UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

A thesis submitted in fulfillment of the requirements for the degree of:

Master's of Science in Nuclear and Particle Physics

The Hunt for Heavy Neutrinos:

Machine Learning Techniques to Probe Heavy Neutral Leptons in the $e\nu jj$ Final State at the e^+e^- Future Circular Collider

Student:

Thomas M. Critchley
Student ID: 22-328-561

Supervisors:

Prof. Anna Sfyrla
Dr. Pantelis Kontaxakis

Département de Physique Nucléaire et Corpusculaire

Submitted: 21 June 2024

Acknowledgments

I would like to thank my Supervisors, Prof. Anna Sfyrla and Dr. Pantelis Kontaxakis for their invaluable support, knowledge and guidance throughout this process: ευχαριστώ πολύ! Thanks are also due to everyone within the research group, for a very fun and supportive environment, and for inspiration to continue into research. I would also like to thank the BSM LLPs at the FCC-ee group at CERN for introducing me to the broader context of my research. To my partner, Willow, and to my friends in Geneva: thank you for all the fun; we had a lot of it! A special thanks to Sebastien Roy-Garand for his expertise in HNL studies at the LHC, and to Mathias Moors for keeping me in good company throughout this degree.

Of course, to my parents, siblings and our constantly growing family: thank you so very much. Dad, I am sorry for choosing to study in such an expensive city. Lastly, thanks are due to Jilly and Anthony Payne for their work in getting me here, and to Christine Schneider and her kind family for welcoming me so wonderfully to this city.

Abstract

In place of traditional cut-and-count analysis, machine learning (ML) techniques offer powerful methods to optimise our searches for new physics. At the FCC-ee, we will probe the highest intensities and energies ever seen at a lepton collider, opening the possibility for the direct discovery of massive right-handed neutrino states. In this work, existing searches for heavy neutral leptons (HNLs) at the FCC-ee are optimised using a Boosted Decision Tree (BDT) and a Deep Neural Network (DNN) for a mass range of 10 GeV $\leq m_N \leq$ 80 GeV with mixing angles $10^{-4} \leq |U_{eN}|^2 \leq 10^{-10}$. We find that both ML approaches are capable of outperforming cut-and-count methods by orders of magnitude in the sensitivity to HNL couplings. Specifically, in the 95% confidence level (CL) limit, the BDT models achieve sensitivity beyond $|U_{eN}|^2 = 10^{-9}$ for the electron dijet final state at 150 ab^{-1}. Furthermore, we demonstrate the value of ML methods in distinguishing between prompt and long-lived HNL signals in our parameter space by training BDT models based on the transverse impact parameter significance.

Contents

1 Introduction **4**

2 Heavy Neutral Leptons **5**

2.1 The Standard Model 5

2.1.1 The Dirac Equation and Chirality 6

2.1.2 The Higgs Mechanism 9

2.2 Beyond the Standard Model 11

2.2.1 Neutrino Oscillations 11

2.2.2 Neutrino Masses 14

2.3 The Seesaw Mechanism 15

2.3.1 The Vanilla seesaw 15

2.3.2 Inverse Seesaw - Dirac HNLs 17

2.4 Solutions to Problems Beyond the Standard Model 18

2.5 Present Searches for HNLs 20

3 The Future Circular Collider **23**

3.1 FCC-ee 25

3.1.1 The IDEA Detector 26

3.2 HNLs at the FCC-ee: Experimental Outlook 28

4 Simulation **33**

4.1 Software setup 33

4.2 Jet algorithms 35

4.3 Sample Generation 36

4.3.1 Background Samples 36

4.3.2 Signal Sample Generation 36

4.3.3 Reconstructed Events 37

5 Analysis **39**

5.1 Normalisation 39

5.2 Sensitivity 40

5.3 Discriminating Variables 42

5.3.1 Angular Variables 43

5.3.2 Energy Variables 47

5.3.3 Vertex Variables 49

5.4 Cut and Count 53

5.5 Boosted Decision Trees 56

5.5.1 Data Preparation 58

5.5.2 Training and Testing . . . 60
5.5.3 Results . . . 68
5.6 Deep Neural Network . . . 70
5.6.1 Data Preparation . . . 71
5.6.2 Training and validation . . . 74
5.6.3 Results . . . 79
5.7 Long Lived vs Prompt HNLs . . . 82

6 Results and Discussion **86**
6.1 Sensitivity study . . . 86
6.2 Outlook . . . 87

7 Conclusion **88**

Bibliography **89**

A Additional Theory **99**
A.1 Gamma Matrices of the Dirac Equation . . . 99
A.2 Neutrino Oscillation Survival Probability . . . 100

B Process Cards **102**
B.1 MG5 Process Card for Dirac HNLs . . . 102
B.2 MG5 Process Card for 4-body background . . . 103

C Additional Event Statistics **104**
C.1 Signal Sample Cross Sections . . . 104
C.2 Filter Efficiencies . . . 106

1 Introduction

When CERN announced the discovery of the Higgs Boson in July 2012 [1, 2], many reported the Standard Model (SM) as being *complete* [3]. This discovery honored the prediction made by Robert Brout, François Englert, and Peter Higgs in 1964, validating the massive nature of the SM bosons through the spontaneous breaking of the electroweak symmetry. However, describing the SM as complete suggests that the story is finished.

Far from it, several vital questions remain unanswered by the SM, including the issues of neutrino mass, dark matter, and the dominance of matter over antimatter in the universe. The Large Hadron Collider (LHC) at CERN is actively searching for new physics (NP) beyond the SM (BSM). However, there is currently no indication of supersymmetry or any other exotic solutions to BSM phenomena. The High-Luminosity LHC (HL-LHC) will begin taking data at the earliest in 2029 [4], and only time will tell if the significantly increased data volume will reveal signs of new physics. For now, we must look ahead to the future landscape of particle physics in Europe.

