Search for Heavy Neutral Leptons from D_s meson decays

As discussed in chapter 1, HNLs are theoretical particles that can be added to the SM with minimal assumptions. They are spin 1/2 fermions with no SM charge that only interact faintly with the SM particles through their coupling with the active left-handed neutrinos. The two relevant parameters of a HNL are its mass m_N and its mixing matrix elements with the active neutrinos V_α , where $\alpha = e, \mu, \tau$. A wide range of different experiments have provided constraints on the parameter space $m_N - |V_\alpha|$ for HNLs. Ref. [5] contains a review of the experimental searches and limits on HNLs. Currently, there are no published results on HNL searches from charmed meson decays at LHC. The work presented in this chapter refers to a HNL search from D_s meson decays in the CMS experiment.

The chapter is organized as follows: in section 3.1 we present the overall analysis strategy; in section 3.2 we present the data and MC samples used in the analysis; in section 3.3 we describe the event selection and in section 3.4 we briefly describe the procedure to extract the expected limits on HNL production.

3.1 Analysis strategy

The most promising channel for HNL search from charmed mesons, as discussed in section 1.3.1, are two-body decays of D_s mesons, because of the higher branching ratio and the relatively wide range of m_N masses that can be studied with it, which is below the D_s mass ($\lesssim 2$ GeV).

The specific channel considered is the $D_s^+ \rightarrow N\mu^+$ decay with the subsequent $N \rightarrow \mu^+\pi^-$, where N is the HNL. Thus, the final state is fully visible and consists of two muons and a pion. Figure 3.1 show a Feynman diagram of the decay under study.

This channel has the following advantages:

- low p_T muons from the D_s decay can be selected profiting from the B-Parking displaced muon triggers (described in section 2.3.2) with low p_T thresholds that allowed collection of a large number of events;
- the excellent muon reconstruction and identification performance of the CMS detector can be exploited;

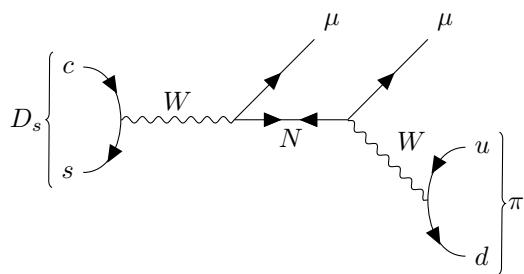


Figure 3.1: Feynman diagram for $D_s \rightarrow \pi \mu \mu$ decay via Majorana neutrino labeled N .

- the secondary vertex formed by the HNL decay products (μ and π) can be reconstructed, allowing to search for HNL decaying within the CMS tracker volume (< 1 m), and the optimal resolution of the $\mu\pi$ invariant mass provides a narrow peak for bump hunting;
- the D_s decays can be fully reconstructed, allowing to put a cut on the reconstructed D_s mass and reject combinatorial background.

The main form of background for this analysis is the combinatorial background, coming from accidental associations of muons and pions into a common vertex.

The displaced vertex of the HNL decay can be used as a signature to reject the combinatorial background but, at the same time, this can be a limiting factor for the analysis, because if the HNL has a large proper decay length, the displaced vertex can be outside the tracker acceptance and thus the HNL decay cannot be reconstructed.

The optimization of the signal selection is performed using MC samples: a cut-based selection is applied to efficiently select MC signal events and reject MC background. On the other hand, a data-driven method is used to make an estimate of the background yield while avoiding potential MC mis-modeling: for each HNL mass point the number of expected background events is estimated by fitting the $\mu\pi$ invariant mass distribution in the “sidebands” i.e. the region outside a window of interest centered around the HNL mass peak.

The analysis doesn’t focus on a single signal hypothesis and takes into account four different values of HNL mass: 1.0, 1.25, 1.5 and 1.8 GeV and three different values of proper mean decay length $c\tau$ (also referred to as “lifetime” for short): 10, 100 and 1000 mm. Because of the proportionality $c\tau \sim m_N^{-5} |V_\mu|^{-2}$ (see section 1.5), considering larger values of $c\tau$ allows exploring smaller values of the mixing parameter $|V_\mu|$.

For all the HNL mass points, the number of signal events is extracted from the data using an unbinned maximum likelihood (UML) fit of the HNL invariant mass. A limit on the value of $|V_\mu|^2$ is set with asymptotic approximations [57]. This limit assumes that the mixing with the other active neutrinos, $|V_e|^2$ and $|V_\tau|^2$, is zero. To avoid biases, at this stage, the analysis uses “data blinding”: the optimization studies are performed on data excluding events compatible with the HNL mass.

3.1.1 Signal yield

To estimate the number of expected signal events a reference “normalization channel” is used. The normalization channel is $D_s^+ \rightarrow \phi (\rightarrow \mu^+ \mu^-) \pi^+$. The expected signal yield in the signal channel $N_{D_s^+ \rightarrow N\mu^+}$ is:

$$N_{D_s^+ \rightarrow N\mu^+} = L_{\text{int}} \sigma(D_s) \mathcal{B}(D_s^+ \rightarrow N\mu^+) \mathcal{B}(N \rightarrow \mu^+ \pi^-) \varepsilon_{D_s^+ \rightarrow N\mu^+}, \quad (3.1)$$

and the number of events in the normalization channel $N_{D_s^+ \rightarrow \phi\mu^+}$ is:

$$N_{D_s^+ \rightarrow \phi\mu^+} = L_{\text{int}} \sigma(D_s) \mathcal{B}(D_s^+ \rightarrow \phi\pi^+) \mathcal{B}(\phi \rightarrow \mu^+ \mu^-) \varepsilon_{D_s^+ \rightarrow \phi\mu^+}, \quad (3.2)$$

where in both equations L_{int} is the integrated luminosity, $\sigma(D_s)$ is the production cross-section of D_s mesons, \mathcal{B} is the branching ratio of the decay and ε is the total selection efficiency. All these quantities can be measured or estimated:

- the efficiencies can be computed from MC samples, by comparing the number of generated events with those that pass the final selection;
- the branching ratios for the normalization channel i.e. $\mathcal{B}(D_s^+ \rightarrow \phi\pi^+)$ and $\mathcal{B}(\phi \rightarrow \mu^+ \mu^-)$ are well known and reported by the PDG [13];
- the branching ratios for the signal channel i.e. $\mathcal{B}(D_s^+ \rightarrow N\mu^+)$ and $\mathcal{B}(N \rightarrow \mu^+ \pi^-)$ come from the theoretical model as described in section 1.5;
- The number of D_s in the normalization channel can be measured from data.

Thus, the signal yield can be computed by combining eqs. (3.1) and (3.2) into:

$$N_{D_s^+ \rightarrow N\mu^+} = N_{D_s^+ \rightarrow \phi\mu^+} \frac{\mathcal{B}(D_s^+ \rightarrow N\mu^+) \mathcal{B}(N \rightarrow \mu^+ \pi^-) \varepsilon_{D_s^+ \rightarrow N\mu^+}}{\mathcal{B}(D_s^+ \rightarrow \phi\pi^+) \mathcal{B}(\phi \rightarrow \mu^+ \mu^-) \varepsilon_{D_s^+ \rightarrow \phi\mu^+}}. \quad (3.3)$$

Measuring the yield with respect to the normalization channel allows the cancellation of systematic uncertainties on the luminosity, the production cross-section, and partially on the selection efficiencies having the same particles in the final state.

3.2 Data and Monte Carlo samples

The analysis uses data from proton-proton collisions collected at the CMS experiment with $\sqrt{s} = 13$ TeV during the 2018 data-taking period with the B-Parking trigger. Monte-Carlo (MC) samples are used for signal selection optimization, background modeling, measuring acceptance and efficiency and estimating the systematic uncertainties.

3.2.1 B-parking dataset

Events are selected with the B-parking trigger (described in section 2.3.2), which is designed to select events with the presence of at least one displaced muon. The Level 1 (L1) trigger requires the presence of either a low p_T muon with pseudorapidity $|\eta| < 1.5$ (the specific threshold on p_T changes based on the instantaneous luminosity) or a higher p_T muon ($p_T > 22$ GeV) without pseudorapidity restrictions. Then, at the HLT level the muon is required to have, in addition to the p_T threshold, an impact parameter significance (IPS) above a certain threshold (which again changes based on the instantaneous luminosity).

The p_T and IPS thresholds range from 7 to 12 GeV and from 3.5 to 6 respectively. Table 3.1 shows a summary of all HLT paths and the integrated luminosity collected by each of them.

HLT path	Total Integrated Luminosity (fb $^{-1}$)
HLT_Mu7_IP4	6.940
HLT_Mu8_IP3	1.583
HLT_Mu8_IP5	8.259
HLT_Mu8_IP6	8.259
HLT_Mu8p5_IP3p5	0.320
HLT_Mu9_IP4	0.010
HLT_Mu9_IP5	20.89
HLT_Mu9_IP6	33.669
HLT_Mu10p5_IP3p5	0.320
HLT_Mu12_IP6	34.791

Table 3.1: Total integrated luminosity for each B-Parking HLT path. The HLT paths are indicated as HLT_MuX_IPY where X is the p_T threshold and Y is the IPS threshold.

To compute the total integrated luminosity we must take into account the fact that some triggers paths are active at the same time and thus share some luminosity sections. The total luminosity collected with the B-parking trigger, after excluding overlaps, is 41.6 fb $^{-1}$.

3.2.2 Monte-Carlo simulated sample

The signal samples are generated with four values of HNL mass m_N : 1.0, 1.25, 1.5 and 1.8 GeV, and three values of the mean proper decay length $c\tau$: 10, 100 and 1000 mm. The D_s in the generated sample is inclusive i.e. it contains both prompt generated D_s mesons and D_s mesons produced from in-flight decays of heavier hadrons. The decay $D_s^+ \rightarrow \phi\pi^+$ with subsequent $\phi \rightarrow \mu^+\mu^-$ is used to measure the acceptance and efficiency of the normalization channel. These MC samples are produced with PYTHIA 8.240 [58] which generates the hard event and handles both the hadronization of the partons and the particles decay, except for the decay of interest, which is handled by the EVTGEN 1.6.0 software [59]. The $D_s \rightarrow N\mu^+$ and subsequent $N \rightarrow \mu^+\pi^-$ are generated assuming no spin correlation between the particles involved.

The background from SM is simulated by generating events with jets in the final states, which we will refer to as QCD background events. The event generation for the QCD background is divided into different samples, each corresponding to a different range of the transverse momentum of the hard process, \hat{p}_T . This is done because the production cross-section depends on \hat{p}_T , and the event generation is more efficient at low \hat{p}_T . The range of \hat{p}_T varies from 20 to 300 GeV. The QCD events are then weighted in a way that is inversely proportional to the cross-section of the specific \hat{p}_T range.

These QCD samples are also “muon enriched” i.e. the events are required to have at least one muon with $p_T > 5$ GeV and $|\eta| < 2.5$. The background samples have also been generated with PYTHIA.

While the QCD MC is used to model the distribution of the combinatorial background in the optimization of the signal selection requirements, the background yield is extracted directly from data, avoiding any MC mismodelling.

Additional requirements, at the generated particle level, are applied for both signal and background MC samples to the final states particles p_T and η to ensure that they are within detector acceptance, meet trigger conditions and meet the kinematic requirements for the decays under study (signal channel or normalization channel).

Details on these additional requirements and the generator settings for the MC samples are listed in appendix A.

3.3 Event Selection

Before finding the best selection for the signal, a loose pre-selection (section 3.3.1) is applied. The MC samples must be corrected for potential differences between data and simulation (section 3.3.2). After this, the signal selection is optimized using the MC samples (section 3.3.3).

3.3.1 Pre-selection

The D_s candidates of the analysis are built from three final states objects: two muons and a pion. A set of loose selection criteria is applied to the final state reconstructed objects, to reduce mis-reconstructed candidates coming from background. In what follows we list: the pre-selection requirements for the individual muons and pions, the trigger matching criteria, the signal candidate pre-selection and the selection for the normalization channel.

Muon pre-selection

The reconstructed muons are required to pass the following conditions:

- (i) $p_T > 3.5(2.0)$ GeV for muons with $|\eta| < 1.2(2.4)$; these are meant to ensure that the muons can reach the muon chambers
- (ii) pass *Soft muon* ID [53]; this is a typical muon identification used by CMS in its B-physics analyses.

Pion pre-selection

Among all candidates reconstructed by the PF algorithm (description in section 2.4.2), the objects that are considered pion candidates must meet the following requirements:

- (i) the track passes the *highPurity* [52] condition;
- (ii) the PDG ID assigned by the reconstruction is a charged hadron;
- (iii) $p_T > 0.5$ GeV and $|\eta| < 2.4$.

Trigger matching

In both data and MC, at least one of the muons must be the one that fired the B-Parking trigger. We consider a muon to match a trigger object by considering the track separation $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$ and the p_T : the muon and the trigger object must have $\Delta R < 0.05$ and $\Delta p_T/p_T < 0.1$.

Signal candidate pre-selection

Other than the pre-selection applied to the individual muons and the pions, the signal candidates have to satisfy the following requirements:

- the muon and the pion forming the HNL candidate are required to:
 - have total $p_T > 1.0$ GeV;
 - have opposite electric charge;
 - form a common vertex with a fit probability greater than 0.01;
 - have an invariant mass $0.2 < m_{\mu\pi} < 2.0$ GeV.

- Once a HNL candidate is found, it is combined with an additional muon and fitted to a common vertex to form a D_s candidate:
 - the invariant mass of the $\mu\mu\pi$ system must be between 1.5 and 2.5 GeV, so that it is compatible with the D_s mass;
 - the vertex fit probability of the D_s candidate must be greater than 0.01.

Normalization channel selection

The D_s candidates for the normalization channel $D_s^+ \rightarrow \phi(\rightarrow \mu^+\mu^-)\pi^+$ are built from the same final state objects as the signal channel, and have to satisfy the following requirements:

- the two muons forming the ϕ candidate are required to:
 - have an invariant mass within 0.05 GeV of the known ϕ mass (taken from PDG [13]);
 - have a fit probability of the common vertex greater than 0.01;
- for each good $\phi \rightarrow \mu^+\mu^-$ candidate, the muons are combined with a π candidate to form a D_s candidate, which is required to:
 - have an invariant mass between 1.75 and 2.15 GeV;
 - have a fit probability of the common vertex greater than 0.01.

3.3.2 Correction factors

A number of correction factors are applied to the MC samples to account for differences between data and simulation:

- pileup distribution: the number of pileup interactions in the MC samples does not match the number of pileup interactions in the data. To correct for this, the MC samples are re-weighted so that the distribution of the number of pileup interactions are the same in data and MC;
- muon reconstruction and identification efficiency: the MC simulation does not reproduce perfectly the muon reconstruction and identification efficiencies of the real detector. To correct for this, the MC samples are re-weighted so that the efficiency of the muon reconstruction and identification is the same in data and MC;
- trigger efficiency: several single muon HLT paths were used to collect the B-Parking data and during the data taking not all the triggers were simultaneously on. The time evolution of the trigger paths activation is not reproduced in MC samples. Correction factors are computed and applied in order to correct the MC samples to the complex trigger composition of the data.

After applying these corrections, the MC simulated samples reproduce the data reasonably well, with some small differences still present, that will be accounted for as systematic uncertainties (see section 3.4.4).

Figure 3.2 shows the comparison between data and MC for the p_T , η and IPS of the two muons and pion. The muon p_T distributions show several structures: one effect is due to the selection cut requirements based on the muon η , i.e. muon $p_T > 3.5$ GeV for $|\eta| < 1.2$ and $p_T > 2$ GeV for $1.2 < |\eta| < 2.4$; other effects for $p_T > 7$ GeV and a small bump around 12 GeV are due to the additional trigger thresholds on the muon p_T . The IPS distribution shows low values and falls off rather quickly.

3.3.3 Signal selection

After the pre-selection, the signal selection is optimized using the MC samples.

Since the search is not focused on a single kind of signature and different HNL masses or lifetime hypotheses can have different signatures, the selection is optimized in different event categories. The categorization is based on:

- relative muon sign: same sign muons or opposite sign muons;
- HNL decay length on the transverse plane (L_{xy}): $0 < L_{xy} < 1$ cm, $1 < L_{xy} < 5$ cm and $L_{xy} > 5$ cm,

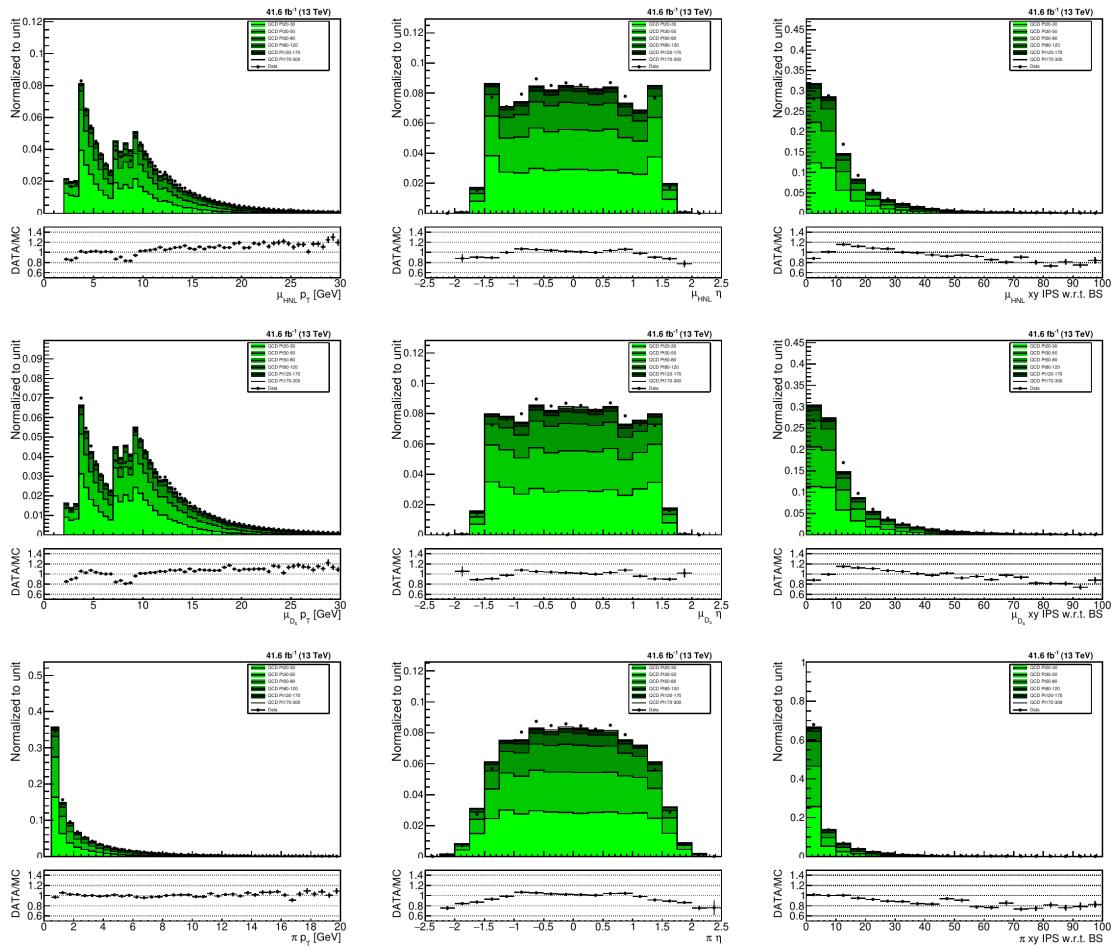


Figure 3.2: Data/MC comparison for p_T (left column), η (middle column) and IPS (right column) for the muon forming the $D_s^+ \rightarrow N\mu^+$ (top row) and the $N \rightarrow \mu\pi$ (middle row) candidate, and for the pion (bottom row). The data corresponds to the full B-Parking data set.

for a total of six categories.