The electron-positron future circular collider (FCC-ee) is a proposed initiative at CERN, set to become the highest energy lepton collider ever built. As the largest and highest-intensity collider of its kind, it will function both as a machine for precision measurements and as a discovery tool for NP. During the Z-pole run of the FCC-ee, approximately 10^{12} Z bosons will be produced in a clean, pileup-free environment, serving as an electroweak factory operating with unprecedented precision.

One interesting possibility of the FCC-ee is its potential to discover a new type of particle that addresses existing Standard Model tensions: heavy neutral leptons (HNLs). The discovery of non-zero neutrino masses has exposed a significant problem in our understanding of neutrinos as massless objects. By introducing a sister particle to the neutrino, we propose a sterile right-handed object heavy enough to generate tiny SM neutrino masses via a low scale Type I seesaw mechanism.

In this thesis, we explore the discovery potential of such a particle in the electron di-jet final state during the Z run of the FCC-ee. Chapter 2 introduces the SM and explores the mechanism through which HNLs can be integrated into our current particle physics theories. Chapter 3 examines the phenomenology of these objects within one proposed FCC-ee detector design, IDEA. Chapter 4 provides a review of the simulation details. Chapter 5 constructs a comprehensive overview of the analysis, including a summary of existing cut-and-count methods and an extension into the use of machine learning (ML) techniques such as boosted decision trees (BDTs) and deep neural networks (DNNs). Chapter 6 presents an overview of the results in the context of existing and future expectations for HNL reach by collider experiments.

2 Heavy Neutral Leptons

> *... this experience made a deep and lasting impression upon me. Something deeply hidden had to be behind things.*
>
> – Albert Einstein

The Standard Model is the most complete theory of fundamental physics ever devised, exemplified by numerous experimental confirmations, such as that of the anomalous magnetic moment of the electron [5], and crowned by the discovery of the Higgs Boson. Yet, in spite of this, successive experimental blows have shown us definitively that SM is not capable of describing nature fully. The observation of neutrino masses demonstrated by baseline neutrino oscillation measurements [6, 7, 8, 9] show us decisively that neutrinos are massive, which is not predicted by the SM. Something deeply hidden must be at work, and it will be the aim of experimental physicists over the next decades to discover it.

Interactions of the SM neutrinos with a unseen massive counterpart, a heavy neutral lepton (HNL), is one proposed method for explaining the origin and scale of neutrino masses. In this section, we will explore the SM, neutrino oscillation experiments and the Seesaw Mechanism with sights to setup our search for HNLs.

2.1 The Standard Model

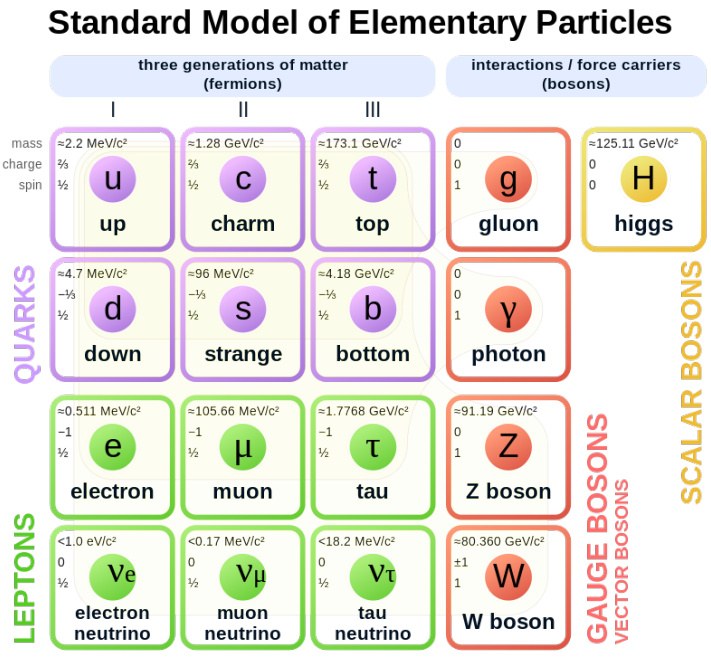


Figure 1: The Standard Model of Particle Physics (Image: Available at Ref. [10]).

The Standard Model of Particle Physics represents the combination of two pillars of twentieth century physics: special relativity and quantum mechanics. It seeks to explain how objects on the smallest scale move when at speeds approaching the universal speed limit. The SM encodes our current understanding of the physical universe, providing a picture of all elementary particles, and indeed of the forces governing their interactions, themselves explained via particle exchange. All matter is composed of *fermions*, and all forces exchanged via *bosons*. A complete picture of the SM is given in Figure 1.

Fermions are distinguished as being *quarks* or *leptons*. The quarks make up atomic matter, such as protons and neutrons, and do not exist as isolated particles, but instead *always* form hadrons, distinguished as *mesons* with two quarks and *baryons* with three. The leptons, however, exist freely in nature and do not hadronise, such as the electrons, orbiting the nucleus in atomic structures.

The bosons, commonly referred to simply as the force-carriers, mediate the interactions between particles in the field theories, which give rise to the mathematics of the SM. They are denoted as gauge bosons or scalar bosons depending on their *spin* properties. The only observed scalar boson is the Higgs Boson, which is coincidentally the most recent experimental addition to the SM.

The SM is governed by three main field theories: Quantum Electrodynamics (QED), Quantum Chromodynamics (QCD), and the weak interaction. QED describes the electromagnetic interactions mediated by photons. QCD explains the interactions between quarks, which are governed by gluons. The weak interaction, responsible for processes involving nuclear decay, is mediated by the W and Z bosons. Indeed, the neutrinos are known to *only* interact via the weak force. QED has been unified with the weak interaction to form the Electroweak Theory (EWT) [11].

The Higgs boson plays a crucial role in the Standard Model (SM). Without the Higgs mechanism, the masses of fermions and bosons would violate the local gauge symmetry, causing the SM to break down. Simply put, the Higgs mechanism allows us to preserve symmetry by treating SM particles as massless. The interaction strength with which particles couple to the Higgs, known as their *Yukawa coupling*, can be conceptualized as the 'drag' they experience through the Higgs field, which manifests as mass.

2.1.1 The Dirac Equation and Chirality

The dyanamics of spin-half, massive, relativistic particles are described by the covariant Dirac Equation given in Equation 1, where $\gamma^\mu = (\gamma^0, \gamma^1, \gamma^2, \gamma^3)$, ∂_μ is the covariant derivative given by $(\partial_0, \partial_1, \partial_2, \partial_3) \equiv \left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$, and ψ is the wave-function solution to the equation, describing particles or antiparticles with mass m. The γ matrices come

directly from the covariant formulation of the Dirac Equation in Equation 1, and their derivations are found in Appendix A.1.

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0. \tag{1}$$

The plane-wave solutions, $\psi(\mathbf{x}, t)$, of Equation 1 are given by $\psi(\mathbf{x}, t) = u(E, \mathbf{p})e^{i(\mathbf{p}\cdot\mathbf{x}-Et)}$, where $u(E, \mathbf{p})$ is known as the Dirac spinor with four components describing particles and antiparticles with spin $\pm 1/2$. The antiparticles are those with E < 0 solutions, as per the Feynman-Stückelberg interpretation[1], and are often denoted with ν_1 and ν_2 as opposed to the particle states u_1 and u_2; in this regime, the plane-wave solutions are written as $\psi_i = u_i e^{+i(\mathbf{p}\cdot\mathbf{x}-Et)}$ for particles, and $\psi_i = \nu_i e^{-i(\mathbf{p}\cdot\mathbf{x}-Et)}$ for antiparticle states.

Particles within the SM are imbued with certain properties, many of which are familiar to every day life, such as electrical charge. One such property is known as *chirality*. Unlike the similar concept of helicity, which can be thought of as simply the projection of spin onto the direction of the particle momentum, chirality is much more subtle. Helicity, though an important concept, is not Lorentz invariant. This means that for massive particles, it is always possible to Lorentz boost to a frame in which the helicity is inverted. Chirality, on the other hand, is Lorentz invariant and is best understood abstractly as the eigenstates[2] of the γ^5 matrix, defined in Equation 2.

$$\gamma_5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \tag{2}$$

The eigenstates of the γ^5-matrix are defined as left- and right-handed chiral states such that we may write: $\gamma^5 u_R = +u_R$ and $\gamma^5 u_L = -u_L$, $\gamma^5 v_R = -v_R$ and $\gamma^5 v_L = +v_L$. In the relativistic limit, $E \gg m$, these eigenstates correspond to solutions of the Dirac equation.

In general, all Dirac Spinors can be decomposed into right- or left-handed chiral states using the operators P_R and P_L shown in Equation 3. Certain combinations of chiral operators dictate the nature of a given interaction.

$$P_R = \frac{1}{2}(1 + \gamma_5), \quad P_L = \frac{1}{2}(1 - \gamma_5). \tag{3}$$

[1] The Feynman-Stückelberg interpretation elicits that the negative energy solutions to the Dirac equation are physical and correspond to positive energy solutions moving backwards through time.

[2] Eigenstates are special states of a quantum system that satisfy the equation $\hat{O}\psi = \lambda\psi$, where $\hat{O}$ is an operator (in this case, the γ^5 matrix) and λ is a scalar known as an eigenvalue and ψ is the eigenstate.

These operators obey the relations $P_R+P_L = 1$, $P_R \cdot P_R = P_R$, $P_L \cdot P_L = P_L$, $P_L \cdot P_R = 0$. P_R projects out right-handed chiral particle states and left-handed chiral antiparticle states, whilst P_L serves to do the inverse. This can be summarised by the relations: $P_R u_R = u_R$, $P_R u_L = 0$, $P_R \nu_R = 0$, $P_R \nu_L = \nu_L$ for the right-handed operator, and $P_L u_R = 0$, $P_L u_L = u_L$, $P_L \nu_R = \nu_R$, $P_L \nu_L = 0$ for the left-handed operator.

In QED, the nature of interactions between charged fermions and photons is given by the form: $\bar{\psi}\gamma^\mu\phi$ where ψ and ϕ are Dirac spinors, with $\bar{\psi}$ representing an antiparticle spinor. The 4-vector current can be decomposed into left and right chiral states using the operators in Equation 3, generating terms such as $\bar{u}_L(p)\gamma^\mu u_R(p')$. We saw earlier that the right-handed chiral operator leaves a right-handed spinor unchanged, $u_R(p') = u_R(p')P_R$ and for the antiparticle state it can be shown using the the relations of the gamma matrices that $\bar{u}_L(p) = \bar{u}_L(p)P_R$.

Considering again the term $\bar{u}_L(p)\gamma^\mu u_R(p')$, with no change to our original expression we can write $\bar{u}_L(p)\gamma^\mu u_R(p') = \bar{u}_L(p)P_R\gamma^\mu P_R u_R(p')$. The gamma matrices are such that $\gamma^5\gamma^\mu = -\gamma^5\gamma^\mu$ and hence: $P_R\gamma^\mu = \frac{1}{2}(1+\gamma^5)\gamma^\mu = \gamma^\mu\frac{1}{2}(1-\gamma^5) = \gamma^\mu P_L$. Thus, finally, we see that $\bar{u}_L(p)\gamma^\mu u_R(p') = \bar{u}_L(p)P_R P_L u_R(p') = 0$. That is to say: only specific combinations of chiral eigenstates give non-zero interactions. For QED, these are interactions involving same chiral signs, i.e. two left-handed states (LL) or two right-handed states (RR), but never a mix of the two.