The selection is based on choosing the best set of cuts on the variables which show a significant separation power between signal and background: this selection is optimized by maximizing the expected signal significance, defined as [60]:

$$Z_0 = \sqrt{2 \left[(S + B) \ln \left(1 + \frac{S}{B} \right) - S \right]}, \quad (3.4)$$

where S and B are the expected yields of signal and background events. This reduces to $S/\sqrt{S+B}$ in the limit of $S \ll B$.

The expected signal yield is computed with eq. (3.3) assuming a fixed mass of $m_N = 1.5$ GeV and a fixed mean proper decay length of $c\tau = 10$ mm.

The expected background yield is the number of background events in the signal region, which is a region of $\pm 3 \sigma_{\text{HNL}}$ around the HNL mass peak, where σ_{HNL} is the width of the HNL mass peak, extracted from a fit to the HNL mass distribution in the MC samples. A different value of σ_{HNL} is extracted for each signal mass hypothesis. To estimate the background yield, the parameters of the exponential distribution are computed with an exponential fit on the $\mu\pi$ invariant mass distribution in the “sidebands” of the HNL mass distribution in the data, i.e. a $\pm 6 \sigma_{\text{HNL}}$ region outside the signal region, and then the background yield is computed by integrating the exponential distribution in the signal region.

The optimization is performed by scanning the cut values on the following variables:

- $L_{xy}/\sigma_{L_{xy}}$: the HNL vertex displacement significance i.e. the HNL decay length divided by its uncertainty;
- $\text{IP}(\pi), \text{IP}(\mu)$: the impact parameter (IP) on the transverse plane, measured with respect to the beam spot (BS) for the pion and the muon forming the HNL candidate;
- $\text{IPS}(\pi), \text{IPS}(\mu)$: the impact parameter significance (IPS) on the transverse plane, i.e. the impact parameter divided by its uncertainty, for the pion and the muon forming the HNL candidate;

The following algorithm is used to find the best set of cuts in each category:

1. the significance is tested separately for each variable by varying the value of the selection cut;
2. the variable giving the largest significance gain is selected and the optimized selection cut is applied;
3. the variable selected in the previous step is excluded;
4. the procedure starts over.

Following this algorithm, the final set of variables used for the signal selection are $L_{xy}/\sigma_{L_{xy}}$, $\text{IP}(\pi)$ and $\text{IPS}(\mu)$. Any further steps in the algorithm provide no further gain in significance.

The final choice on the cut values defines a working point for the selection, with a fixed signal and background efficiency. The signal efficiency is defined as the fraction of signal events that pass the selection. Instead of the background efficiency, we can refer to the background rejection defined as the fraction of background events that are rejected by the selection.

Table 3.2 shows the working points for all the signal mass hypotheses with a fixed mean proper decay length of $c\tau = 10$ mm.

Signal mass hypothesis (GeV)	Signal efficiency	Background rejection
1.0	54.7 %	
1.25	62.8 %	
1.5	67.9 %	97.9 %
1.8	62.9 %	

Table 3.2: Signal efficiency and background rejection for the working points of the signal selection for each mass hypothesis with fixed $c\tau = 10$ mm.

There is a general trend of increasing signal efficiency with increasing mass, except for the 1.8 GeV mass hypothesis. This is because this is a peculiar limit case, very close to maximum

allowed mass of the decay (the mass of the D_s meson), which makes the HNL decay products very soft and thus harder to reconstruct.

The primary emphasis of this thesis work is improving these selection efficiencies. This involves the use of Machine Learning techniques to optimize the process of signal selection. The details of these techniques, their application and resulting selection efficiencies are discussed in chapter 4.

3.4 Limits extraction

The analysis consists of a search for HNLs of unknown mass and lifetime. A scan in the displaced muon and pion invariant mass spectrum is performed, where, in each window centered around a mass hypothesis, signals of different lifetime are searched over the background. The signal process we are looking for is rare, and the expected signal yield is very low. Thus, instead of looking for a significant excess of events, we aim to set an upper limit. To obtain the results, a simultaneous fit of signal and background to the $\mu\pi$ invariant mass spectrum is performed to determine the signal strength.

3.4.1 Lifetime reweighting

The simulated MC signal samples are generated only for fixed values of $c\tau$ and m_N . In order to have a fine scan of the mean proper decay length of the HNL, the simulated samples are reweighted to cover the $c\tau$ spectrum ranging from 10 to 1000 mm. To reweight a simulated sample with lifetime $c\tau_0$ to a target lifetime $c\tau_1$, the events must be multiplied by the following weight:

$$w(ct, \tau_0 \rightarrow \tau_1) = \frac{\frac{1}{c\tau_1} \exp\left(-\frac{ct}{c\tau_1}\right)}{\frac{1}{c\tau_0} \exp\left(-\frac{ct}{c\tau_0}\right)}, \quad (3.5)$$

where ct is the generator-level proper lifetime of the HNL in the event, calculated as:

$$ct = \frac{LM}{p}, \quad (3.6)$$

with L being the decay length, M the mass and p the momentum at generator level.

The signal selection described in section 3.3.3 is optimized for a fixed $c\tau = 10$ mm, but this reweighting procedure allows to extract the corresponding efficiencies for the other $c\tau$ values.

3.4.2 Normalization channel yield

The estimated yield in the signal channel can be computed with eq. (3.3) once the expected yield in the normalization channel $D_s^+ \rightarrow \phi(\rightarrow \mu^+\mu^-)\pi^+$ is known. The expected yield in the normalization channel is computed with an unbinned maximum likelihood fit of the reconstructed $\mu\mu\pi$ invariant mass distribution for the D_s candidates of the normalization channel.

Figure 3.2 shows the $\mu\mu\pi$ invariant mass distribution of the normalization channel candidates, with the corresponding fit.

3.4.3 Signal and background fit

The limit extraction method requires the number of signal and background events, which can be extracted from a binned maximum likelihood fit of the $\mu\pi$ invariant mass distribution for the muon and pion forming the HNL candidate.

The signal region is defined as a window of $\pm 3\sigma_{\text{HNL}}$ around the HNL mass peak. The sideband region is defined as a window of $\pm 6\sigma_{\text{HNL}}$ outside the signal region. σ_{HNL} is the width of the HNL mass peak, extracted from a Voigtian fit of the HNL mass distribution in the signal MC samples, for each mass hypothesis. Figure 3.4 shows as illustrative example the $\mu\pi$ invariant mass distribution for a signal MC sample with $m_N = 1.0$ GeV and $c\tau = 10$ mm, in the category of same sign muons with $0 < L_{xy} < 1$ cm.

Thus, the signal and background events in the signal region can be extracted with a simultaneous fit of the signal and background probability density functions (PDFs) to the $\mu\pi$ invariant mass distribution. The signal is modeled with a Voigtian PDF and the background with an

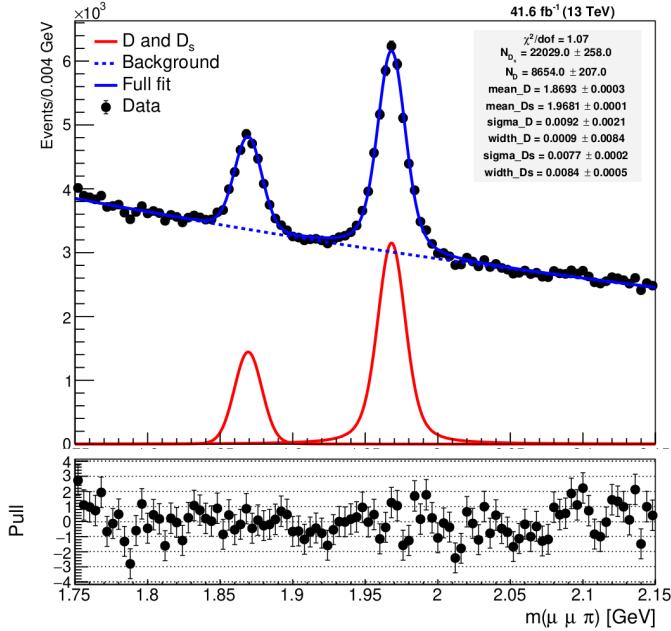


Figure 3.3: The $\mu\mu\pi$ invariant mass distribution with fits of the D and D_s peaks and the background. The bottom panel shows the pulls of the data with respect to the fit curve.

exponential PDF. However, since the analysis is blinded at this stage, the observed exclusion limits cannot be extracted yet, but the number of signal and background events can be estimated to get an expected limit.

To estimate the number of background events in the signal region, the parameters of the exponential are computed with an exponential fit on the $\mu\pi$ invariant mass distribution in the sidebands of the HNL mass distribution in the data, and then the background yield is computed by integrating the exponential distribution in the signal region.

As an illustrative example, fig. 3.5 shows the $\mu\pi$ invariant mass distribution centered around 1.0 GeV, for the category of same sign muons with $0 < L_{xy} < 1$ cm. At this stage of the analysis the signal region is blinded.

3.4.4 Systematic uncertainties

Before describing the procedure to extracting the estimated limits, we briefly discuss the systematic uncertainties that affect the analysis. The estimated yield of the signal channel is affected by the following sources:

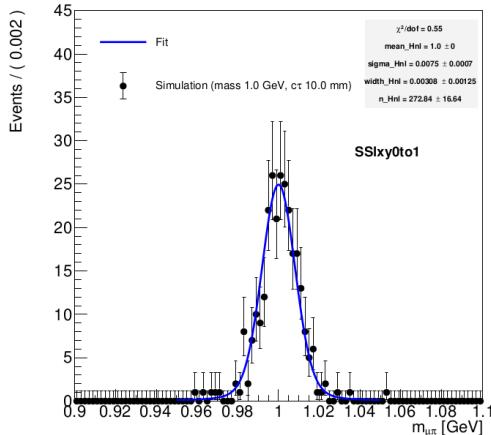


Figure 3.4: The $\mu\pi$ invariant mass distribution for MC generated signal with $m_N = 1.0$ GeV and $c\tau = 10$ mm, with signal fit. The data corresponds to the category of same sign muons with $0 < L_{xy} < 1$ cm.

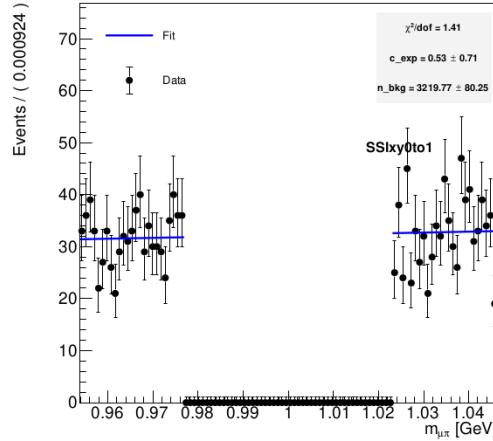


Figure 3.5: The $\mu\pi$ invariant mass distribution centered around 1.0 GeV, with the exponential fit in the sidebands of the HNL mass distribution to estimate the background yield in the signal region. The signal region is blinded. The data corresponds to the category of same sign muons with $0 < L_{xy} < 1$ cm.

- normalization channel yield:

the uncertainty on the normalization channel yield is estimated by using alternative functions to fit the peak of the D_s and D mesons: a Gaussian and a Crystal Ball function. The background is still fitted with an exponential function in all cases. The systematic error is taken as the maximum relative difference between the yields obtained with the Voigtian compared to the yields obtained with the alternative functions, which is 10%;

an additional uncertainty is assigned to the normalization channel yield to account for the difference in the yield based on using a fixed or free mass for the D_s mass in the fit, which is 0.2%;

- data and MC simulation discrepancies:

even after applying the corrections described in section 3.3.2, there are residual differences between data and MC. A re-weighting of the MC samples is performed to correct for these differences. The systematic uncertainty is estimated by comparing the expected signal yield with and without the re-weighting. The uncertainties vary depending on the event category and on the reweighting variable and range between few % to 15%;

- vertex reconstruction efficiency:

the efficiency of the vertex reconstruction as a function of the displacement and p_T is estimated by comparing the number of reconstructed vertices for K_s^0 decays to charged particles in data and MC [61]. The maximum difference in the efficiency is taken as the systematic uncertainty, which is 7%;

- muon ID/reconstruction efficiency:

the correction factors for muon ID and reconstruction have an associated uncertainty. The systematic uncertainty on the yield is estimated by varying the correction factors within their uncertainties. It ranges between 0.1% and 1% for different event categories;

The yield of the background is affected by the choice of the PDF used to model the background. The systematic uncertainty on the background yield is estimated by using alternative functions to fit the background: a power law and a first order polynomial. The systematic error is taken as the maximum relative difference between the yields obtained with the exponential function compared to the yields obtained with the alternative functions. It ranges between 0.1% and 7% for different event categories.

Table 3.3 shows a summary of the systematic uncertainties that have been mentioned and their estimated values.

Source	Systematic uncertainty (%)
$D_s^+ \rightarrow \phi\pi^+$ PDF	10
$D_s^+ \rightarrow \phi\pi^+$ fixed/floating mass	0.2
Data/MC agreement	1-15
Vertex reconstruction efficiency	7
Muon ID/reconstruction efficiency	0.1-1
Background PDF	0.1-7

Table 3.3: Sources of systematic uncertainty and their estimated values.

3.4.5 Expected limit extraction procedure

The expected background yield can be extracted from data by fitting the HNL mass sidebands (see section 3.4.3), while the number of expected signal events can be computed with eq. (3.3). The systematics described in section 3.4.4 are taken as inputs to the fit. The MC reweighting described in section 3.4.1 can be used to span over the $c\tau$ spectrum from 10 to 1000 mm.

Since the analysis is blinded, the observed exclusion limits cannot be extracted. However, the expected limits is extracted with the CLs method, using as test statistic a binned profile likelihood ratio in the asymptotic approximation [57].

Chapter 4

Machine Learning algorithms for the search for Heavy Neutral Leptons

One of the margins of improvement in the search for Heavy Neutral Leptons (HNL) seen in chapter 3 is the efficiency of the selection of signal events. Thus, we attempt a different approach, compared to the current cut-based analysis, exploiting different Machine Learning (ML) models and evaluating the possible improvement in the selection efficiency [7].

The chapter is organized as follows: section 4.1 is an introduction to the basic concepts of ML models that are relevant for this work; in section 4.2 we describe the topological characteristics of the signal channel under study and the variables used for the training of the ML models; section 4.3 explains the data preparation, the hyperparameter configuration of the ML models and the monitoring of the training process; in section 4.4 we present the results of the training: the choice of the best models, which physical variables are the most important for the discrimination of signal from background, and the performance of the chosen models. Finally, we conclude with section 4.5 where we summarize the results and the future prospects of this work.

4.1 Brief introduction to Machine Learning

The ML models we focus on in this work are *supervised classifiers*.

The input data is provided to the model in a table-like format: the *input features* are the columns of the table, and they represent the characteristics of each data point, while the rows of the table are the *events*, and they represent the individual data points. For each event there is a *label* that represents the true class of the event. The purpose of the model is to learn the relationship between the input data and the labels, so that it can make predictions on new data for which the label is unknown [62].

In our case the input features are the physical variables that describe the particles in the event, and the label is either signal or background, which makes this a binary classification problem.

The main goal typically consists in finding the optimal parameters of the model that minimize a so-called *loss function*. A loss function (also known as cost or objective function) is a measure of the difference between the model's predictions and the labels. Thus minimizing the loss function means that the model's predictions are as close as possible to the labels.

The training process involves two main steps: *forward propagation* and *backward propagation*. In the forward propagation, the input data is passed through the model to generate predictions. These predictions are then used to compute the loss function. Then, in the backward propagation, the gradient of the loss function with respect to the model's internal parameters is computed. This gradient is then used to update the parameters in a way that minimizes the loss function. This process is repeated for a fixed number of iterations, or until a stopping criterion is met [63].

Other than the models' internal parameters, the *hyperparameters* [64] of the models are configurable settings that are set before the training process. They define the learning process

and the architecture of the model, for instance: the specific choice of loss function, the number of trees in an ensemble for a decision tree model or the number of nodes/layer in an artificial neural network.

A common problem with ML is *overtraining* (or overfitting). This particular issue occurs when a trained model matches the data used for training very precisely, thus achieving a high accuracy, but lacks in generalization capability on new, previously unseen data (see as an illustrative example fig. 4.1). To keep this issue under control, the input data is divided into three sets: the *training set*, the *validation set*, and the *test set*. The training set is used to train the model, the validation set is used to monitor the performance of the model during the training process, and the test set is used to evaluate the performance of the model after the training process is complete [63].

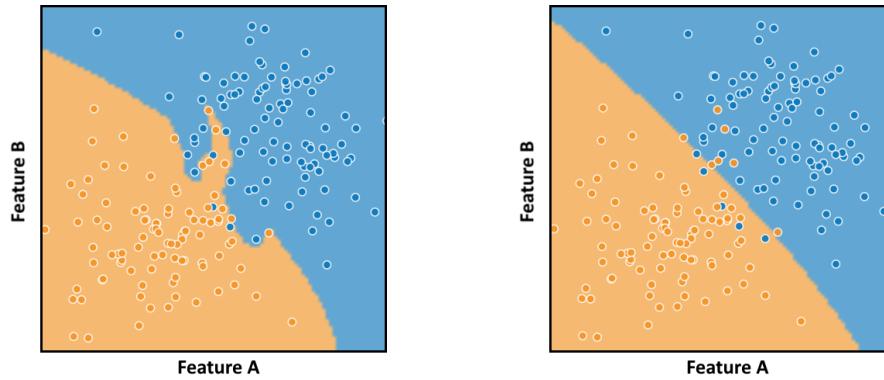


Figure 4.1: Example of overtraining (left) and no overtraining (right). The dots represent the training data, blue and orange for the two classes. The colored areas represent the decision boundary of the model. In the left plot, the decision boundary is very complex and follows the training data very closely, while in the right plot the decision boundary is simpler. Images are generated with Tensorflow playground [65].

A technique used in this work to mitigate overtraining is called *early stopping*. A metric, called *early stopping metric*, is evaluated on the validation set at each iteration of the training. In this work, we use a loss function as the early stopping metric, and thus refer to it as *validation loss*. When the validation loss stops improving for a number of iterations (called *patience*), the training is stopped. The model at the iteration with the lowest validation loss (typically referred to as *early stopping point*) is then chosen as the final model [66].

Logistic loss

As previously mentioned, the class of ML models we focus on are binary classifiers, i.e. they are trained to distinguish between two classes of events, in our case signal and background. One commonly used loss function in binary classification problems is the *Logistic loss* or log-loss for short [63]. The log-loss function is defined as:

$$L_{\log}(y, p) = -[-y \log(p) - (1 - y) \log(1 - p)] \quad (4.1)$$

where $y \in \{0, 1\}$ is the true class label, and $p \in [0, 1]$ is the model's predicted probability that the event belongs to the class 1. It is particularly suitable for binary classification problems because it penalizes strongly erroneous predictions e.g. if the true class label is 1 and the predicted probability is 0, the loss is negative infinity. Due to its suitability for binary classification problems, the log-loss is used as the loss function and early stopping metric for the training of all ML models in this work.

4.1.1 Boosted Decision Trees

Decision trees are a simple type of ML classifier that work by recursively splitting the input feature space into regions, and assigning a class label to each region. The regions are defined by a series of binary decisions, based on the values of the input features. The decision tree is built

by iteratively choosing the input feature and the threshold that best splits the input feature space, based on a criterion that measures the purity of the regions [63]. The specific criterion used to measure the purity can be different based on the specific implementation of the decision tree algorithm.

A key hyperparameter of decision trees is the *maximum depth* of the tree, which is the maximum number of binary decisions that can be made before a class label is assigned. Figure 4.2 shows an example of a simple decision tree.