This is the case for QED, which has a *vector* like interaction $\bar{\psi}\gamma^\mu\phi$. For the weak interaction, the vertex is a combination of vector and axial-vector (V-A), which we understand since the weak interaction is known to violate parity conservation[3]. Hence, the interaction vertex has the form $u(p')\frac{1}{2}\gamma^\mu(1-\gamma^5)u(p)$. We recognise this as containing the left-handed chiral operator, $P_L = \frac{1}{2}(1-\gamma^5)$. As a consequence, under the same treatment of the vertex as we did for QED, the only non-zero surviving terms now exclusively involve left-handed chiral *particles* and right-handed *antiparticles*.

Due to the V-A nature of the weak interaction, no right-handed (left-handed) particle (antiparticle) states interact via the weak force. In the case of the neutrinos, which only interact via the weak force, this means that any postulated right handed neutrino must be completely sterile. In this context, sterile is meant to mean that this object would only interact via the Higgs and gravity, and would have no coupling to the weak, strong or electromagnetic forces. The result of this is that the SM does not currently admit the existence of right-handed neutrinos, since we cannot probe them in known SM

[3]Parity refers to the symmetry of physical laws under spatial inversion (i.e., flipping the sign of spatial coordinates). Parity violation in weak interactions, discovered in the 1950s, showed that nature does not treat left-handed and right-handed particles the same way. This was a groundbreaking discovery because it demonstrated that the weak force does not conserve parity, unlike other fundamental forces such as electromagnetism.

interactions.

2.1.2 The Higgs Mechanism

The dynamics and properties of all objects within the Standard Model are encoded into its Lagrangian. The Lagrangian is a compact representation of the underlying dynamics of the SM. At its simplest level, a classical Lagrangian is given by L = T - U, as the difference between the kinetic and potential energies. Using the principle of least action allows us to find the Euler-Lagrange equations to elegantly describe the mechanics. For the SM Lagrangian, we actually refer to the Lagrangian *density*, which is such that we integrate the Lagrangian over a differential volume element, $\mathcal{L} = \int L d^3x$. This formulation is much more useful in quantum mechanical systems, where the number of players in a given interaction is not fixed.

The Standard Model Lagrangian is given below in Equation 4 where *h.c.* denotes the hermitian conjugate term, e.g. for antiparticles as opposed to particle objects:

$$\mathcal{L}_{\text{SM}} = \underbrace{-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{(a) Gauge Fields}} + \underbrace{i\bar{\psi}\gamma^{\mu}D_{\mu}\psi}_{\text{(b) Fermion Kinetic}} - \underbrace{y\bar{\psi}_L\phi\psi_R + \text{h.c.}}_{\text{(c) Yukawa Interactions}} + \underbrace{|D_{\mu}\phi|^2 - V(\phi)}_{\text{(d) Higgs Potential}}. \tag{4}$$

It is beyond the scope of this thesis to describe the Higgs Mechanism in full detail[4], but for the purposes of understanding the neutrino masses presented in Section 2.2, it is necessary to understand the principle of how the other SM fermions derive their mass. Figure 2 below shows once again a depiction of the players within the SM with an emphasis on the role of the electroweak symmetry breaking (EWSB) and the Higgs non-zero vacuum expectation on the generation of the masses within the SM.

The SM is constructed on the principles of Lorentz invariance and local gauge symmetries. Lorentz invariance ensures that the laws of physics remain the same under rotations and boosts in spacetime, while local gauge symmetries guarantee the consistency of interactions at every point in space and time. The SM is based on the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$, which describes the interactions of particles through the strong, weak, and electromagnetic forces, respectively.

The masses of massive SM particles are generated via the EWSB of the Higgs mechanism. The SM Lagrangian is built upon steadfast principles of Lorentz invariance and local gauge symmetries. For the vector bosons, mass is introduced via the spontaneous

[4]Full derivations can be found in e.g. Griffiths, David J. Introduction to Elementary Particles. John Wiley & Sons, 2008. or Halzen, Francis, and Alan D. Martin. Quarks and Leptons: An Introductory Course in Modern Particle Physics. John Wiley & Sons, 1984.

symmetry breaking of the $U(1)_Y \times SU(2)_L$ group. After this breaking, the W and Z bosons become massive, while the gluons and photons remain massless. This mechanism is crucial for maintaining the consistency and predictive power of the SM.

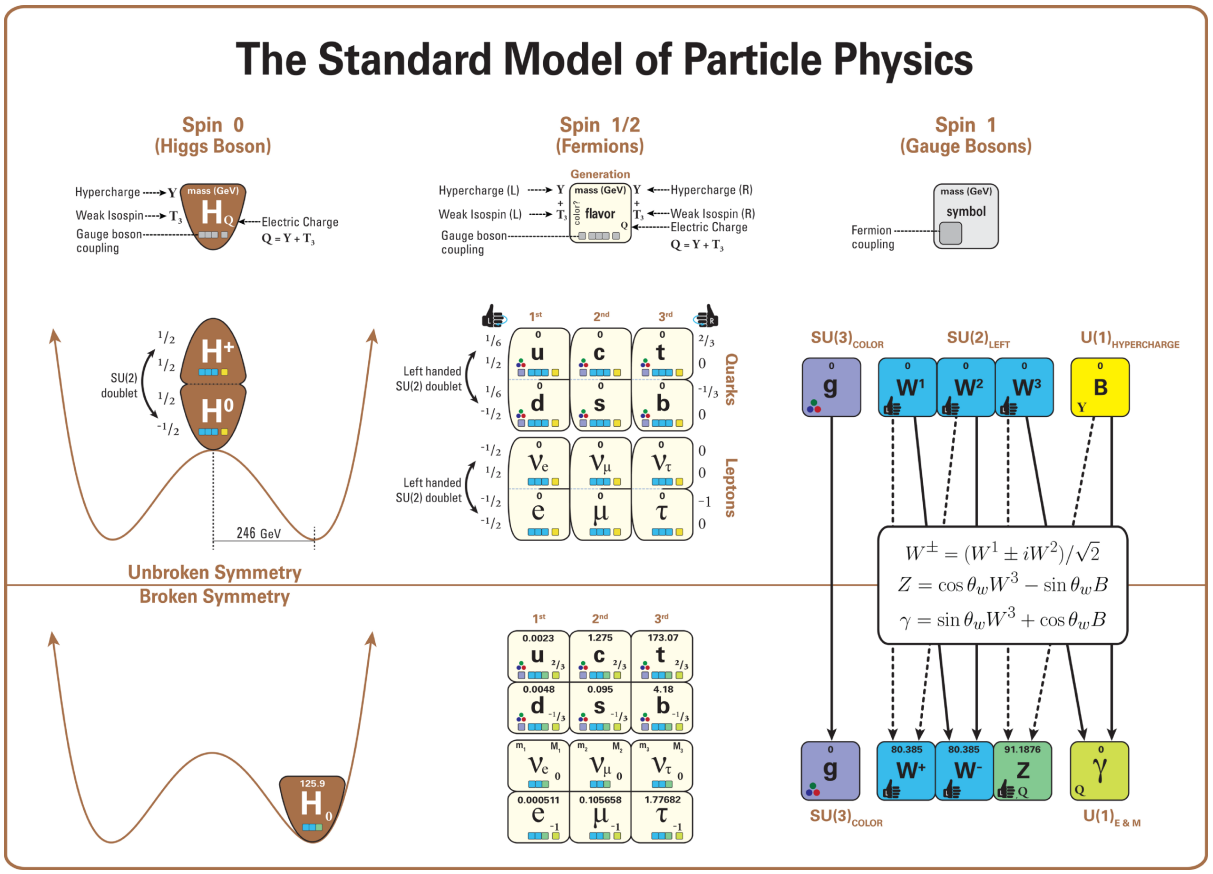


Figure 2: Diagram of the Standard Model with emphasis on the role of the Higgs Mechanism and EWSB (Image: L. Boyle, 2014).

The fermions similarly derive their masses from the Higgs Mechanism, necessarily since the fermion mass term in the Dirac Lagrangian: $-m\bar{\psi}\psi = -m\left(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R\right)$, does not respect the $U(1)_Y \times SU(2)_L$ gauge symmetry. The crux of the mechanism is that after the EWSB, there exists a coupling of left-handed and right-handed massless fermions though the interaction with the non-zero expectation value of the Higgs field. Since, however, the non-zero VEV of the Higgs only occurs for the lower component of the Higgs doublet - such a mechanism can only generate masses for the fermion in the lower component of an $SU(2)_L$ doublet, e.g. the charged leptons (i.e. not the neutrinos) and the down-type quarks.

For the up-type quarks and the neutrinos, a subtlety arises: it is necessary to construct a conjugate doublet ϕ_c to ensure that the corresponding mass terms respect gauge symmetry. A conjugate doublet is formed by applying charge conjugation to the Higgs field, effectively flipping the sign of its hypercharge. Thus, for all Dirac fermions, gauge-

invariant mass terms can be constructed as shown in Equation 5, where g_f represents the fermion Yukawa coupling to the Higgs, L and R respectively denote chiral left and right spinors, and ϕ represents the Higgs field.

$$\mathcal{L} = -g_{\mathrm{f}} \left[\bar{L}\phi R + (\bar{L}\phi R)^{\dagger}\right] \quad or \quad \mathcal{L} = -g_{\mathrm{f}} \left[\bar{L}\phi_c R + (\bar{L}\phi_c R)^{\dagger}\right] \tag{5}$$

After the EWSB, the Yukawa couplings of the fermions, g_f are given by the relationship between the mass of the fermion, m_f and the Higgs VEV, $\nu = 246\,\mathrm{GeV}$. This is shown below in Equation 6. There is no a priori way to motivate the value of the Yukawa couplings, and they are free-parameters of the Standard Model, ascertained via measurements of the fermion masses in particle experiments. Indeed, the large number of free parameters entering the SM serve to undermine its place as a final theory for elementary physics. For the top quark, it is somewhat interesting that the Yukawa coupling is almost exactly unity, whilst the smallest coupling comes from the electron with $\mathcal{O}(10^{-6})$.

$$g_f = \sqrt{2}\ \frac{m_f}{\nu} \tag{6}$$

We saw in the previous section that the nature of the weak interaction dictates that there are no right-handed (left-handed) neutrinos (anti-neutrinos) that can participate in the weak interaction. Since both a left and right-handed state are required to generate the mass terms per Equation 5, we cannot introduce a Dirac mass for the SM neutrinos without the admission of a presently undetected right-handed state.

As it stands, the SM does not admit this and hence neutrinos are assumed to be massless in direct conflict with the observation of neutrino oscillations. For further discussion, we must explore mechanisms currently beyond the SM.

2.2 Beyond the Standard Model

2.2.1 Neutrino Oscillations

When we study a beam of just one neutrino flavour, over a sufficient distance we are able to observe that the beam is no longer pure, but is instead a combination of several different flavour states. This is explained via *neutrino oscillations* [12, 13, 14].