Boosted Decision Trees, in particular, are a type of ensemble learning method, which is very commonly used in the field of HEP [67]. The idea of boosting is to combine several decision trees, that alone have a relatively low predictive power, to create a new classifier, called ensemble, with a significantly improved predictive power.

One of the first popular examples of boosting algorithms for ensembles of decision trees is the *AdaBoost* (which stands for Adaptive Boosting) algorithm [68]. The key idea of AdaBoost is to train a weak learner on the dataset, then re-weight the dataset in order to give more importance to the misclassified events, and then train another weak learner on these re-weighted dataset. This process is repeated for a fixed number of iterations, and the weak learners are combined to form a stronger learner, in a way that gives more importance to the weak learners that perform better.

Another class of boosting techniques is *Gradient Boosting* [69]. There are several implementations of gradient boosting, but the core idea is that at each iteration of the training, new weak learners are trained to minimize the gradient of the loss function. One popular implementation of gradient boosting is the *XGBoost* (short for eXtreme Gradient Boosting) algorithm [70]. This specific gradient boosting algorithm uses second-order gradients to improve the convergence of the training process, and it is well known for its overall performance and stability during training.

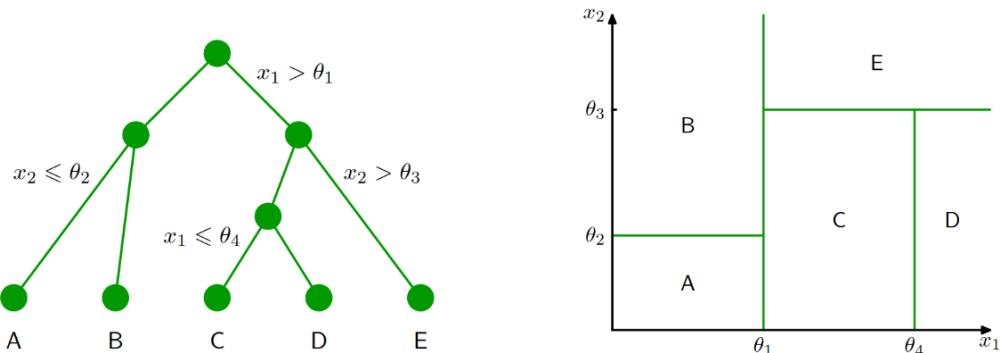


Figure 4.2: Example of a simple decision tree (left) and corresponding two-dimensional input feature space partitioning (right). x_i are the input features and θ_j are the splitting thresholds. This tree has a maximum depth of 3.

4.1.2 Artificial Neural Networks

Artificial Neural Networks are a class of ML models that are based on a series of interconnected nodes, called *neurons*, that are organized in layers: the input layer, composed by the entirety of input features, the output layer for the output score computation, and the hidden layers in between, for increasing the complexity of the parameter matrix. The specific choice of arrangement of the nodes and layers is called the architecture of the neural network, and a typical choice, which we consider in this work, is a fully connected architecture, where each node in a layer is connected to each node in the next layer. Each node receives as input the weighted sum of the outputs of the nodes in the previous layer, applies an activation function to this sum, and then passes the result to the nodes in the next layer. The activation function is a non-linear function, and it is what allows the neural network to learn non-linear patterns from the input data [62].

The architecture of the ANN models we use is the following (shown in fig. 4.3):

- Input layer: a normalization layer, which scales the input features to have zero mean and unit variance; this is a typical preprocessing step for neural networks [71];

- Hidden layer: a single layer with 64 nodes and a ReLU activation function [72] to allow non-linear output capabilities;
- Output layer: single layer with one node and a sigmoid activation function, which is a common activation function for binary classification problems, because it is defined in a fixed range [0,1], and prevents jumps in the output values by being always differentiable.

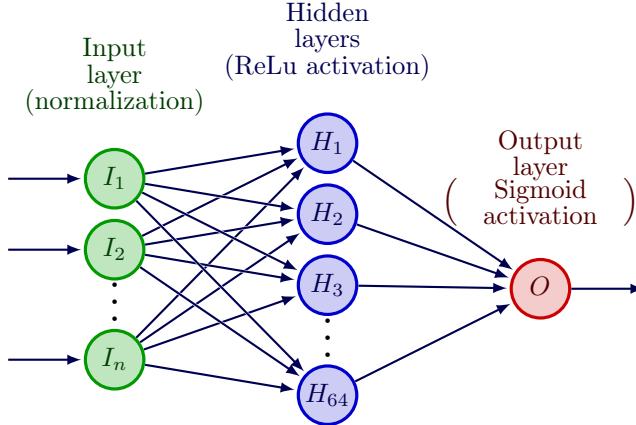


Figure 4.3: Architecture used for the ANN models. The input layer is followed by a single hidden layer with 64 nodes and the final output layer.

Using a more complex architecture with more hidden layers and nodes did not show any significant improvement in performance, thus we chose this relatively shallow one for its simplicity and discrete overall efficiency.

An important hyperparameter of ANN models is the number of *epochs*. An epoch is a single pass through the entire training dataset with a forward and backward propagation step.

The optimization algorithm used in this work for the ANN models is the *Adam* algorithm [73].

4.1.3 Evaluating the performance of the models

Finally, a trained model will produce an output score for each candidate event and, if trained properly, the score for signal and background events should be significantly different. The classification of the events (as signal or background) is then done by choosing a cut value on such score: all events with a score above the cut value ($N^{\text{score} > \text{cut}}$) are classified as signal, while all events with a score below the cut value ($N^{\text{score} < \text{cut}}$) are classified as background. Different cut values will result in a different signal efficiency and background rejection, which are defined as the fraction of signal events that are correctly classified as signal and the fraction of background events that are correctly classified as background i.e. :

$$\begin{aligned} \text{signal efficiency} &= \frac{N_{\text{true signal}}^{\text{score} > \text{cut}}}{N_{\text{true signal}}^{\text{total}}} \\ \text{background rejection} &= 1 - \frac{N_{\text{true background}}^{\text{score} > \text{cut}}}{N_{\text{true background}}^{\text{total}}} . \end{aligned} \quad (4.2)$$

The receiver operating characteristic (ROC) curve is a useful tool that visualizes the signal efficiency against the background rejection for all possible cut values. The ideal ROC curve has a background rejection of 1 and a signal efficiency of 1: this means that the model can perfectly separate signal and background. The area under the ROC curve (AUC) is a metric that can be used to compare the performance of different binary classification models. The closer the AUC is to 1, the better the performance of the model, since it means that it is close to the ideal case [62]. Figure 4.4 shows illustrative examples of ROC curves for differently performing models.

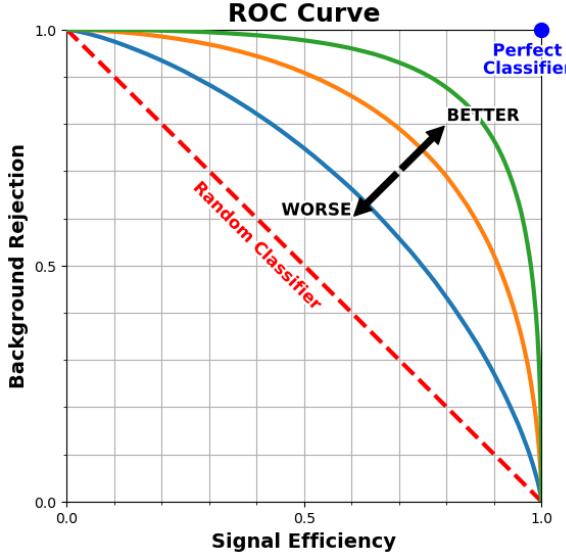


Figure 4.4: Examples of different ROC curves, comparing better and worse performing models. The dashed red line is a random classifier, the blue dot is the ideal classifier.

4.2 Signal topology and variable selection

The process under study is the decay $D_s^+ \rightarrow N (\rightarrow \mu^\pm \pi^\mp) \mu^+$ where N refers to the HNL. The final state consists in two muons and a pion. To distinguish them, we will refer to the muon coming from the D_s as μ_D and the muon coming from the HNL as μ_N . Figure 4.5 shows a schematic representation of the topology of this process, where a reconstructed displaced vertex formed by the HNL decay products (μ_N and π) is sketched. The muons and pion in the event are in the low transverse momentum (p_T) regime. The main background comes from QCD processes at low p_T , since contribution of physics at high p_T is negligible. The background is combinatorial in nature, arising from unrelated combinations of two muons and a track.

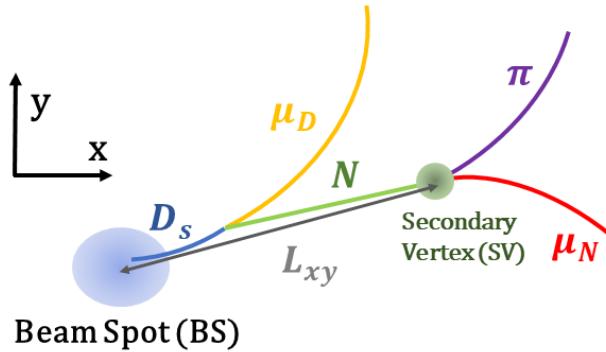


Figure 4.5: Topology of the decay under study $D_s^+ \rightarrow N (\rightarrow \mu^\pm \pi^\mp) \mu^+$ projected on the $x-y$ plane; N refers to the HNL; L_{xy} is the HNL decay length measured with respect to the beam spot.

The datasets used for the training of the ML models are MC generated (see appendix A for details). The signal datasets take into account four different values of HNL mass m_N : 1.0, 1.25, 1.5 and 1.8 GeV and mean proper decay length (or lifetime for short) $c\tau$: 10 mm.

A set of loose pre-selection criteria are applied to reduce the number of candidates arising from background while keeping the majority of those arising from signal events, as described in section 3.3.

4.2.1 Discriminating variables

The variables used for discriminating signal from background in the cut-based approach (see section 3.3.3) are the following:

- the HNL vertex displacement significance (L_{xy}/σ) on the transverse plane, i.e. the HNL decay length divided (see fig. 4.5) by its uncertainty measured with respect to the beam spot (BS);
- the impact parameter (IP) of the pion on the transverse plane, measured with respect to the BS;
- the impact parameter significance (IPS) of μ_N on the transverse plane i.e. the IP of μ_N measured with respect to the BS, divided by its uncertainty.

In a ML approach, we can include more variables for the signal selection, which could help achieve a better discrimination. The set of variables that are used as input features (see section 4.1) for the training of the ML models are the following:

- L_{xy}/σ : HNL vertex displacement significance on the transverse plane, measured with respect to Beam Spot (BS) (see fig. 4.5);
- $p_T(D_s)$, $p_T(\mu_D)$, $p_T(\mu_N)$, $p_T(\pi)$: transverse momentum of D_s , μ_D , μ_N , π ; the background should mostly populate the soft spectrum of p_T while the signal should populate slightly higher p_T spectrum, especially for HNLs with higher mass;
- $\text{IPS}(\mu_D)$, $\text{IPS}(\mu_N)$, $\text{IPS}(\pi)$, $\text{IP}(\pi)$: IPS of μ_D , μ_N , π , and IP of π on the transverse plane, measured with respect to the BS; because of the long-lived nature of the HNL, the IP of its decay products (μ_N and π) should be large compared to the combinatorial background; $\text{IPS}(\mu_D)$ has an appreciable half-life which should lead to a distinct signature for signal events that distinguish them from combinatorial background;
- $\cos \theta_{pt}(D_s)$, $\cos \theta_{pt}(N)$: cosine of the pointing angle of D_s and HNL on the transverse plane; the pointing angle θ_{pt} is the angle formed by the fitted p_T at the vertex and the vector joining the BS with the position of the vertex (see fig. 4.6); signal events should have an angle closer to 0 compared to the accidental background;
- vertex prob(D_s), vertex prob(N): fit probability of the D_s and HNL vertices; since background events are combinatorial, they should have a worse vertex fit probability;
- $\Delta R(\mu_D, \mu_N)$, $\Delta R(\mu_D, \pi)$: track separation between μ_D and μ_N , and between μ_D and π , where $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$;
- $N_{pixel}(\mu_N)$, $N_{pixel}(\pi)$: number of hits in the pixel system for μ_D , μ_N , and π ; because of the long-lived nature of the HNL, the number of pixel hits for its decay products should be small compared to the combinatorial background;
- Iso(μ_N): isolation of μ_N (see section 2.4.1); isolation is computed as $\text{Iso}(\mu) = \sum p_T^{\text{track}}/p_T^\mu$ where the sum is on the tracks from the primary vertex in a cone of $\Delta R = 0.3$ around the muon. Muons from signal events tend to be more isolated than muons from combinatorial background.

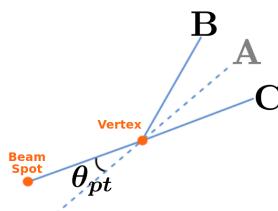


Figure 4.6: Sketch of the pointing angle $\theta_{pt}(A)$ for the generic decay $A \rightarrow BC$;

Correlation of input features

We initially studied a larger set of input features, but highly correlated features constitute redundant information for the algorithm, so some have been removed. The correlation matrix for the background dataset is shown in fig. 4.7 while the correlation matrices for each signal mass hypothesis dataset are shown in figs. 4.8 to 4.11. In the final set of input features there is no correlation coefficient above 80%. An interesting thing to remark is the correlation between $\text{IPS}(\mu_N)$ and $\text{IPS}(\pi)$, which is > 70 % for the signal datasets and 30 % for the background. This is expected, since the muon and the pion come from the same secondary vertex in the signal, while in the background the association between μ_N and π is random.

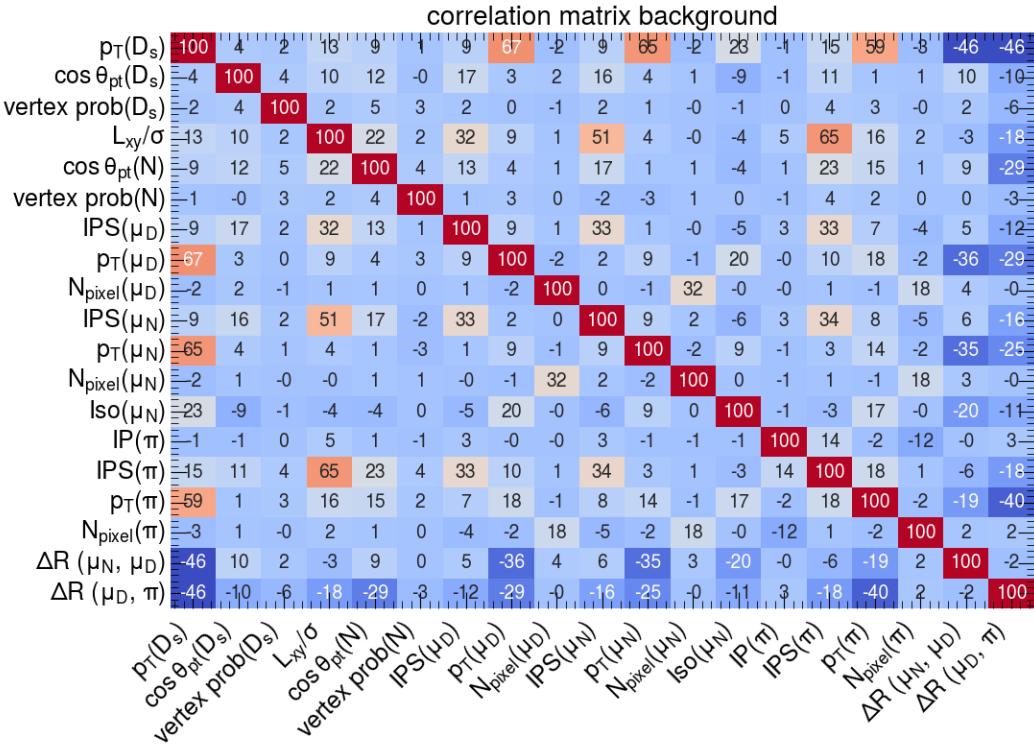


Figure 4.7: Correlation matrix of the input features for the QCD background dataset.

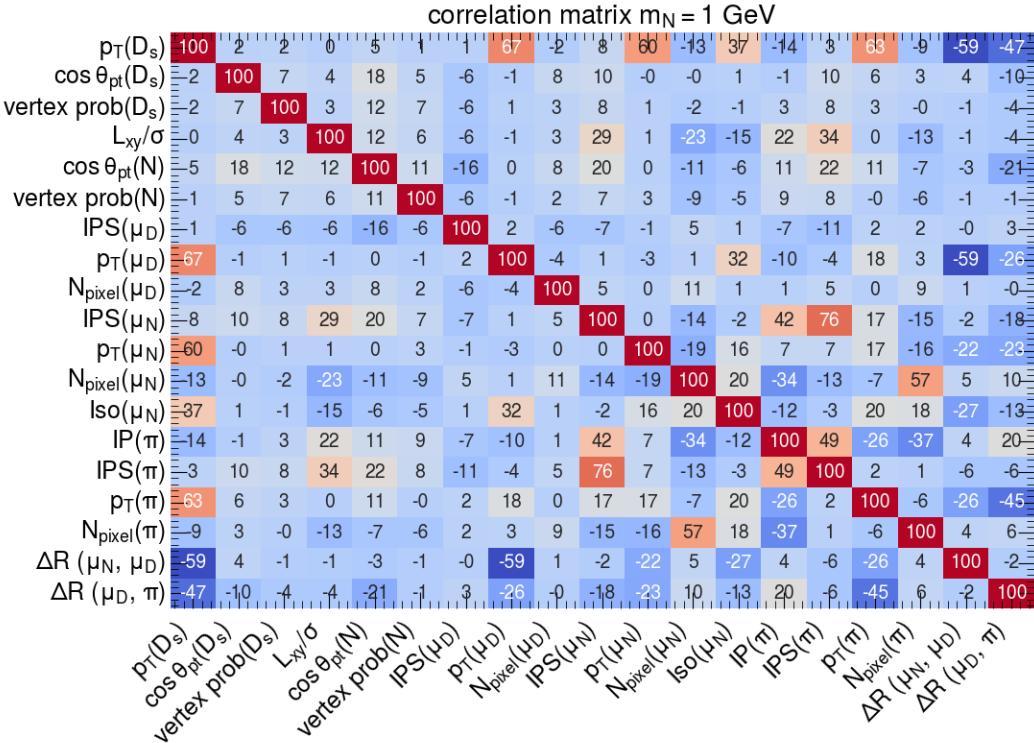


Figure 4.8: Correlation matrix of the input features for the signal dataset with mass hypothesis $m_N = 1.0 \text{ GeV}$.

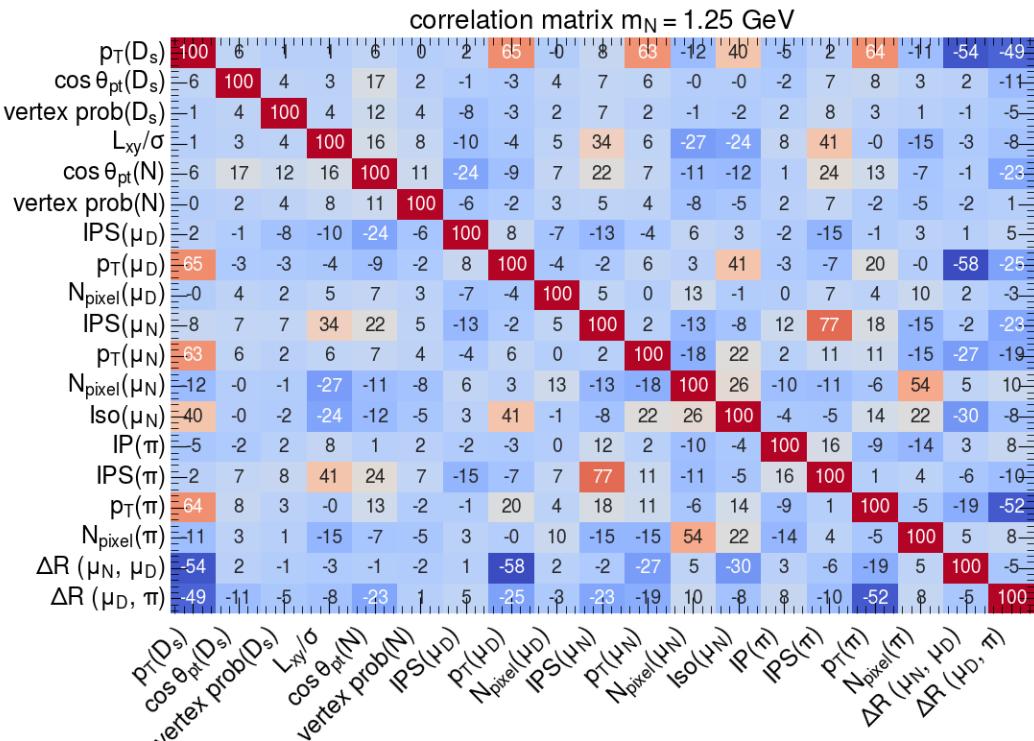


Figure 4.9: Correlation matrix of the input features for the signal dataset with mass hypothesis $m_N = 1.25 \text{ GeV}$.

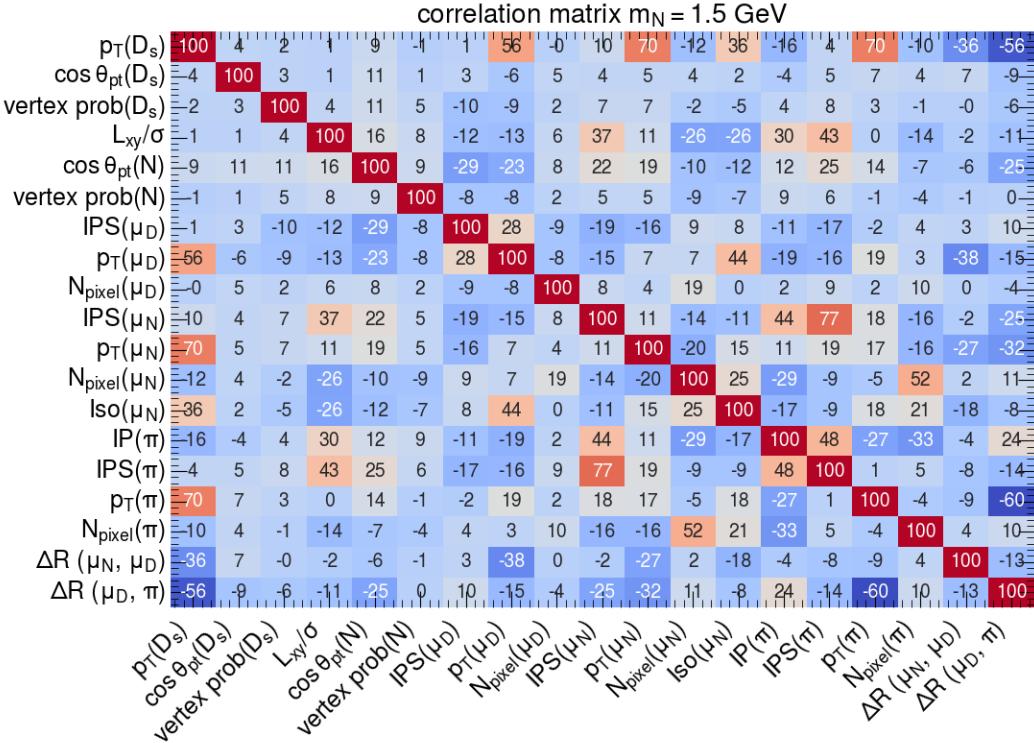


Figure 4.10: Correlation matrix of the input features for the signal dataset with mass hypothesis $m_N = 1.5 \text{ GeV}$.

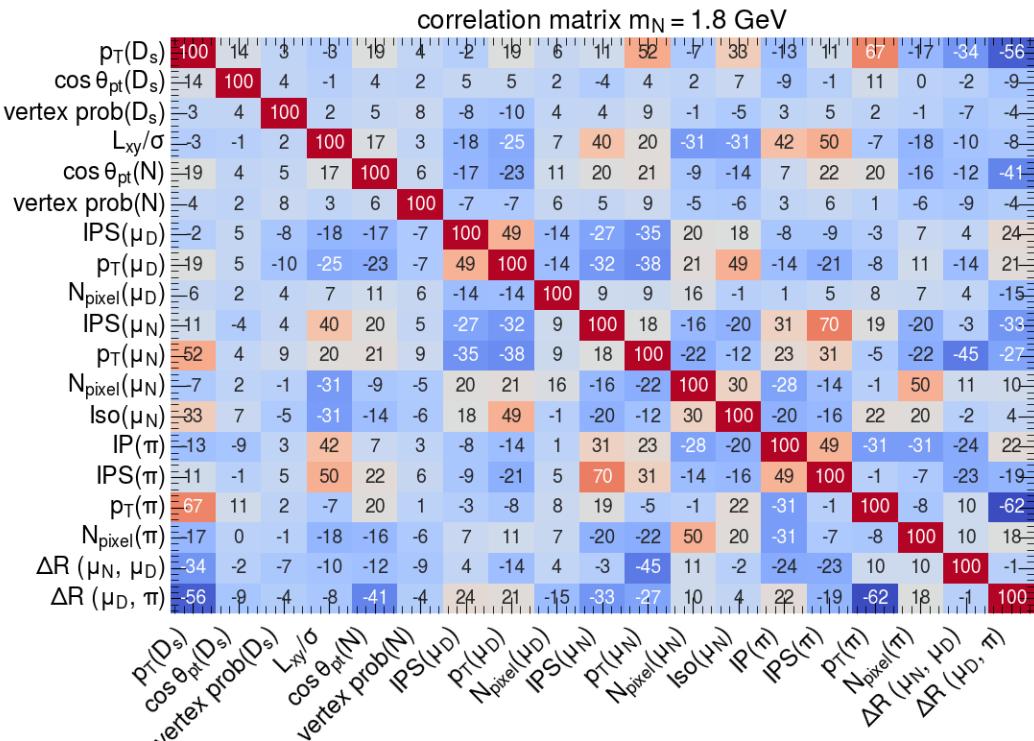


Figure 4.11: Correlation matrix of the input features for the signal dataset with mass hypothesis $m_N = 1.8 \text{ GeV}$.

4.2.2 Categorization of events

The search for HNL is not focusing on a single signal signature and different HNL masses or lifetime hypotheses can have different signatures. The selection strategy should aim to discriminate against the background for as many signatures as possible. The selection is therefore optimized in different event categories. The events are divided into 6 different categories based on the HNL vertex displacement significance (L_{xy}/σ) and the relative sign of the muons.

The background comes from combinatorics that are mostly low-displaced, while signal events can have very large displacements. This is reflected in the L_{xy}/σ distribution of the signal and background, shown in fig. 4.12. This categorization allows studying different signal lifetime hypotheses, since long-lived signals should mostly populate the high L_{xy}/σ categories, while short-lived signals should mostly populate the low L_{xy}/σ categories. Three different categories are defined based on the displacement significance:

- Low L_{xy}/σ category: $L_{xy}/\sigma \leq 50$;
- Medium L_{xy}/σ category: $50 < L_{xy}/\sigma \leq 150$;
- High L_{xy}/σ category: $L_{xy}/\sigma > 150$.

The categorization on relative muon sign is:

- muons have same sign (SS);
- muons have opposite sign (OS).

This categorization is physically significant because same sign muons violate lepton number conservation and are only allowed if the neutrino is of Majorana kind, while opposite sign muons conserve lepton number and are allowed by both Dirac and Majorana kind neutrinos. Table 4.1 shows the naming we use for each of the six categories of the analysis.

	Same Sign muons	Opposite Sign muons
$L_{xy}/\sigma \leq 50$	lowDisp_SS	lowDisp_OS
$50 < L_{xy}/\sigma \leq 150$	mediumDisp_SS	mediumDisp_OS
$L_{xy}/\sigma > 150$	highDisp_SS	highDisp_OS

Table 4.1: Summary table of all the six categories used in the analysis, with their respective names. L_{xy}/σ is the HNL vertex displacement significance.

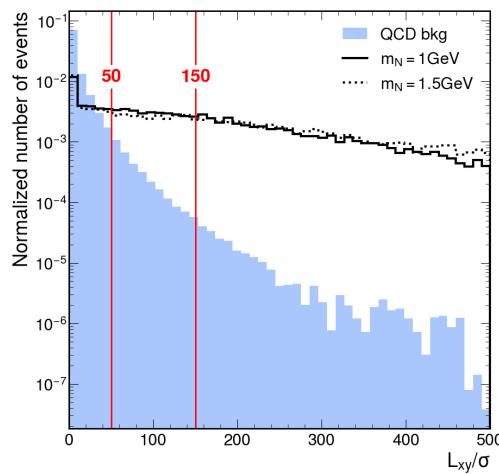


Figure 4.12: Distribution of the HNL vertex displacement significance L_{xy}/σ for the QCD background and signal mass hypotheses $m_N = 1.0$ GeV (solid), $m_N = 1.5$ GeV (dotted). The red vertical lines at $L_{xy}/\sigma = 50$ and $L_{xy}/\sigma = 150$ refer to the boundaries of the event categories definition. All histograms are normalized to unit area.

4.3 Training the models

The signal datasets take into account four different values of HNL mass m_N : 1.0, 1.25, 1.5 and 1.8 GeV and lifetime $c\tau$: 10 mm. The background dataset comes from combinatorial QCD events. The full details on the signal and background datasets is described in section 3.2.

Like for the selection of current cut-based analysis, we use the signal datasets with lifetime $c\tau = 10$ mm for the training of the ML models. This ensures consistency with previous work, and should also represent the most conservative choice: if we can select particles with short lifetimes, it's likely we can also select those with longer lifetimes, while the reverse is less likely to be true.

4.3.1 Data preparation

The datasets, derived from the preselection stage of the current (see section 3.3) analysis, are in the format of ROOT NTuples. To facilitate further analysis in Python, we adopt the *uproot* [74] library. This library allows us to convert these ROOT NTuples [75] into a tabular format that Python can handle more efficiently. Once the data is in a Python-friendly format, we can seamlessly interface with a wide range of ML libraries via their Python programming interfaces. These libraries provide us with the tools necessary for advanced data analysis and model training. The events of all the data samples are also weighted to account for several correction factors (e.g. cross-sections, scale factors, efficiencies, ...) described in section 3.3.2.

The analysis uses four different datasets for the signal, one for each signal mass hypothesis, and one for the QCD background. Each of these datasets is split into 6 category datasets, based on the displacement of the HNL vertex and the relative sign of the muons, ad describe in table 4.1 above.

The events in the category datasets are imbalanced, i.e. the total weights of the signal class and background class are different. Since this can negatively impact the training process, a constant multiplicative factor is added to the weights of each class to ensure that the total weights are equal. Taking into account the split in categories, we have a total of 24 signal-background pairs of datasets to train the models on.

Training, validation and test splitting

As mentioned in section 4.1, it is common practice to split the dataset into three different sets: the training set, the testing set, and the validation set. The splitting fractions chosen for this work are:

- training set: 65% of the total;
- testing set: 25% of the total;
- validation set: 10% of the total.

After dividing the dataset into training, testing, and validation sets, it is crucial to maintain an unbiased distribution of events. Since the events in our datasets have different weights, randomly choosing events for the training, testing, and validation sets could lead to a biased distribution of signal and background classes across the sets. The splitting is done in such a way that the weights of the signal and background classes are equal in the training, testing, and validation sets, preserving the original proportions of the dataset.

4.3.2 Configuration of the ML models

In section 4.1 we introduced the basic concepts of the ML models used in this work. The specific implementation of the models used is:

- Boosted Decision trees with AdaBoost boosting algorithm, implemented with the *Scikit-learn* python module [76], we refer to this as BDT;
- Gradient tree boosting, implemented with the open-source software library *XGBoost* [70], we refer to this as XGB;
- Artificial Neural networks, implemented with the *Keras* high-level API of the *TensorFlow* platform [77], we refer to this as ANN.

The hyperparameters used for the training of the models are listed in table 4.2 for BDT, table 4.3 for XGB, and in table 4.4 for ANN. Further explanations of the hyperparameters were already provided in sections 4.1.1 and 4.1.2.

The chosen hyperparameters are determined through a series of scans, as a compromise between performance and computation time.

4.3.3 Monitoring training progress

As previously mentioned in section 4.1, monitoring the performance of the model on the validation sets during the training process is important to keep overtraining under control. The example considered in fig. 4.13 shows the training behavior of the XGB models for the signal mass hypothesis $m_N = 1.0$ GeV in the category lowDisp_SS.

The y-axis shows the loss value, computed using the log-loss defined in eq. (4.1), that represents the degree of correctness of the predictions (see section 4.1), while the x-axis shows the iteration number, in this case represented by the number of trees in the model (see section 4.1.1). Lower values of the loss translate to a more accurate prediction. The plot shows the loss for both the training and validation sets. As the number of trees increases, the loss on the training set keeps decreasing, while the loss on the validation set starts flattening out.

Since the validation set is not used for training, the performance on the validation set is an indicator of the model’s generalization capability. Thus, the training is stopped when the performance on the validation set does not improve anymore, because if the training continues, there is a risk of overfitting the model to the training set i.e. the model will not be able to generalize well to new, unseen data.

In this specific case, in the region where the validation loss is relatively flat, the model reaches the lowest loss value for a number of trees equal to 43, and since all 10 subsequent values of the validation loss are higher, this is chosen as early stopping point.

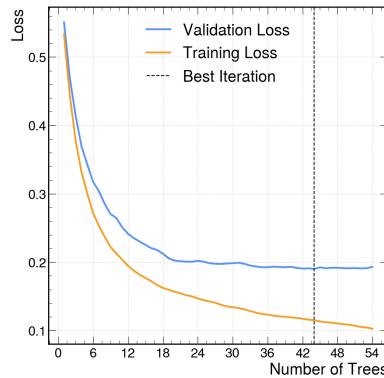


Figure 4.13: Performance of the XGB models on training and validation sets during the training process, for the signal mass hypothesis $m_N = 1.0$ GeV in the category lowDisp_SS. The blue line represents the performance on the validation set, while the orange line represents the performance on the training set. The black dashed line represents the early stopping point.

An analogous behavior is seen in fig. 4.14, which shows the training behavior of the XGB models on all categories for the signal mass hypothesis $m_N = 1.0$ GeV.

Boosted Decision Trees (BDT)	
Hyperparameter	Value
Maximum depth	3
Maximum number of trees	10
Boosting type	AdaBoost
Early stopping metric	Logistic loss
Patience	3

Table 4.2: Summary table of the hyperparameters used for the BDT models.

XGBoost (XGB)	
Hyperparameter	Value
Maximum depth	3
Maximum number of trees	100
Loss function	Logistic loss
Early stopping metric	Logistic loss
patience	10

Table 4.3: Summary table of the hyperparameters used for the XGBoost model.

Artificial Neural Network (ANN)	
Hyperparameter	Value
hidden layers	1
hidden layer nodes	64
epochs	10
loss function	Logistic loss
optimization algorithm	Adam
Early stopping metric	Logistic loss
patience	3

Table 4.4: Summary table of the hyperparameters used for the ANN models.

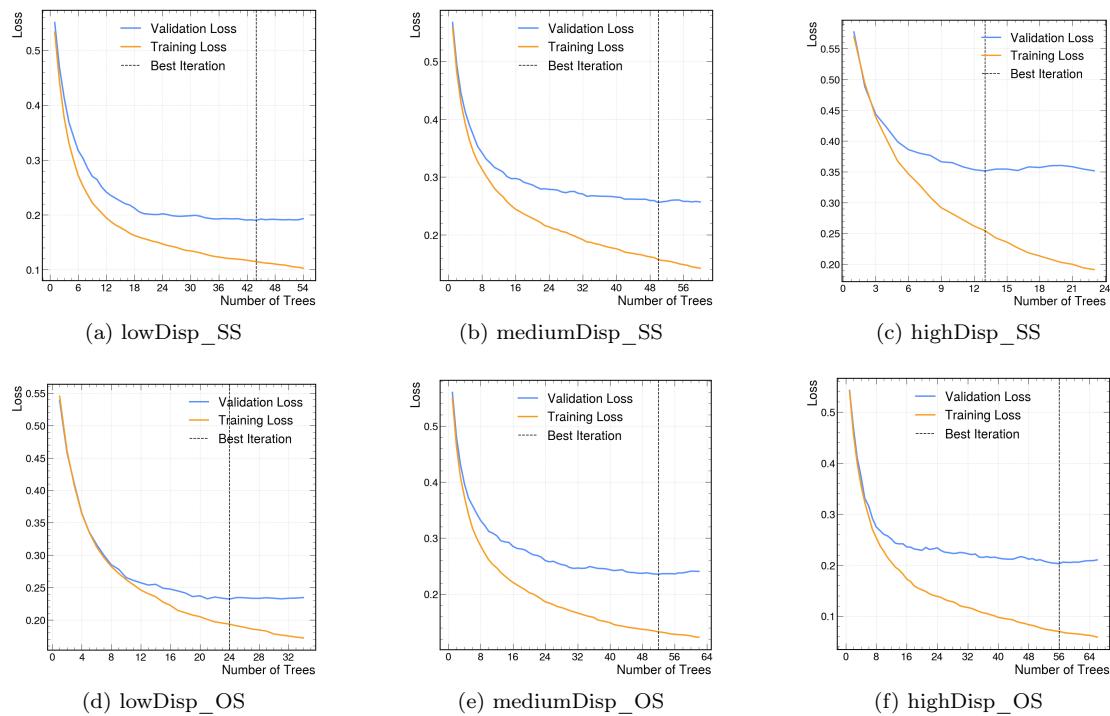


Figure 4.14: Performance of the XGB models on training and validation sets during the training process, for the signal mass hypothesis $m_N = 1.0$ GeV. Each plot represents a different category. The blue line represents the performance on the validation set, while the orange line represents the performance on the training set. The black dashed line represents the early stopping point.

4.4 Results

One model of each type (XGB, BDT, ANN) is trained on each of the six category datasets for each of the four signal mass hypotheses, for a total of 24 signal-background pairs of datasets and 72 models. After the training of the models, their performance is evaluated on the test set. Our purpose is to find the optimal models that maximize the signal-to-background discrimination, and compare the signal efficiency and background rejection we obtain with the ones from the cut-based analysis.

4.4.1 Selecting the best algorithm

As a metric to compare the performance of the models, we use the AUC of the ROC curve on the test set (see section 4.1.3). As an example, fig. 4.15 shows the ROC curve for all models, for signal mass hypothesis $m_N = 1.0$ GeV in the category mediumDisp_SS. The AUC values for each model are shown in the legend.

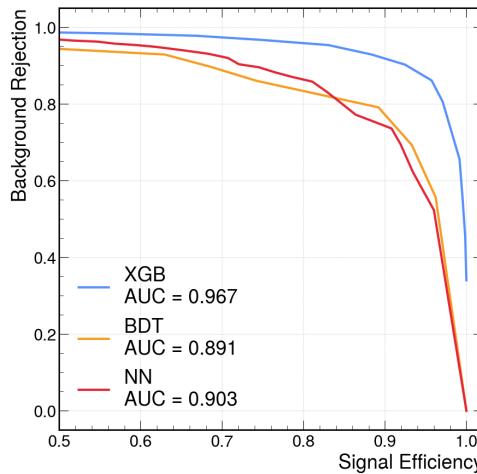


Figure 4.15: ROC curve for all models, category mediumDisp_SS, signal mass hypothesis $m_N = 1.0$ GeV. The AUC values for each model are shown in the legend.