In particle physics, the *mass eigenstates* are the stationary states that obey the Schrödinger equation, and can be thought of as the mathematically fundamental object arising from the predictions of the SM. They are labelled ν_1, ν_2 and ν_3.

There is no a priori reason one should believe that these mathematical objects correspond

to the *observed* neutrino states, coined as *weak eigenstates*. These weak eigenstates are those we observe with the corresponding flavour of lepton. They are denoted as ν_e, ν_μ and ν_τ.

Introducing this subtle but critical difference affords us the freedom to say that we do not know which of the mass eigenstates couples to the corresponding lepton in the weak interaction. As is known in quantum mechanics, we therefore consider the weak eigenstate to be a linear, coherent combination of the mass eigenstates. The neutrino state that obeys the mechanics of the Schrödinger equation thus propagates as this superposition until the wavefunction collapses into an observed weak eigenstate. We can express the relationship between the two eigenstates via Equation 7, where the elements U_{ij}, $i = e, \mu, \tau$, $j = 1, 2, 3$ convey the values of the unitary matrix elements describing the relationship. The unitary matrix entering Equation 7 is referred to as the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{7}$$

In this way, we might decompose the weak eigenstate ν_e into the combination of the mass eigenstates as in Equation 8, where $|\nu_i\rangle$ is the weak eigenstate for a given lepton flavour i=e, μ, τ and U^*_{ij} represents the complex conjugate of the elements entering Equation 7:

$$|\nu_i\rangle = U^*_{i1}|\nu_1\rangle + U^*_{i2}|\nu_2\rangle + U^*_{i3}|\nu_3\rangle. \tag{8}$$

If the mass of all three of these mass eigenstates were equal, there would be no phase difference between ν_1, ν_2 and ν_3. A phase difference arises due to the different masses and energies of the neutrino mass eigenstates, which cause them to evolve differently over time. This evolution can be described by a phase factor $\exp(-iE_jt)$ for each mass eigenstate $|\nu_j\rangle$, where E_j is the energy of the eigenstate and t is time. The result of this is that, if the three masses were equal, a given flavour eigenstate would remain unchanged as it propagates through time. If this were the case, we would not observe neutrino flavour beams to oscillate. The PMNS matrix is known to be of the form in Equation 9, where θ_{ij} are the mixing angles between different neutrino flavours,$s_{ij} = \sin(\theta_{ij})$, $c_{ij} = \cos(\theta_{ij})$ and δ is a complex phase responsible for CP violation[5].

[5]CP violation refers to the violation of the combined symmetry of charge conjugation (C) and parity (P). It is a phenomenon that allows the distinction between matter and antimatter.

$$\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \quad (9)$$

The PMNS matrix is a special type of matrix in that it is unitary, meaning that $U^{-1} = U^{\dagger} \equiv (U^{*})^{T}$. This is a key requirement in quantum mechanics since the probabilities must add to 1, forcing our matrix to take a unitary form such that e.g. $|U_{e1}|^2 + |U_{e2}|^2 + |U_{e3}|^2 = 1$. In light of the unitary nature of the PMNS matrix, we can place nine constraints on the parameters entering, since $U^{\dagger}U = I$. To get the time evolution of the state $|\nu_i\rangle$ entering Equation 8, we take the time evolution of the mass eigenstates. From here, we can use our unitary constraints to show that[6] the survival probability of an electron neutrino state is given by Equation 10.

$$\begin{aligned} P\left(\nu_{\mathrm{e}} \to \nu_{\mathrm{e}}\right) = 1 &- 4\left|U_{\mathrm{e1}}\right|^2 \left|U_{\mathrm{e2}}\right|^2 \sin^2 \Delta_{21} \\ &- 4\left|U_{\mathrm{e1}}\right|^2 \left|U_{\mathrm{e3}}\right|^2 \sin^2 \Delta_{31} - 4\left|U_{\mathrm{e2}}\right|^2 \left|U_{\mathrm{e3}}\right|^2 \sin^2 \Delta_{32}. \end{aligned} \quad (10)$$

In Equation 10, $\Delta_{ji} = \frac{\left(m_j^2 - m_i^2\right)L}{4E_{\nu}}$, where L is the distance traversed by the propagating state. Hence, the survival probability of electron neutrinos is dependent upon three differences of squared masses: $\Delta m^2_{21} = m^2_2 - m^2_1$, $\Delta m^2_{31} = m^2_3 - m^2_1$, and $\Delta m^2_{32} = m^2_3 - m^2_2$. Among these differences, only two are independent, since $\Delta m^2_{31} = \Delta m^2_{32} + \Delta m^2_{21}$. The neutrino oscillation probability is known to be of the form given in Equation 38.

$$P\left(\nu_{\mathrm{e}} \to \nu_{\mu}\right) = \sin^2(2\theta_{ij}) \sin^2\left(1.27 \frac{\Delta m^2_{ij}\left[\mathrm{eV}^2\right] L[\mathrm{km}]}{E_{\nu}[\mathrm{GeV}]}\right) \quad (11)$$

The result of the long baseline neutrino experiments have placed upper bounds on the relative scale of the neutrino masses, but can make no comment on the hierarchy of the masses beyond the absolute square of the mass difference between two states. This is shown in Figure 3.

[6] A full derivation is found in Appendix A.2.