Each ROC curve shows the trade-off between signal efficiency and background rejection: we can reach a higher signal efficiency at the cost of a lower background rejection, and vice versa. To choose the best model we use as a metric the AUC of the ROC curve (see section 4.1.3) and thus the best model is XGB: it can be seen that its curve is “above” the two other curves, which means that it achieves a better background rejection for any given signal efficiency.

The ROC curves for all models and categories are shown for each signal mass hypothesis: fig. 4.16 ($m_N = 1.0$ GeV), fig. 4.17 ($m_N = 1.25$ GeV), fig. 4.18 ($m_N = 1.5$ GeV), fig. 4.19 ($m_N = 1.8$ GeV). The AUC values for all models, categories, and signal mass hypotheses are summarized in table 4.5.

Since the AUC of the XGB models is consistently higher than the AUC of the BDT and ANN models for all categories and signal mass hypotheses, from now on we will only consider the XGB models.

The ROC curves for the signal mass hypothesis $m_N = 1.8$ GeV (fig. 4.19) show a discontinuous and erratic behavior. This is because the dataset for this signal mass hypothesis has low statistics, which influences the distribution of scores and consequently the ROC curve. This can be seen in the score distribution plots that will be shown and commented in section 4.4.3.

category	AUC of model			category	AUC of model		
	XGB	BDT	ANN		XGB	BDT	ANN
lowDisp_SS	0.961	0.925	0.917	lowDisp_SS	0.966	0.938	0.958
mediumDisp_SS	0.967	0.891	0.903	mediumDisp_SS	0.966	0.934	0.919
highDisp_SS	0.944	0.902	0.843	highDisp_SS	0.97	0.937	0.922
lowDisp_OS	0.957	0.914	0.91	lowDisp_OS	0.978	0.956	0.957
mediumDisp_OS	0.964	0.926	0.924	mediumDisp_OS	0.978	0.912	0.931
highDisp_OS	0.974	0.892	0.939	highDisp_OS	0.985	0.963	0.947

(a) signal mass hypothesis $m_N = 1.0$ GeV

category	AUC of model			category	AUC of model		
	XGB	BDT	ANN		XGB	BDT	ANN
lowDisp_SS	0.994	0.977	0.975	lowDisp_SS	0.997	0.984	0.901
mediumDisp_SS	0.99	0.964	0.959	mediumDisp_SS	0.995	0.972	0.993
highDisp_SS	0.991	0.969	0.965	highDisp_SS	1.000	0.998	0.974
lowDisp_OS	0.989	0.919	0.965	lowDisp_OS	0.946	0.704	0.902
mediumDisp_OS	0.99	0.973	0.958	mediumDisp_OS	0.997	0.972	0.995
highDisp_OS	0.99	0.955	0.965	highDisp_OS	0.998	0.956	0.953

(b) signal mass hypothesis $m_N = 1.25$ GeV

category	AUC of model			category	AUC of model		
	XGB	BDT	ANN		XGB	BDT	ANN
lowDisp_SS	0.997	0.984	0.901	lowDisp_SS	0.997	0.984	0.901
mediumDisp_SS	0.995	0.972	0.993	mediumDisp_SS	0.995	0.972	0.993
highDisp_SS	1.000	0.998	0.974	highDisp_SS	1.000	0.998	0.974
lowDisp_OS	0.946	0.704	0.902	lowDisp_OS	0.946	0.704	0.902
mediumDisp_OS	0.997	0.972	0.995	mediumDisp_OS	0.997	0.972	0.995
highDisp_OS	0.998	0.956	0.953	highDisp_OS	0.998	0.956	0.953

(c) signal mass hypothesis $m_N = 1.5$ GeV

(d) signal mass hypothesis $m_N = 1.8$ GeV

Table 4.5: AUC of the ROC curves for each model and category, for all signal mass hypotheses.

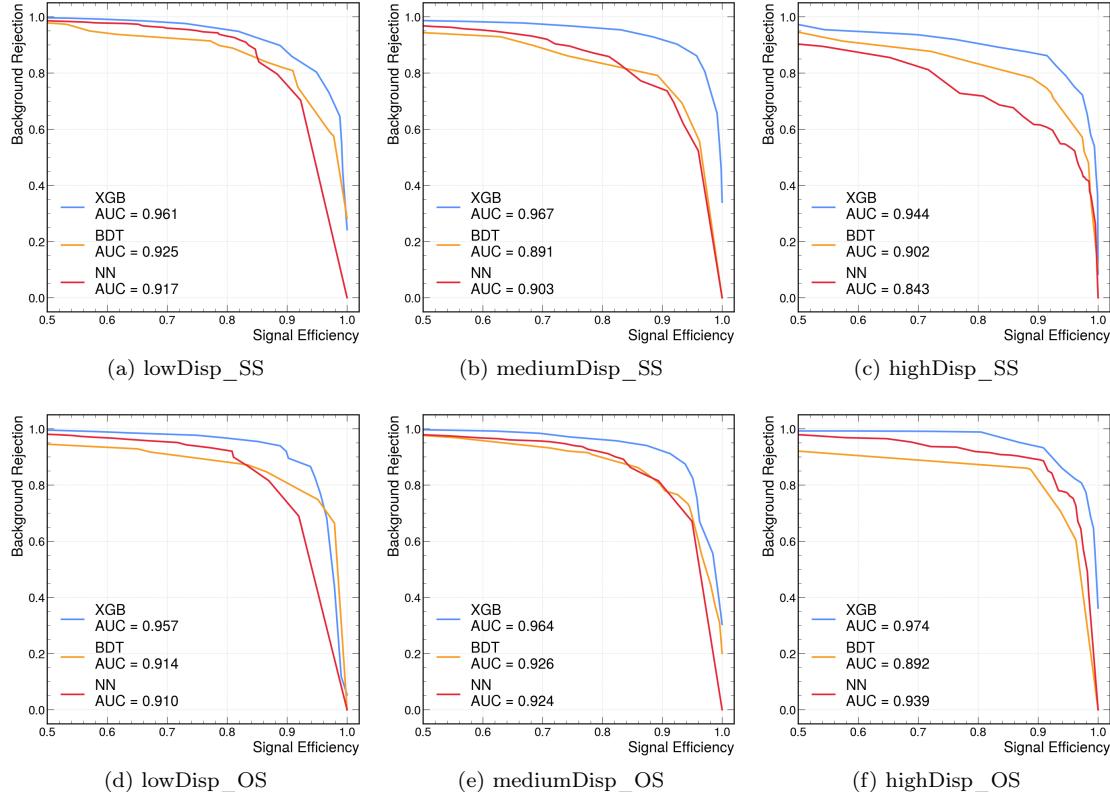


Figure 4.16: ROC curves of all models for the signal mass hypothesis $m_N = 1.0$ GeV, for each category.

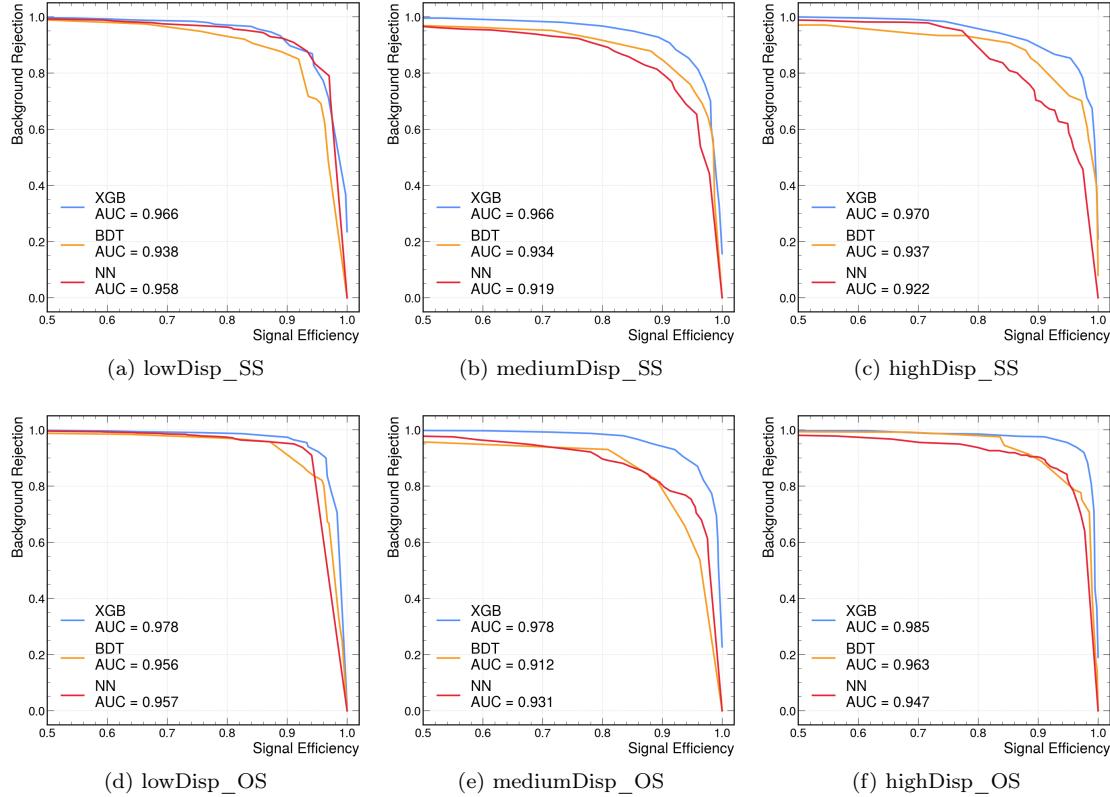


Figure 4.17: ROC curves of all models for the signal mass hypothesis $m_N = 1.25$ GeV, for each category.

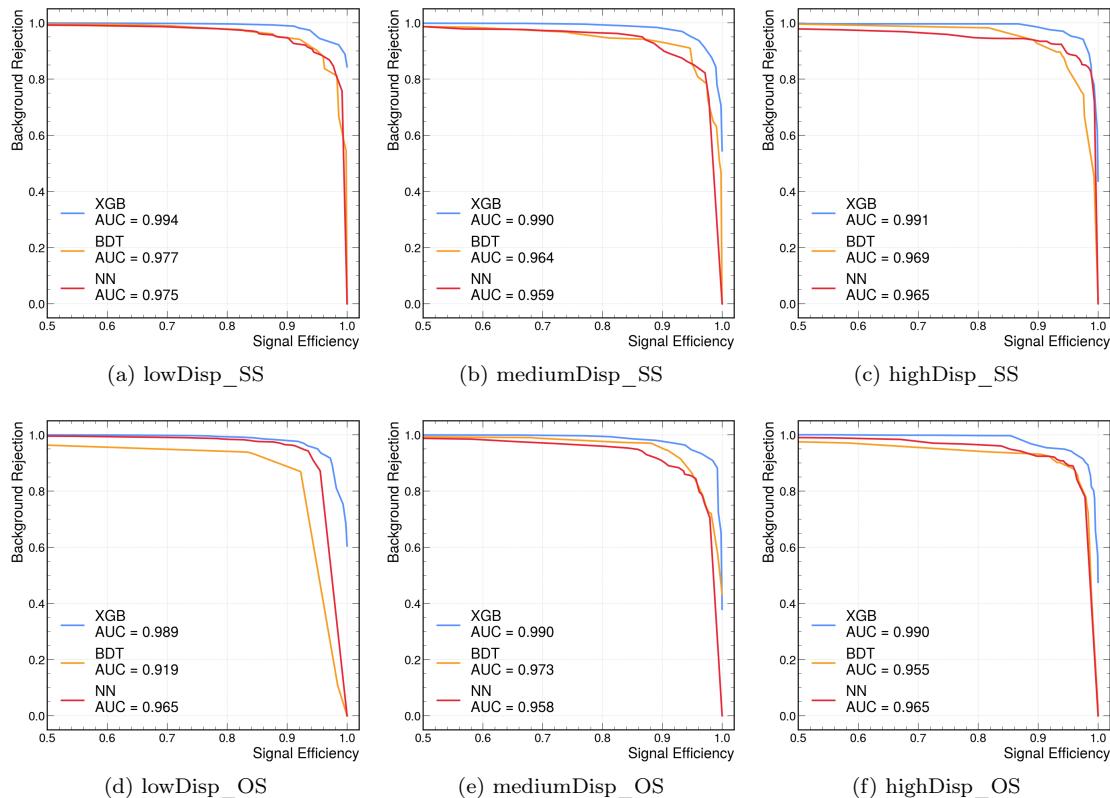


Figure 4.18: ROC curves of all models for the signal mass hypothesis $m_N = 1.5$ GeV, for each category.

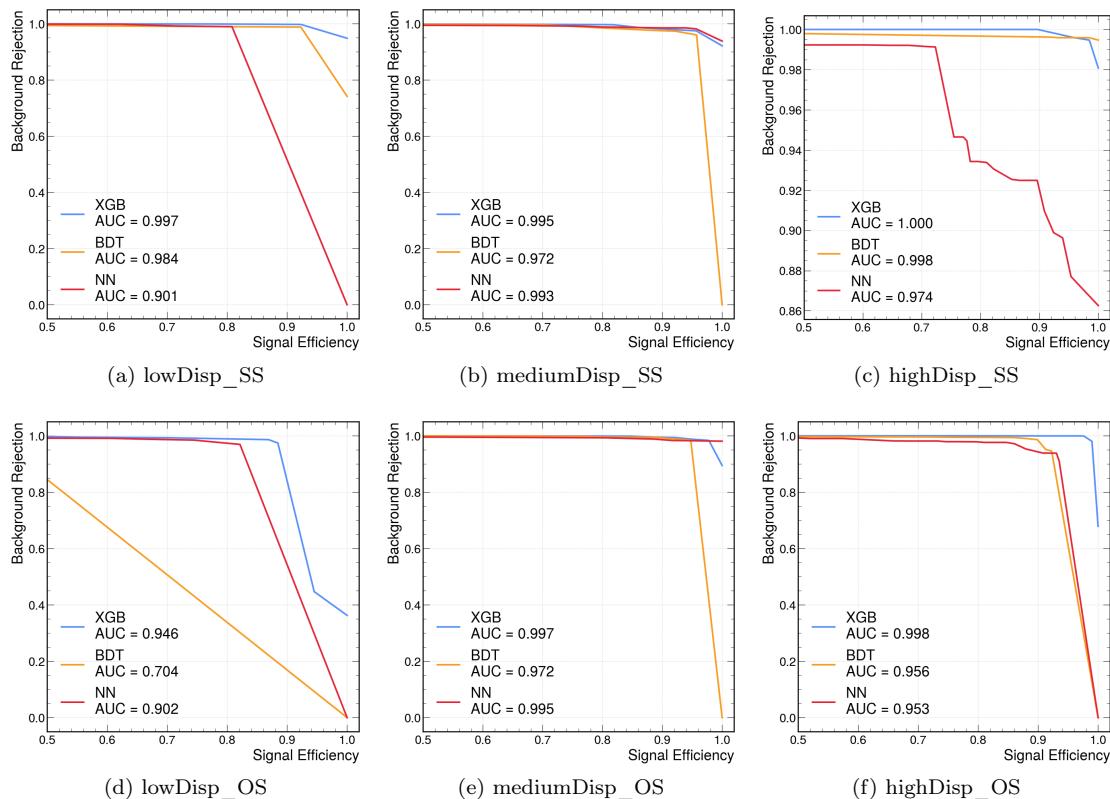


Figure 4.19: ROC curves of all models for the signal mass hypothesis $m_N = 1.8$ GeV, for each category.

4.4.2 Most discriminating variables

The variables we have used as input features can have a different impact on the discrimination of signal from background. Tree ensemble models can provide a measure of the importance of each variable in the discrimination process [62]. This constitutes a useful explainability feature of the models, as it allows us to understand which physical variables are used by the model to discriminate signal from background.

From now on, we will refer to this metric as “importance”. Here we consider the importance as computed by the XGB models, since we have chosen it as the best performing model (as reported in section 4.4.1). The four variables with the highest importance for each category, for signal mass hypotheses $m_N = 1.0$ GeV and $m_N = 1.5$ GeV, are shown in tables 4.6 and 4.7.

variable	importance	variable	importance	variable	importance
$\cos \theta_{\text{pt}}(N)$	0.3212	$\cos \theta_{\text{pt}}(N)$	0.279	$\cos \theta_{\text{pt}}(N)$	0.2117
$\text{IP}(\pi)$	0.0991	$\text{IPS}(\mu_D)$	0.1219	$\text{IPS}(\mu_D)$	0.11
$\cos \theta_{\text{pt}}(D_s)$	0.0837	L_{xy}/σ	0.0821	$p_T(\mu_D)$	0.0989
$\text{IPS}(\mu_N)$	0.0756	$\text{Iso}(\mu_N)$	0.0703	$\text{IPS}(\pi)$	0.0981
(a) Category lowDisp_SS		(b) Category mediumDisp_SS		(c) Category highDisp_SS	
variable	importance	variable	importance	variable	importance
$\cos \theta_{\text{pt}}(N)$	0.2746	$\cos \theta_{\text{pt}}(N)$	0.2844	$\cos \theta_{\text{pt}}(N)$	0.2027
$\text{IP}(\pi)$	0.1151	$\text{IPS}(\mu_D)$	0.1108	L_{xy}/σ	0.1038
$N_{\text{pixel}}(\pi)$	0.101	L_{xy}/σ	0.0752	$\text{IPS}(\mu_D)$	0.1006
L_{xy}/σ	0.0909	$\text{IP}(\pi)$	0.0675	$\text{IPS}(\pi)$	0.1002
(d) Category lowDisp_OS		(e) Category mediumDisp_OS		(f) Category highDisp_OS	

Table 4.6: Four most important variables for the XGB models in each category, for the signal mass hypothesis $m_N = 1.0$ GeV.

variable	importance	variable	importance	variable	importance
$\cos \theta_{\text{pt}}(N)$	0.4112	$\cos \theta_{\text{pt}}(N)$	0.4193	$\cos \theta_{\text{pt}}(N)$	0.2693
$\text{IPS}(\mu_N)$	0.085	$\text{IPS}(\mu_D)$	0.128	$\text{IPS}(\mu_D)$	0.1455
$\text{IP}(\pi)$	0.0793	$p_T(\mu_D)$	0.069	$p_T(\mu_D)$	0.1397
$p_T(\mu_D)$	0.0627	$\text{Iso}(\mu_N)$	0.055	$p_T(\mu_N)$	0.088
(a) Category lowDisp_SS		(b) Category mediumDisp_SS		(c) Category highDisp_SS	
variable	importance	variable	importance	variable	importance
$\cos \theta_{\text{pt}}(N)$	0.493	$\cos \theta_{\text{pt}}(N)$	0.3836	$\cos \theta_{\text{pt}}(N)$	0.2477
$\text{IP}(\pi)$	0.0913	$\text{IPS}(\mu_D)$	0.1232	$p_T(\mu_N)$	0.1835
$p_T(\mu_N)$	0.0789	$p_T(\mu_N)$	0.0758	$\text{IPS}(\mu_D)$	0.1628
$p_T(\mu_D)$	0.054	$\text{IPS}(\pi)$	0.065	$\text{IPS}(\pi)$	0.1166
(d) Category lowDisp_OS		(e) Category mediumDisp_OS		(f) Category highDisp_OS	

Table 4.7: Four most important variables for the XGB models in each category, for the signal mass hypothesis $m_N = 1.5$ GeV.

Among the most discriminating variable, as illustrative examples, we focus on the distribution of $\cos \theta_{\text{pt}}(N)$, $\text{IP}(\pi)$, $\text{IPS}(\mu_N)$ and L_{xy}/σ . These variables are shown in figs. 4.20 to 4.23 for the signal mass hypotheses $m_N = 1.0$ GeV and $m_N = 1.5$ GeV in three categories with same sign muons and displacement ranging from low to high, in order to appreciate possible patterns as a function of displacement.