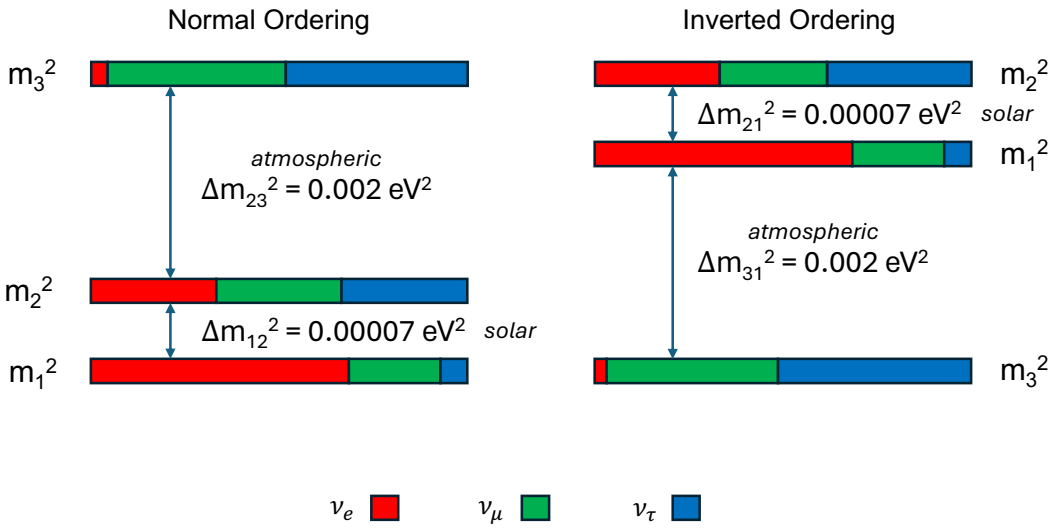


Figure 3: Neutrino mass square differences found from solar and atmospheric neutrino oscillations shown for normal ordering and inverted hierarchy (Image: adpated from [15]).

From cosmological observations of the large-scale structure of the universe [16] it can be determined that $\sum_{i=1}^{3} m_{\nu_i} \lesssim 1\text{eV}$. Which implies exceptionally small neutrino masses relative to all other standard model leptons, the smallest of which is the electron with $m_e = 0.511$ MeV, some $\mathcal{O}(10^6)$ greater than the neutrino mass upper bound.

Neutrino oscillations are direct evidence of BSM physics, and the longstanding assumption of massless neutrinos must be replaced with a mechanism to explain the origin and scale of the tiny neutrino mass.

2.2.2 Neutrino Masses

The discovery of neutrino oscillations indicates categorically that neutrino masses are non-degenerate, in contradiction with the SM assumption that the neutrinos are massless.

We must modify the SM in order to account for this in the Lagrangian of the theory. One might expect that it would be easiest to add a *Dirac mass term*, in the same way as we did for other fermions by using the Higgs doublet. Analogously to the up-type quarks, after spontaneous symmetry breaking, one finds:

$$\mathcal{L}_D = -m_D \bar{\nu}\nu = -m_D(\bar{\nu}_L \nu_R + \bar{\nu}_R \nu_L) \tag{12}$$

Where m_D in Equation 12 refers to the *Dirac mass* term. The inclusion of the Dirac Mass term introduces sterile (with a weak hypercharge of 0, they do not interact even via the weak force) right-handed neutrinos ν_R which only enter the Langrangian through the mass term. If neutrinos acquire their masses in this way, we admit the existence of right-handed

massive neutrinos, and remain bemused at the smallness of the Yukawa coupling of the neutrinos relative to all other fermions, at $\mathcal{O}(10^{-12})$, aggravating the SM flavour puzzle [17]. Instead, we might use this is a hint that there is more complexity to the mass generation. Indeed, because right-handed neutrinos and left-handed antineutrinos are singlets under SM gauge transformations, without imposition of any additional symmetry one is permitted to write the most general form of the lagrangrian by including a *Majorana* term for the RH neutrino.

Being electrically neutral, neutrinos do not immediately impose themselves on the same footing as any other fermions. Since we cannot differentiate antineutrinos from neutrinos without assuming lepton-number conservation, it is possible that, as with the photons, the neutrinos are eigenstates of the charge conjugation operator and are hence *Majorana* type particles, being that they are their own antiparticles. This would allow for lepton number violation, permitting processes such as neutrinoless double beta decay[7]. Since we have lost two degrees of freedom compared with the Dirac spinors, we have to rely on only one chiral state in order to generate the Majorana neutrino mass term, $\mathcal{L}_M$.

$$\mathcal{L}_M = -\frac{1}{2}M(\bar{\nu}_R\nu_R^c + \bar{\nu}_R^c\nu_R) \quad \Delta L = \pm 2 \tag{13}$$

In Equation 13, M is the RH neutrino mass. By generating a mass term with only two degrees of freedom, one must admit the possibility of lepton number violation by two units (denoted by $\Delta L = \pm 2$ in our Equation), since we allow for direct coupling between a particle and its corresponding antiparticle.

2.3 The Seesaw Mechanism

2.3.1 The Vanilla seesaw

The most general renormalisable[8] Lagrangian for neutrino masses includes both Dirac and Majorana terms, shown in Equation 14 and illustrated in Figure 4, where m_D is the Dirac mass, and M is the Majorana mass. The same Lagrangian is given again in Equation 15.

[7] Majorana neutrinos, being their own antiparticles, allow for the exchange of virtual neutrinos, leading to the emission of two electrons without the corresponding neutrinos. This process is significant for understanding the absolute neutrino mass scale and the nature of neutrino mass [18].

[8] A theory is renormalisable if its predictions remain finite and well-defined at all energy scales after the infinities that arise in quantum field theory calculations are systematically removed. This involves adjusting the values of a finite number of parameters in the theory. Renormalisability is crucial for a theory to make consistent and accurate physical predictions.