- $\cos \theta_{\text{pt}}(N)$ (fig. 4.20):

signal events and background events are both peaked towards 1, but the signal peak is higher and drops off more quickly for lower values of cosine; this is quite consistent across all categories;

- $\text{IP}(\pi)$ (fig. 4.21):

the high $\text{IP}(\pi)$ values are populated by more signal events than background events; if we compare the behavior across the three categories, we can see that the signal distribution is shifted towards higher $\text{IP}(\pi)$ values for higher displacement categories; the background distribution on the other hand drops off more quickly in the higher displacement categories;

- L_{xy}/σ (fig. 4.22):

signal events can have very large L_{xy}/σ while the combinatorial background distribution drops off more quickly; if we compare the behavior across the three categories, we can see that the background distribution drops off more quickly in the higher displacement categories;

- $\text{IPS}(\mu_N)$ (fig. 4.23):

the signal distribution is shifted towards higher $\text{IPS}(\mu_N)$ especially in the higher displacement categories;

The behaviors described are analogous for the opposite sign categories, see as illustrative example fig. 4.24 that shows the distribution of $\cos \theta_{pt}(N)$ for all displacement categories but with opposite sign muons. The fact that the behavior is similar across the opposite sign and same sign categories reinforces our hypothesis that the background is mainly combinatorial.

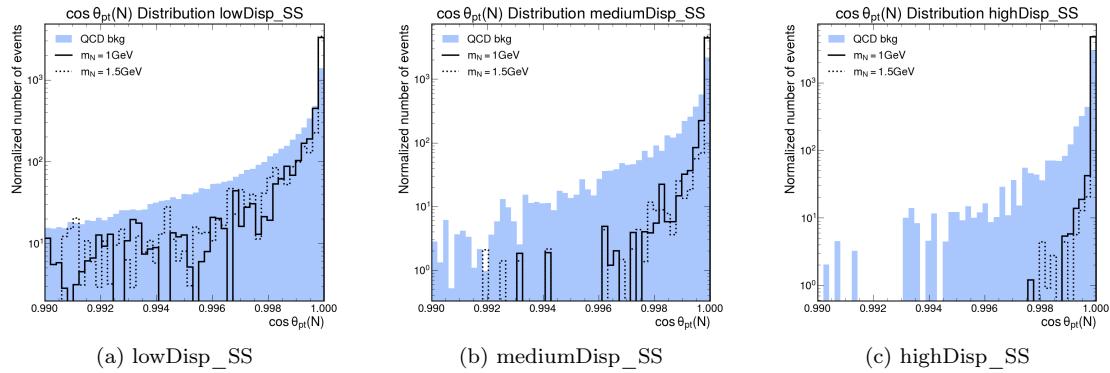


Figure 4.20: Distribution of the variable $\cos \theta_{pt}(N)$ for the QCD background (blue) and for the signal mass hypotheses $m_N = 1.0$ GeV (solid) and $m_N = 1.5$ GeV (dotted) across three different displacement categories.

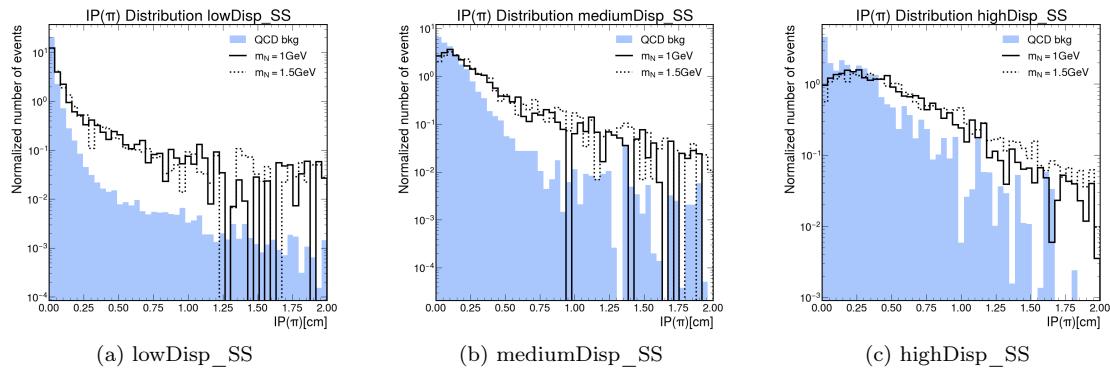


Figure 4.21: Distribution of the variable $\text{IP}(\pi)$ for background for the QCD background (blue) and signal for the signal mass hypotheses $m_N = 1.0$ GeV (solid) and $m_N = 1.5$ GeV (dotted) across three different displacement categories.

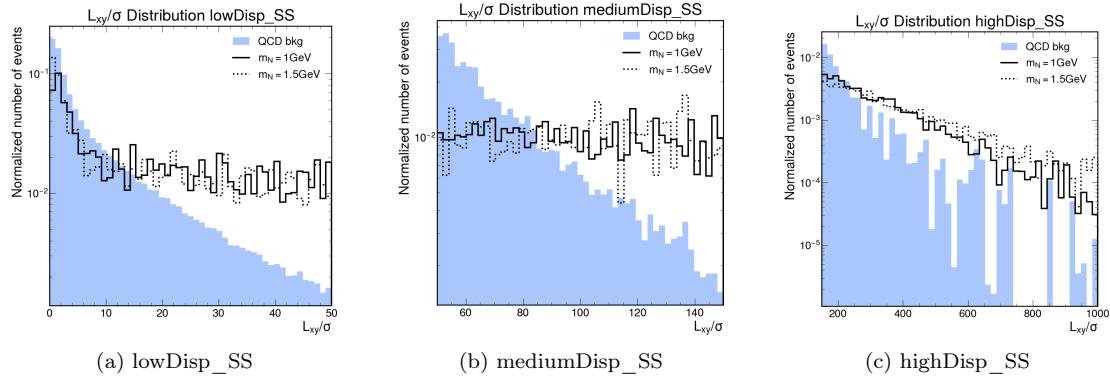


Figure 4.22: Distribution of the variable L_{xy}/σ for background for the QCD background (blue) and signal for the signal mass hypotheses $m_N = 1.0$ GeV (solid) and $m_N = 1.5$ GeV (dotted) across three different displacement categories.

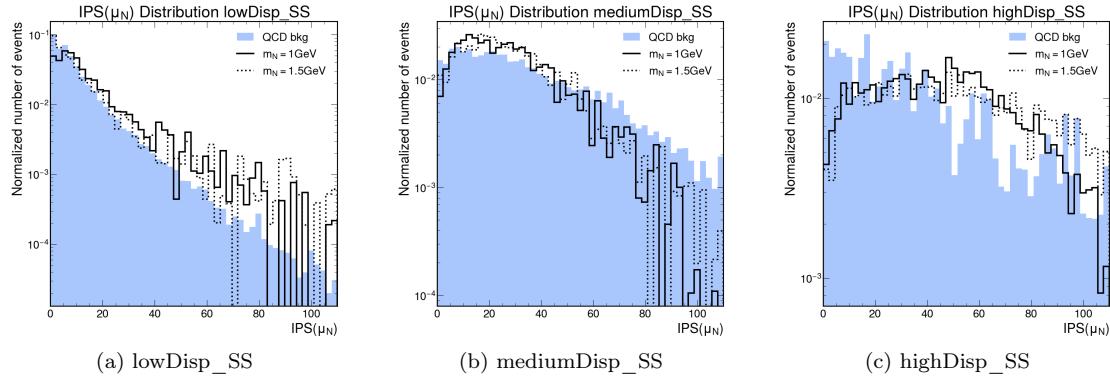


Figure 4.23: Distribution of the variable $\text{IPS}(\mu_N)$ for background for the QCD background (blue) and signal for the signal mass hypotheses $m_N = 1.0$ GeV (solid) and $m_N = 1.5$ GeV (dotted) across three different displacement categories.

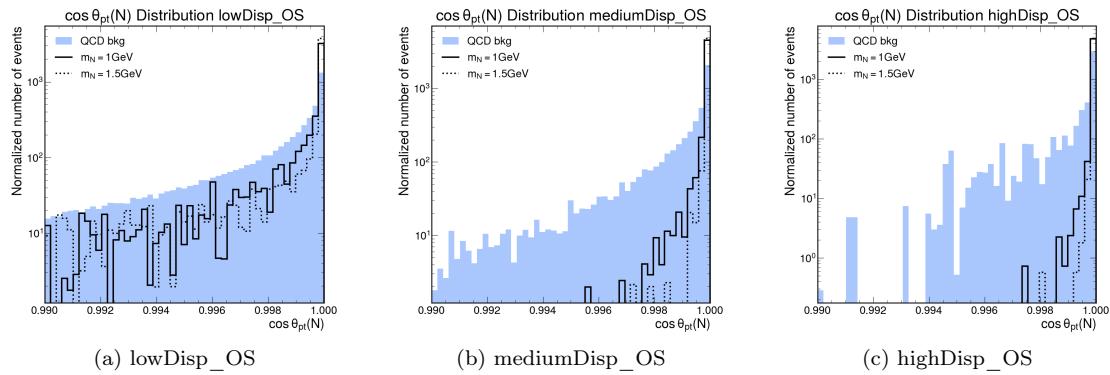


Figure 4.24: Distribution of the variable $\cos \theta_{pt}(N)$ for background for the QCD background (blue) and signal for the signal mass hypotheses $m_N = 1.0$ GeV (solid) and $m_N = 1.5$ GeV (dotted) across the three different displacement categories with opposite sign muons.

4.4.3 Model scores and cut optimization

We have determined that the XGB algorithm provides the best models for the analysis, and we have outlined the most important variables for the discrimination of signal from background. However, we still need to choose a cut on the output score produced by the model to classify the events as signal or background. In order to do this, we need to analyze the distribution of the scores produced by the models for each of the four signal mass hypotheses in each category.

As an example consider fig. 4.25, which shows the distribution of scores produced by the XGB models on the training and test sets for the signal mass hypothesis $m_N = 1.0$ GeV and category mediumDisp_OS. The orange histogram represents the distribution of scores for the background class, while the blue histogram represents the distribution of scores for the signal class. The filled histogram represents the test set, while the empty histogram represents the training set. Comparing the distribution of scores on the test and training sets is important as an additional overtraining check. In our case, the distributions of the scores on the test and training sets are satisfactorily similar, which indicates that the model is fairly not overtrained. We can see that the distribution of scores for the signal class is significantly separated from the distribution of scores for the background class, which reflects the good performance of the model in discriminating signal from background.

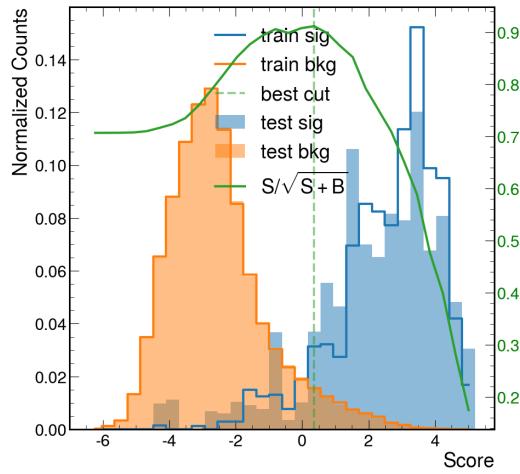


Figure 4.25: Distribution of scores of the XGB models on training and test sets for the signal hypothesis $m_N = 1.0$ GeV and category mediumDisp_SS. All histograms are normalized to unit area.

For completeness, the distribution of output scores of the XGB models on the training and test sets for each of the four signal mass hypotheses, for each category, are shown in figs. 4.26 to 4.29. Some fluctuations in the distribution of scores between the training and test sets are visible, that are due to the limited statistics: the signal has reduced statistics in the lower displacement categories; the background, on the other hand, is expected to have reduced statistics in the higher displacement categories, since the combinatorial QCD background is expected to decrease significantly with displacement. Since we have implemented early stopping to mitigate overtraining, and in most cases the differences in the distributions are not extreme, we do not consider this to be a sign of significant overtraining.

This issue of reduced statistics is particularly true for the signal mass hypothesis $m_N = 1.8$ GeV (see fig. 4.29): this is a peculiar limit case, very close to maximum allowed mass of the decay (the mass of the D_s meson), which makes the HNL decay products very soft and thus harder to reconstruct.

As a criterion for the choice of the cut on the output score, we consider the expected discovery significance (significance for short) Z_0 of the signal over the background [60], in its limit for a small number of signal events, which is given by:

$$Z_0 = \sqrt{2 \left[(S + B) \log \left(1 + \frac{S}{B} \right) - S \right]} \xrightarrow{S \ll B} \frac{S}{\sqrt{S + B}} , \quad (4.3)$$

where S and B are the number of signal and background events that pass the cut. The green curve in fig. 4.25 shows the significance as a function of the cut on the score, and we choose

the cut that maximizes the significance, indicated by the dashed green line. Table 4.8 shows the optimal cut on the score for each category and signal mass hypothesis, according to the maximum expected discovery significance.

category	Signal mass hypothesis (GeV)			
	1.0	1.25	1.5	1.8
lowDisp_SS	-0.938	-1.160	-1.592	-0.933
mediumDisp_SS	-0.136	0.086	1.147	-0.686
highDisp_SS	0.168	-0.225	0.062	-0.062
lowDisp_OS	0.361	0.136	-0.829	-0.564
mediumDisp_OS	-0.243	-0.338	0.364	-0.684
highDisp_OS	1.248	0.620	1.145	0.507

Table 4.8: Optimal cut on the score for each category and signal mass hypothesis, according to the maximum expected discovery significance.

From the output score distributions in figs. 4.26 to 4.29, it is visually clear that the signal and background histograms are more separated for the higher m_N signal hypothesis: this seems to indicate the fact that higher m_N signals have a more distinct signature, and are thus easier to discriminate.

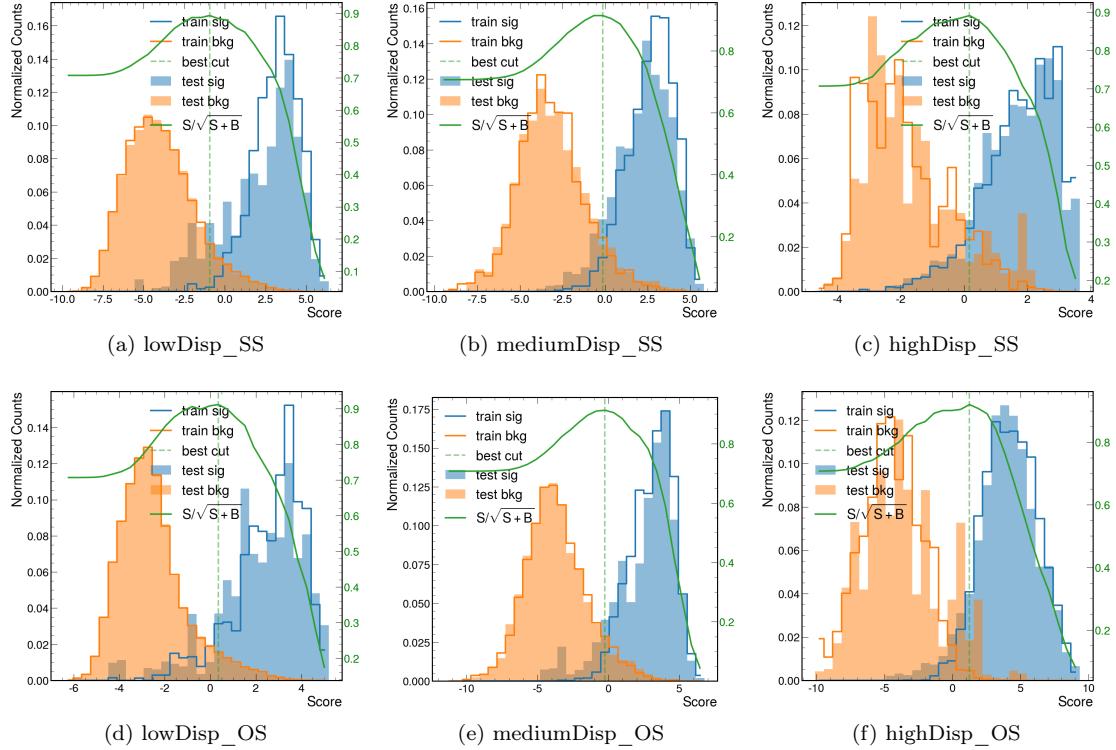


Figure 4.26: Distribution of scores of the XGB models on training and test sets for the signal mass hypothesis $m_N = 1.0$ GeV. All histograms are normalized to unit area. Distributions are shown for each category.

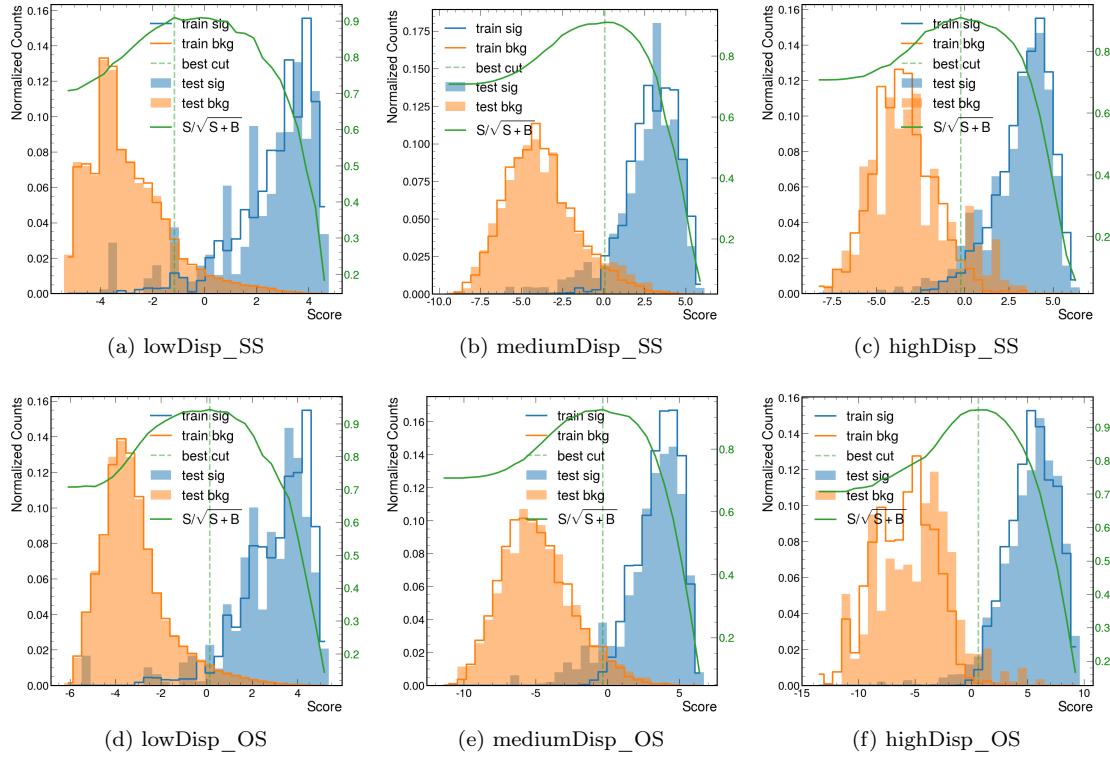


Figure 4.27: Distribution of scores of the XGB models on training and test sets for the signal mass hypothesis $m_N = 1.25$ GeV. All histograms are normalized to unit area. Distributions are shown for each category.

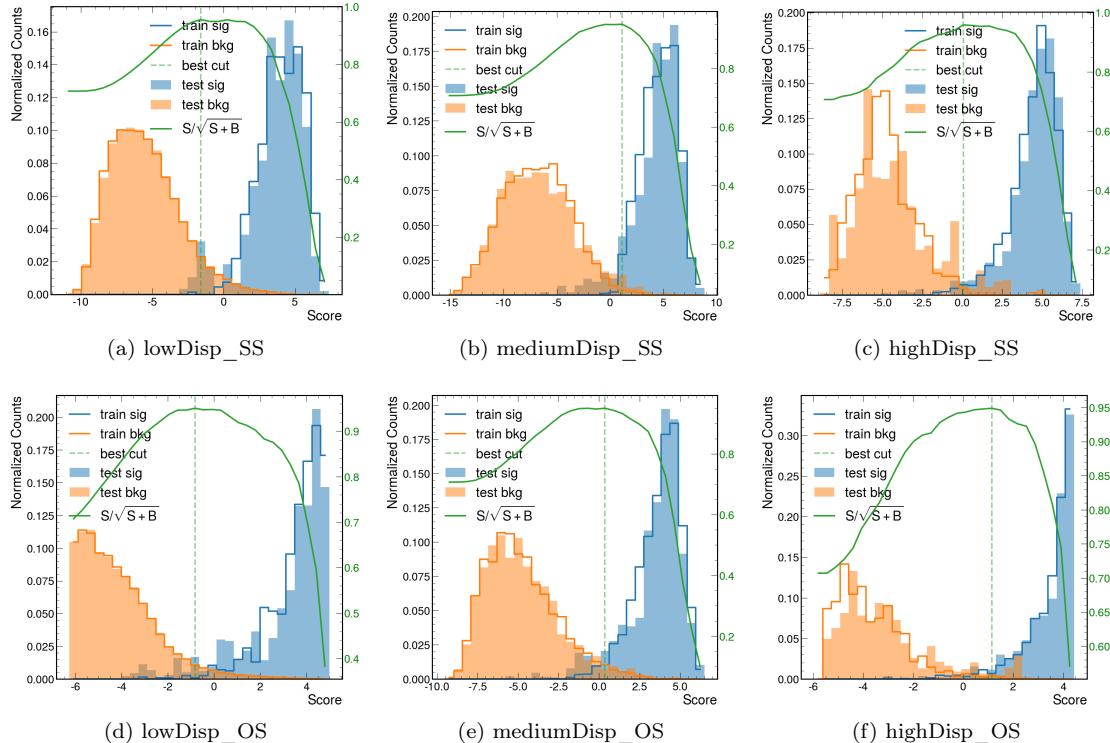


Figure 4.28: Distribution of scores of the XGB models on training and test sets for the signal mass hypothesis $m_N = 1.5$ GeV. All histograms are normalized to unit area. Distributions are shown for each category.

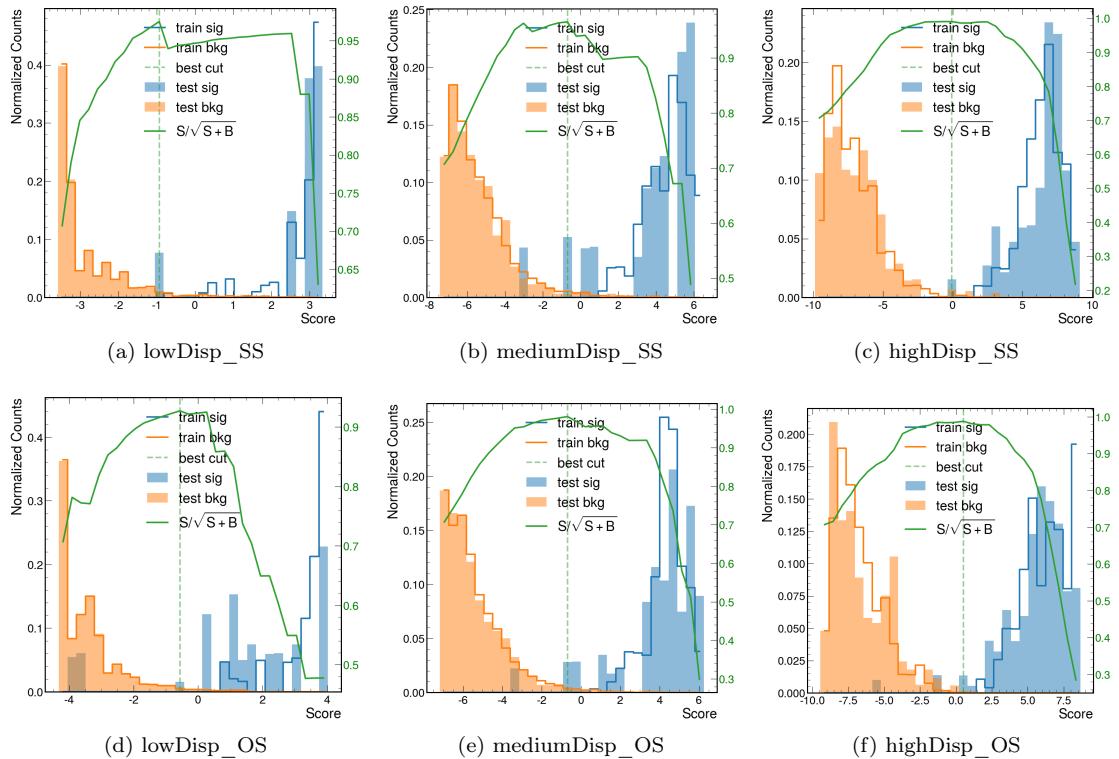


Figure 4.29: Distribution of scores of the XGB models on training and test sets for the signal mass hypothesis $m_N = 1.8$ GeV. All histograms are normalized to unit area. Distributions are shown for each category.

4.4.4 Results and comparison with the cut-based analysis

With the optimal cut determined (section 4.4.3), we have a model and a fixed working point for each category and signal mass hypothesis. Table 4.9 shows the signal efficiency and background rejection for the chosen working points of the XGB models, for all signal hypotheses and categories.

category	Signal efficiency	Background rejection	category	Signal efficiency	Background rejection
lowDisp_SS	0.873	0.930	lowDisp_SS	0.942	0.868
mediumDisp_SS	0.948	0.876	mediumDisp_SS	0.912	0.907
highDisp_SS	0.962	0.814	highDisp_SS	0.954	0.853
lowDisp_OS	0.888	0.937	lowDisp_OS	0.952	0.953
mediumDisp_OS	0.924	0.909	mediumDisp_OS	0.923	0.921
highDisp_OS	0.909	0.932	highDisp_OS	0.928	0.966
(a) signal mass hypothesis $m_N = 1.0$ GeV			(b) signal mass hypothesis $m_N = 1.25$ GeV		
category	Signal efficiency	Background rejection	category	Signal efficiency	Background rejection
lowDisp_SS	0.962	0.944	lowDisp_SS	1.000	0.961
mediumDisp_SS	0.953	0.955	mediumDisp_SS	0.957	0.971
highDisp_SS	0.975	0.941	highDisp_SS	1.000	0.987
lowDisp_OS	0.939	0.969	lowDisp_OS	0.885	0.978
mediumDisp_OS	0.986	0.909	mediumDisp_OS	0.978	0.987
highDisp_OS	0.955	0.941	highDisp_OS	0.976	0.999
(c) signal mass hypothesis $m_N = 1.5$ GeV			(d) signal mass hypothesis $m_N = 1.8$ GeV		

Table 4.9: Signal efficiency and background rejection for the chosen working point of the XGB models, for all signal mass hypotheses and categories.

For visual reference, we take the example of the category mediumDisp_OS and show the ROC curves of the XGB models for all signal mass hypotheses in fig. 4.30. This plot shows the good performance of the XGB models in discriminating signal from background, with AUC values close to 1 for all signal mass hypotheses. It is interesting to remark, as we had noticed visually by looking at the scores distributions (section 4.4.3), that the models perform better for higher signal mass hypotheses, as shown by the higher AUC values for the higher signal mass hypotheses. For completeness, fig. 4.31 shows the same plot for all categories.

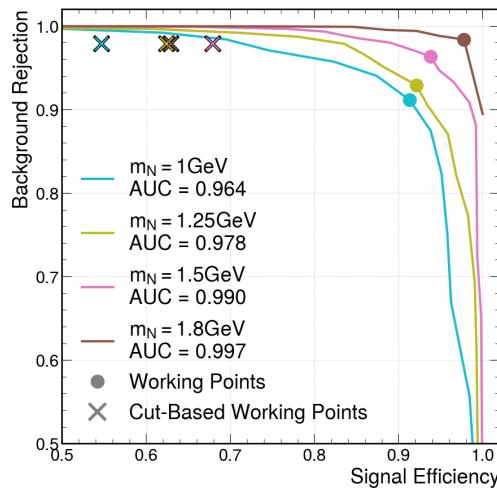


Figure 4.30: ROC curves of XGB models comparing all signal mass hypotheses for the category mediumDisp_OS. The working points of the XGB models are shown as dots on the curves. The crosses show the working points of the cut-based analysis. The legend shows the AUC of each ROC curve.

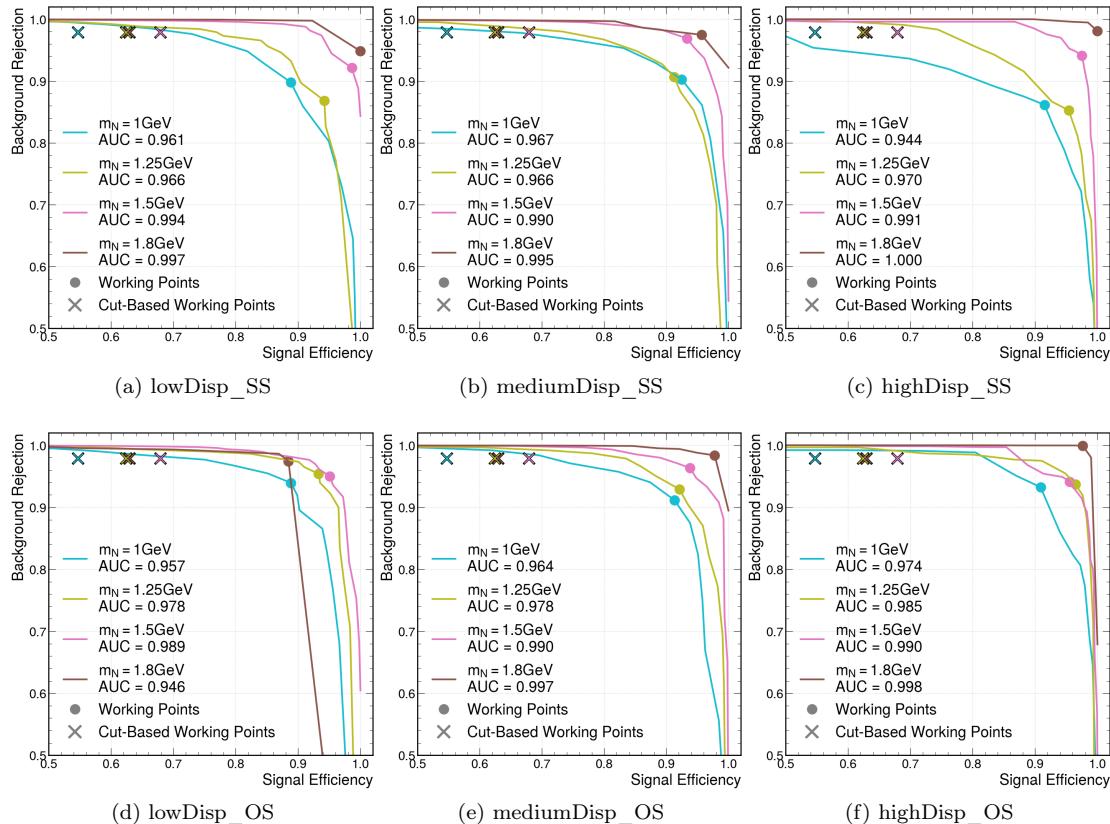


Figure 4.31: ROC curves of XGB models comparing all signal mass hypotheses for each category. The working points of the XGB models are shown as dots on the curves. The crosses show the working points of the cut-based analysis. The legend shows the AUC of each ROC curve.

In order to compare the performance of the XGB models with the current cut-based analysis (described in section 3.3.3), the signal efficiency and background rejection of the cut-based analysis are summarized in table 4.10. For a visual comparison, the working points of the cut-based analysis are also shown as crosses on the previously mentioned figs. 4.30 and 4.31.

m_N (GeV)	Signal efficiency	Background rejection
1.0	54.7 %	
1.25	62.8 %	
1.5	67.9 %	97.9 %
1.8	62.9 %	

Table 4.10: Signal efficiency and background rejection for the cut-based analysis, for all signal hypotheses.

The working points of the XGB models show a significant improvement: they have a signal efficiency roughly 30 percent higher than the cut-based analysis, allowing to reach values of 90 percent or above; the background rejection is only slightly (a few percent) lower.

To quantify the improvement in signal efficiency, we compare the significance of the cut-based selection with the one from this ML approach, summarized in table 4.11. The significance is computed as $S/\sqrt{S+B}$, where S and B are the number of signal and background events selected by each approach.

Signal mass hypothesis (GeV)	Cut-based significance	ML significance	Improvement (%)
1.0	58.69	70.23	20
1.25	62.48	72.8	17
1.5	67.65	76.96	14
1.8	22.13	26.25	19

Table 4.11: Comparison of the cut-based and ML significance ($S/\sqrt{S+B}$) for different signal mass hypotheses. The improvement is computed as the percentage increase in significance from the cut-based to the ML approach.

The ML approach shows a significant improvement in significance over the cut-based analysis, with an increase of the order of 14-20 percent for all signal mass hypotheses. This higher significance in the selection should lead to an improvement in the expected limits on the mixing parameters $|V_{\mu}|^2$ with respect to those that can be obtained with the cut-based analysis [7].

4.5 Summary and outlook

The focus of this work was to improve the selection efficiency of the current cut-based analysis using ML techniques. To summarize, we have trained our models exploring the usage of three different algorithms: XGB, BDT and ANN. We have found that the XGB algorithm provides the models that best discriminate between signal and background. The results of the XGB models show a significant improvement in signal efficiency (reaching values of 90 % or above) compared to the cut-based analysis, while the background rejection is only slightly lower.

This work is part of an ongoing analysis effort within CMS, and there are still opportunities to explore certain areas for improvement.

One margin of improvement for this analysis is the generation of higher statistics MC samples, especially for the signal datasets. This is because the robustness of the ML models is limited by the number of events in the training set, and the signal data samples used in this analysis have low statistics in some categories. As far as background is concerned, an alternative to our approach could be to use a ‘data-driven’ background, obtained from data excluding the signal region of the $\mu\pi$ invariant mass. This has the twofold advantage of relying on data statistics, instead of generating more MC events, and feeding the ML model with data from real data-taking conditions.

The signal MC samples used for the training have been generated with a proper decay length $c\tau = 10$ mm, as a conservative choice that is also consistent with the current cut-based analysis.

It could be worth cross-checking the behavior of the XGB model trained also on different $c\tau$ hypotheses.

From a technical perspective another approach worth investigating is the use of more advanced ML algorithms, such as a parametrized neural network [78], which uses the mass hypothesis parameter as an additional feature. This kind of model can interpolate predictions for any arbitrary value of the signal mass hypothesis, instead of relying on a discrete set of models for a finite number of signal mass hypotheses. This would also reduce the number of models to be trained from N to only one (per category).

Thanks to the higher signal efficiency, the techniques presented in this work will also be used in the future steps of the ongoing CMS analysis to calculate the new expected limits on the mixing parameters $|V_\mu|^2$ of the HNL, which we expect to be improved over the current ones obtained with the cut-based analysis.

Conclusions

This thesis work focused on the use of Machine Learning (ML) techniques for the search of HNLs from D_s meson decays in the CMS experiment. This study has been carried out in the context of an ongoing analysis, that relies on a cut-based event selection, and its aim is to explore a different approach, based on ML techniques, to improve the event selection.

The decay channel we have studied is $D_s^+ \rightarrow N (\rightarrow \mu^\pm \pi^\mp) \mu^+$. The signature of this channel comes from the displaced $N \rightarrow \mu\pi$ vertex whereas the main background is caused by accidental combination of muons and a track into a common vertex. The search is not focused on a single kind of HNL signature: different HNL masses or proper decay length hypotheses can have different signature. The signal datasets that have been used in this work take into account four different HNL mass hypotheses: $m_N = 1.0$ GeV, 1.25 GeV, 1.5 GeV and 1.8 GeV. Since the selection strategy should discriminate signal against background for different signatures, the events are divided into different categories. In particular, we have chosen to use six different event categories based on: the HNL vertex displacement significance, that allows to enhance signals with different proper mean decay length; the relative sign of the muons which is related to lepton number violation (conservation) for same (opposite) sign.

In our ML approach, the models we focus on are supervised classifiers, which are trained to distinguish events of different classes, in our case signal and background. We have explored three different ML algorithms: boosted decision trees (BDT), gradient boosted decision trees (XGB) and artificial neural networks (ANN). The ML models are trained using the Monte Carlo samples (one for background and one for each HNL mass hypothesis) that have also been used in the cut-based selection optimization. These Monte Carlo samples reproduce the data-taking conditions of the Run-2 2018 data collected by the CMS experiment.

We have evaluated several physical variables to be used as inputs for the training of the ML models. We studied the correlation of these variables in order to reduce the number of inputs, and avoid feeding the algorithm with redundant information. In our approach, we have trained one model of each type on each of the six category datasets for each of the four signal mass hypotheses. This process has resulted in a total of 24 signal-background pairs of datasets, thus 72 ML models in total.

After training the models, we have evaluated their performance. The best models we have chosen are the ones built with the XGB algorithm, because compared to the other algorithms, they achieve a better signal efficiency for a fixed background rejection, for all categories and signal mass hypotheses. The ML models provide an output score that is able to discriminate between signal and background. In order to choose the best cut on the model's output score, we have chosen the discovery significance metric. With this approach we have derived the “working points” in the selection efficiency-background rejection plane for all categories and signal mass hypotheses.

The signal efficiencies of the XGB models are 90% or higher, which significantly improves the efficiencies of the cut-based approach, with a small compromise in background rejection. The overall impact of the selection can be evaluated using the significance. The significance is estimated as $S/\sqrt{S+B}$, where $S(B)$ are the number of signal (background) events selected. The improvement in the event selection achieved with the ML approach translates into 14-20% higher significances of the XGB models selection over the cut-based one.

The results of this work show that the developed ML-based strategy is an effective contribution to the ongoing CMS analysis. Therefore, replacing the cut-based selection could lead to an improvement of the upper limits extraction active-sterile neutrino mixing parameter $|V_\mu|^2$.

Bibliography

- [1] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [2] Q. R. Ahmad, R. C. Allen, and T. C. Andersen. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”. In: *Physical Review Letters* 89.1 (June 2002). DOI: 10.1103/physrevlett.89.011301. URL: <https://doi.org/10.1103/physrevlett.89.011301>.
- [3] André de Gouvêa. “Neutrino Mass Models”. In: *Ann. Rev. Nucl. Part. Sci.* 66 (2016), pp. 197–217. DOI: 10.1146/annurev-nucl-102115-044600.
- [4] Takehiko Asaka, Steve Blanchet, and Mikhail Shaposhnikov. “The ν MSM, dark matter and neutrino masses”. In: *Physics Letters B* 631.4 (Dec. 2005), pp. 151–156. DOI: 10.1016/j.physletb.2005.09.070. URL: <https://doi.org/10.1016/j.physletb.2005.09.070>.
- [5] Frank F Deppisch, P S Bhupal Dev, and Apostolos Pilaftsis. “Neutrinos and collider physics”. In: *New Journal of Physics* 17.7 (Aug. 2015), p. 075019. DOI: 10.1088/1367-2630/17/7/075019. URL: <https://dx.doi.org/10.1088/1367-2630/17/7/075019>.
- [6] “Recording and reconstructing 10 billion unbiased b hadron decays in CMS”. In: (2019). URL: <https://cds.cern.ch/record/2704495>.
- [7] Leonardo Lunerti. “Search for heavy neutrinos in D_s decays with the CMS detector at LHC”. PhD thesis. Bologna U., 2023. DOI: 10.48676/unibo/amsdottorato/10924.
- [8] Alexey Boyarsky, Oleg Ruchayskiy, and Mikhail Shaposhnikov. “The Role of Sterile Neutrinos in Cosmology and Astrophysics”. In: *Annual Review of Nuclear and Particle Science* 59.1 (Nov. 2009), pp. 191–214. DOI: 10.1146/annurev.nucl.010909.083654. URL: <https://doi.org/10.1146%2Fannurev.nucl.010909.083654>.
- [9] Matthew Dean Schwartz. *Quantum field theory and the standard model*. New York: Cambridge University Press, 2014. ISBN: 9781107034730.
- [10] Chris Quigg. *Gauge theories of the strong, weak, and electromagnetic interactions*. Second edition. Princeton, New Jersey: Princeton University Press, 2013. ISBN: 9780691135489.
- [11] Carlo Giunti and Chung Wook Kim. *Fundamentals of neutrino physics and astrophysics*. OCLC: ocm76935610. Oxford ; New York: Oxford University Press, 2007. ISBN: 9780198508717.
- [12] M. Goldhaber, L. Grodzins, and A. W. Sunyar. “Helicity of Neutrinos”. In: *Physical Review* 109.3 (Feb. 1958), pp. 1015–1017. DOI: 10.1103/physrev.109.1015. URL: <https://doi.org/10.1103/physrev.109.1015>.
- [13] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.
- [14] B. J. P. Jones. *The Physics of Neutrinoless Double Beta Decay: A Primer*. 2022. arXiv: 2108.09364 [nucl-ex].
- [15] Michael Edward Peskin and Daniel V. Schroeder. *An introduction to quantum field theory*. Reading, Mass: Addison-Wesley Pub. Co, 1995. ISBN: 9780201503975.
- [16] Kyrylo Bondarenko et al. “Phenomenology of GeV-scale heavy neutral leptons”. In: *Journal of High Energy Physics* 2018.11 (Nov. 2018). DOI: 10.1007/jhep11(2018)032. URL: <https://doi.org/10.1007%2Fjhep11%282018%29032>.

BIBLIOGRAPHY

- [17] Elena Graverini, Eric Van Herwijnen, and Thomas Ruf. “Mass dependence of branching ratios into HNL for FairShip”. In: (2016). URL: <https://cds.cern.ch/record/2133817>.
- [18] André de Gouvêa. “Seesaw energy scale and the LSND anomaly”. In: *Phys. Rev. D* 72 (3 Aug. 2005), p. 033005. DOI: 10.1103/PhysRevD.72.033005. URL: <https://link.aps.org/doi/10.1103/PhysRevD.72.033005>.
- [19] M. Cirelli et al. “Probing oscillations into sterile neutrinos with cosmology, astrophysics and experiments”. In: *Nuclear Physics B* 708.1-3 (Feb. 2005), pp. 215–267. DOI: 10.1016/j.nuclphysb.2004.11.056. URL: <https://doi.org/10.1016/j.nuclphysb.2004.11.056>.
- [20] Amand Faessler et al. “Arbitrary mass Majorana neutrinos in neutrinoless double beta decay”. In: *Phys. Rev. D* 90 (9 Nov. 2014), p. 096010. DOI: 10.1103/PhysRevD.90.096010. URL: <https://link.aps.org/doi/10.1103/PhysRevD.90.096010>.
- [21] J. Barea, J. Kotila, and F. Iachello. “Limits on sterile neutrino contributions to neutrinoless double beta decay”. In: *Phys. Rev. D* 92 (9 Nov. 2015), p. 093001. DOI: 10.1103/PhysRevD.92.093001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.92.093001>.
- [22] D. Liventsev, I. Adachi, and H. Aihara. “Search for heavy neutrinos at Belle”. In: *Phys. Rev. D* 87 (7 Apr. 2013), p. 071102. DOI: 10.1103/PhysRevD.87.071102. URL: <https://link.aps.org/doi/10.1103/PhysRevD.87.071102>.
- [23] Andrew Kobach and Sean Dobbs. “Heavy neutrinos and the kinematics of tau decays”. In: *Phys. Rev. D* 91 (5 Mar. 2015), p. 053006. DOI: 10.1103/PhysRevD.91.053006. URL: <https://link.aps.org/doi/10.1103/PhysRevD.91.053006>.
- [24] LBNE Collaboration and Corey Adams. *The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe*. 2014. arXiv: 1307.7335 [hep-ex].
- [25] P. Abreu et al. “Search for neutral heavy leptons produced in Z decays”. In: *Zeitschrift für Physik C Particles and Fields* 74.1 (Apr. 1997), pp. 57–71. DOI: 10.1007/s002880050370. URL: <https://doi.org/10.1007/s002880050370>.
- [26] O. Adriani and M. Aguilar-Benitez. “Search for isosinglet neutral heavy leptons in Z0 decays”. In: *Physics Letters B* 295.3-4 (Dec. 1992), pp. 371–382. DOI: 10.1016/0370-2693(92)91579-x. URL: [https://doi.org/10.1016/0370-2693\(92\)91579-x](https://doi.org/10.1016/0370-2693(92)91579-x).
- [27] Michael Benedikt et al. *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Future Circular Collider*. Tech. rep. 2. Geneva: CERN, 2019. DOI: 10.1140/epjst/e2019-900045-4. URL: <https://cds.cern.ch/record/2651299>.
- [28] G. Aad, B. Abbott, and J. Abdallah. “Search for heavy Majorana neutrinos with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Journal of High Energy Physics* 2015.7 (July 2015). DOI: 10.1007/jhep07(2015)162. URL: <https://doi.org/10.1007%2Fjhep07%282015%29162>.
- [29] V. Khachatryan and A.M. Sirunyan. “Search for heavy Majorana neutrinos in mu+-mu+- + jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *Physics Letters B* 748 (Sept. 2015), pp. 144–166. DOI: 10.1016/j.physletb.2015.06.070. URL: <https://doi.org/10.1016%2Fj.physletb.2015.06.070>.
- [30] *Heavy Neutral Lepton, Search for*. from R.L. Workman et al.(Particle Data Group), Prog.Theor.Exp.Phys.2022, 083C01 (2022) and 2023 update. URL: <https://pdg.lbl.gov/2023/listings/rpp2023-list-heavy-neutral-leptons.pdf>.
- [31] The CMS collaboration, A. M. Sirunyan, and A. Tumasyan. “Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV”. en. In: *Journal of High Energy Physics* 2019.1 (Jan. 2019), p. 122. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2019)122. URL: [https://link.springer.com/10.1007/JHEP01\(2019\)122](https://link.springer.com/10.1007/JHEP01(2019)122) (visited on 02/29/2024).
- [32] The CMS collaboration et al. “Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV”. en. In: *Journal of High Energy Physics* 2022.7 (July 2022), p. 81. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2022)081. URL: [https://link.springer.com/10.1007/JHEP07\(2022\)081](https://link.springer.com/10.1007/JHEP07(2022)081) (visited on 02/29/2024).

BIBLIOGRAPHY

- [33] A. Tumasyan, W. Adam, and J. W. Andrejkovic. “Probing Heavy Majorana Neutrinos and the Weinberg Operator through Vector Boson Fusion Processes in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. Lett.* 131 (1 July 2023), p. 011803. DOI: 10.1103/PhysRevLett.131.011803. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.131.011803>.
- [34] “Search for long-lived heavy neutral leptons with lepton flavour conserving or violating decays to a jet and an electron, muon, or tau lepton”. In: (2023).
- [35] A. Tumasyan et al. “Search for Long-Lived Particles Decaying in the CMS End Cap Muon Detectors in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. Lett.* 127 (26 Dec. 2021), p. 261804. DOI: 10.1103/PhysRevLett.127.261804. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.127.261804>.
- [36] CERN. *The accelerator complex / CERN*. URL: <https://home.cern/science/accelerators/accelerator-complex> (visited on 01/14/2024).
- [37] R Alemany-Fernandez et al. “Operation and Configuration of the LHC in Run 1”. In: (2013). URL: <https://cds.cern.ch/record/1631030>.
- [38] Jorg Wenninger. “Operation and Configuration of the LHC in Run 2”. In: (2019). URL: <https://cds.cern.ch/record/2668326>.
- [39] O. Brüning, H. Burkhardt, and S. Myers. “The large hadron collider”. en. In: *Progress in Particle and Nuclear Physics* 67.3 (July 2012), pp. 705–734. ISSN: 01466410. DOI: 10.1016/j.ppnp.2012.03.001. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0146641012000695> (visited on 01/14/2024).
- [40] The CMS Collaboration et al. “The CMS experiment at the CERN LHC”. en. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08004. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004> (visited on 01/14/2024).
- [41] CMS Collaboration. *Detector / CMS Experiment*. URL: <https://cms.cern/detector> (visited on 01/14/2024).
- [42] The Tracker Group of the CMS Collaboration. *The CMS Phase-1 Pixel Detector Upgrade*. Tech. rep. Geneva: CERN, 2020. URL: <https://cds.cern.ch/record/2745805>.
- [43] L. Viliani. “CMS tracker performance and readiness for LHC Run II”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 824 (2016). Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors, pp. 67–69. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2015.09.046>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900215011018>.
- [44] A.M. Sirunyan et al. “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 15.05 (May 2020), P05002. DOI: 10.1088/1748-0221/15/05/P05002. URL: <https://dx.doi.org/10.1088/1748-0221/15/05/P05002>.
- [45] Hyunyong Kim. “Track-based muon system alignment of the CMS detector”. In: *PoS ICHEP2020* (2021), p. 745. DOI: 10.22323/1.390.0745.
- [46] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- [47] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.020>. URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- [48] V. Khachatryan et al. “The CMS trigger system”. In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. URL: <https://dx.doi.org/10.1088/1748-0221/12/01/P01020>.

- [49] A.M. Sirunyan et al. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. URL: <https://dx.doi.org/10.1088/1748-0221/15/10/P10017>.
- [50] Swagata Mukherjee. “Data Scouting and Data Parking with the CMS High level Trigger”. In: *PoS EPS-HEP2019* (2020), p. 139. DOI: 10.22323/1.364.0139. URL: <https://cds.cern.ch/record/2766071>.
- [51] The CMS collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”. In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. URL: <https://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- [52] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. URL: <https://dx.doi.org/10.1088/1748-0221/9/10/P10009>.
- [53] A.M. Sirunyan et al. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/06/p06015. URL: <http://dx.doi.org/10.1088/1748-0221/13/06/P06015>.
- [54] Andy Buckley et al. “General-purpose event generators for LHC physics”. In: *Physics Reports* 504.5 (2011), pp. 145–233. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2011.03.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157311000846>.
- [55] Albert M Sirunyan et al. “Pileup mitigation at CMS in 13 TeV data”. In: *JINST* 15.09 (2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].
- [56] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [57] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. en. In: *The European Physical Journal C* 71.2 (Feb. 2011), p. 1554. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: <http://link.springer.com/10.1140/epjc/s10052-011-1554-0> (visited on 02/21/2024).
- [58] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2015.01.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442>.
- [59] David J. Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4). URL: <https://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [60] Glen Cowan. “Statistics for Searches at the LHC”. In: *69th Scottish Universities Summer School in Physics: LHC Physics*. July 2013, pp. 321–355. DOI: 10.1007/978-3-319-05362-2_9. arXiv: 1307.2487 [hep-ex].
- [61] The CMS collaboration et al. “Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV”. en. In: *Journal of High Energy Physics* 2022.7 (June 2022), p. 81. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2022)081. URL: [https://link.springer.com/10.1007/JHEP07\(2022\)081](https://link.springer.com/10.1007/JHEP07(2022)081) (visited on 03/03/2024).
- [62] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York, 2009. ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7. URL: <http://link.springer.com/10.1007/978-0-387-84858-7> (visited on 02/20/2024).

BIBLIOGRAPHY

- [63] Christopher M. *Pattern Recognition and Machine Learning*. en. 1st ed. Information Science and Statistics. New York, NY: Springer, Aug. 2006.
- [64] Li Yang and Abdallah Shami. “On hyperparameter optimization of machine learning algorithms: Theory and practice”. In: *Neurocomputing* 415 (Nov. 2020), pp. 295–316. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2020.07.061. URL: <http://dx.doi.org/10.1016/j.neucom.2020.07.061>.
- [65] Tensorflow. *TensorFlow Playground*. 2022. URL: <https://playground.tensorflow.org/> (visited on 01/14/2024).
- [66] Lutz Prechelt. “Early Stopping — But When?” In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 53–67. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_5. URL: https://doi.org/10.1007/978-3-642-35289-8_5.
- [67] Yann Coadou. “Boosted Decision Trees”. In: *Artificial Intelligence for High Energy Physics*. WORLD SCIENTIFIC, Feb. 2022, pp. 9–58. DOI: 10.1142/9789811234033_0002. URL: http://dx.doi.org/10.1142/9789811234033_0002.
- [68] Yoav Freund and Robert E Schapire. “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: <https://doi.org/10.1006/jcss.1997.1504>. URL: <https://www.sciencedirect.com/science/article/pii/S002200009791504X>.
- [69] Jerome H. Friedman. “Greedy function approximation: A gradient boosting machine.” In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: <https://doi.org/10.1214/aos/1013203451>.
- [70] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: <http://dx.doi.org/10.1145/2939672.2939785>.
- [71] Charu Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Jan. 2018. ISBN: 978-3-319-94462-3. DOI: 10.1007/978-3-319-94463-0.
- [72] Vinod Nair and Geoffrey E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines”. In: *International Conference on Machine Learning*. 2010. URL: <https://api.semanticscholar.org/CorpusID:15539264>.
- [73] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *CoRR* abs/1412.6980 (2014). URL: <https://api.semanticscholar.org/CorpusID:6628106>.
- [74] Jim Pivarski et al. *Uproot*. Version v5.3.0rc1. Feb. 2024. DOI: 10.5281/zenodo.10659250. URL: <https://doi.org/10.5281/zenodo.10659250>.
- [75] ROOT Team. *ROOT Files*. URL: https://root.cern/manual/root_files/ (visited on 01/14/2024).
- [76] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [77] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/>.
- [78] Tommaso Diotallevi. “Application of Deep Learning techniques in the search for BSM Higgs bosons in the $\mu\mu$ final state in CMS”. PhD thesis. alma, June 2022. URL: <http://amsdottorato.unibo.it/10356/>.

BIBLIOGRAPHY

Appendix A

Monte Carlo

After event generation, additional conditions are required for the event to be finally stored. These requirements depend on the generated samples and usually are requirements on p_T and η of the final-state particles to make them fall into detector acceptance and to meet the trigger conditions, or requirements for final-state particles to be generated from the decay of interest. These requirements are often referred to as generator filters, and the fraction of events that survive such requirements is called filter efficiency. The simulated samples used in the analysis are summarized in table A.1. For each sample we report the number of events that pass the generator filters, the cross-section and the filter efficiency. The reported cross-section was also calculated using the PYTHIA software, and it refers to the production process only. A summary of the generator settings and filters used for each of the simulated processes can be found in Tab. A.2.

Process	N_{TOT}	σ (pb)	$\varepsilon_{\text{filter}}$ (%)
$D_s^\pm \rightarrow N(\rightarrow \mu\pi)\mu^\pm$			
$m_{\text{HNL}} = 1.8 \text{ GeV}; c\tau_{\text{HNL}} = 10 \text{ mm}$	309857	—	0.005
$m_{\text{HNL}} = 1.8 \text{ GeV}; c\tau_{\text{HNL}} = 100 \text{ mm}$	311533	—	0.005
$m_{\text{HNL}} = 1.8 \text{ GeV}; c\tau_{\text{HNL}} = 1000 \text{ mm}$	309867	—	0.004
$m_{\text{HNL}} = 1.5 \text{ GeV}; c\tau_{\text{HNL}} = 10 \text{ mm}$	215946	—	0.005
$m_{\text{HNL}} = 1.5 \text{ GeV}; c\tau_{\text{HNL}} = 100 \text{ mm}$	213340	—	0.005
$m_{\text{HNL}} = 1.5 \text{ GeV}; c\tau_{\text{HNL}} = 1000 \text{ mm}$	214422	—	0.003
$m_{\text{HNL}} = 1.25 \text{ GeV}; c\tau_{\text{HNL}} = 10 \text{ mm}$	211501	—	0.004
$m_{\text{HNL}} = 1.25 \text{ GeV}; c\tau_{\text{HNL}} = 100 \text{ mm}$	211400	—	0.004
$m_{\text{HNL}} = 1.25 \text{ GeV}; c\tau_{\text{HNL}} = 1000 \text{ mm}$	213100	—	0.004
$m_{\text{HNL}} = 1.0 \text{ GeV}; c\tau_{\text{HNL}} = 10 \text{ mm}$	277288	—	0.004
$m_{\text{HNL}} = 1.0 \text{ GeV}; c\tau_{\text{HNL}} = 100 \text{ mm}$	255234	—	0.005
$m_{\text{HNL}} = 1.0 \text{ GeV}; c\tau_{\text{HNL}} = 1000 \text{ mm}$	259243	—	0.003
$D_s^\pm \rightarrow \phi(\mu\mu)\pi^\pm$	2964234	—	0.0160
Muon enriched QCD dijet			
$20 < \hat{p}_T < 30 \text{ GeV}$	60640516	3.977×10^8	0.65
$30 < \hat{p}_T < 50 \text{ GeV}$	58627984	1.070×10^8	1.25
$50 < \hat{p}_T < 80 \text{ GeV}$	40022458	1.573×10^7	2.35
$80 < \hat{p}_T < 120 \text{ GeV}$	45566475	2.341×10^6	3.81
$120 < \hat{p}_T < 170 \text{ GeV}$	39114558	4.086×10^5	5.17
$170 < \hat{p}_T < 300 \text{ GeV}$	71926577	1.036×10^5	6.83

Table A.1: Total events that pass generator filters(N_{TOT}), cross-section (σ) and filter efficiency ($\varepsilon_{\text{filter}}$) of simulated samples. Whenever the cross-section is not used in the analysis, it is indicated with the “—” symbol.

Process	PYTHIA settings	EVTGEN settings	Generator filters
$D_s \rightarrow N(\rightarrow \mu\pi)\mu$	SoftQCD:inelastic = on	$\mathcal{B}(D_s \xrightarrow{\text{PHSP}} N\mu) = 1.0$ $\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^+\pi^-) = 0.5$ $\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^-\pi^+) = 0.5$	Final state μ having: At least a $N \rightarrow \mu^+\pi^-$ decay having: <ul style="list-style-type: none"> • $p_T^\mu > 6.8;$ • $\eta^\mu < 1.55.$
$D_s \rightarrow \phi(\rightarrow \mu\mu)\pi$	SoftQCD: inelastic = on	$\mathcal{B}(D_s \xrightarrow{\text{SVS}} \phi\pi) = 1.0$ $\mathcal{B}(\phi \xrightarrow{\text{VLL}} \mu\mu) = 1.0$	At least a D_s . Final state μ having: <ul style="list-style-type: none"> • $p_T > 2.7 \text{ GeV};$ • $\eta < 2.45.$
QCD dijet	ParticleDecays: limitTau0 = off ParticleDecays: limitCylinder = on ParticleDecays: xyMax = 2000 ParticleDecays: zMax = 4000 HardQCD: all = on PhaseSpace: pTHatMin = p_T^{\min} PhaseSpace: pTHatMax = p_T^{\max} 130: mayDecay = on 211: mayDecay = on 321: mayDecay = on	-	Final state μ having: <ul style="list-style-type: none"> • $p_T^\mu > 5.;$ • $\eta^\mu < 2.5;$ • decay length < 2 m on xy; • decay length < 4 m on z.

Table A.2: Summary of the generator parameters and generator filter settings used for the simulation of each of the processes.