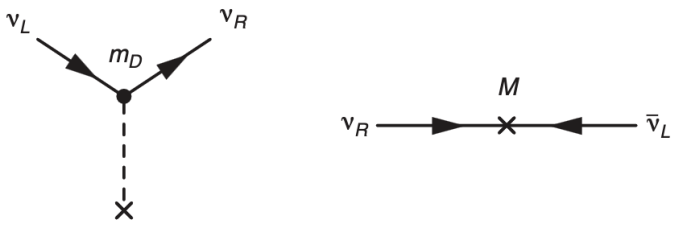


Figure 4: Feynman depiction of Dirac (left) and Majorana (right) mass terms [19]. On the left, the mixing of the left and right-handed neutrino states generates a mass term, m_D, via the Yukawa coupling to the Higgs field. On the right, a Majorana mass term, M, is introduced via the coupling of a neutrino to its own antiparticle, resulting in lepton number violation.

$$\mathcal{L}_{DM} = -\frac{1}{2}\left[m_D(\bar{\nu}_L\nu_R + \bar{\nu}_R^c\nu_L^c) + M(\bar{\nu}_R^c\nu_R)\right] + h.c. \tag{14}$$

$$\mathcal{L}_{DM} = -\frac{1}{2}\left[\overline{\nu_L}\ \overline{\nu_R^c}\right] \underbrace{\begin{bmatrix} 0 & m_D \\ m_D & M \end{bmatrix}}_{\text{mass matrix}} \begin{bmatrix} \nu_L^c \\ \nu_R \end{bmatrix} + h.c. \tag{15}$$

The Dirac term alone, seen in Equation 12, admits the possibility for mass generation via the EWSB of the Higgs mechanism. Adding the Majorana term in, through Equation 14, is automatically gauge invariant and respects the symmetries of the SM, hence it is permitted given the fact that neutrinos are electrically neutral objects. Specifically, for the candidate Lagrangian to be viable, it must obey the gauge symmetries of the SM, $SU(3)_C \times SU(2)_L \times U(1)_Y$. Neutrinos are electrically neutral making them singlets under the $U(1)_Y$ hypercharge gauge group, which means that a Majorana mass term can be introduced, allowing for same lepton sign self-coupling without violating charge conservation.

The physical states of this system are found via the eigenvalues of the mass matrix shown in Equation 15. In two regimes one finds:

$$m_\pm \approx \frac{1}{2}M \pm \frac{1}{2}\left(M + \frac{2m_D^2}{M}\right) \quad M \gg m_D \tag{16}$$

$$|m_\nu| \approx \frac{m_D^2}{M} \quad M \approx m_D \tag{17}$$

This proposed generation of the neutrino masses is known as the *seesaw mechanism*, first proposed in [20, 21, 22, 23]. It is so called rather intuitively, since it is the heaviness of the right-handed neutrino which is permitted to give the SM neutrinos their very small

masses due to the inverse relationships shown in Equations 16 and 17. The new right-handed neutrino species is coined as a *heavy neutral lepton*, often known as *sterile* since they have no charge or weak isospin, and have only their Yukawa couplings to the Higgs boson and leptons.

The seesaw model can be thought of as Type I [24] Type II [25] or Type III [26, 27]. Type I is the mechanism we have up to now been describing. Type II introduces a heavy scalar triplet which leads to small SM neutrino masses via the coupling of the lepton doublets to the Higgs field. The Type III seesaw mechanism similarly adds a heavy fermion triplet in order to achieve the same means. These mechanisms are shown at tree level in Figure 5.

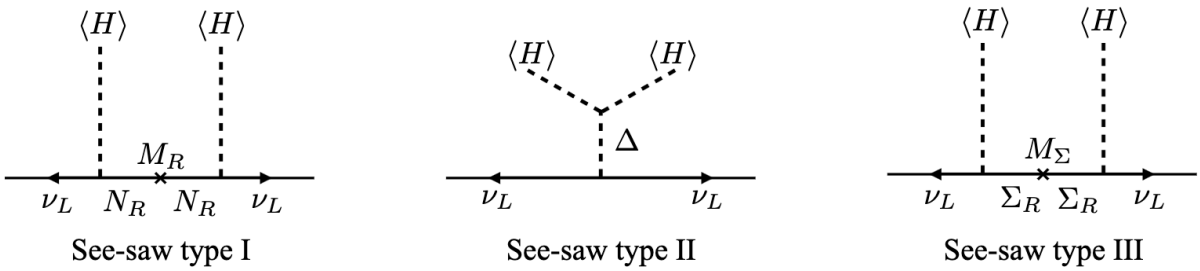


Figure 5: Mathematical schematic of Type I, II and II seesaw mechanisms [28].

2.3.2 Inverse Seesaw - Dirac HNLs

If we wish to protect lepton number conservation, by trying to find a seesaw mechanism that permits Dirac type HNLs (possibly degenerate with equal masses), we must do so through an inverse seesaw mechanism.

In this instance, we consider at least two HNLs, and we can call them N and S. Setting once again both Majorana and Dirac masses we can write the mass matrix, M, in the basis of $(\nu_L\ S_L\ N_L^c)$.

$$M = \begin{bmatrix} 0 & m_D & 0 \\ m_D & \mu_N & m_R \\ 0 & m_R & \mu_S \end{bmatrix} \tag{18}$$

In Equation 18, m_{D} corresponds to the Dirac mass terms involving v and N, representing the direct coupling between the neutrino field N and the Higgs field v. On the other hand, m_{R} corresponds to Dirac mass terms of a different form, $m_{\mathrm{R}}((N^c)^\dagger S^c + (S^c)^\dagger N^c)$, which involve a scalar singlet field S and the conjugates of the right-handed neutrino field N^c. The Majorana mass terms for N and S are μ_N and μ_S respectively.

We assume a mass hierarchy where μ_N, $\mu_S \ll m_D \ll m_R$. Furthermore, the contribution

from μ_N to the masses of ν is minimal and is hence set to 0. With these assumptions, the mass matrix can be diagonalised as:

$$\begin{pmatrix} m_\nu & 0 & 0 \\ 0 & m_N & 0 \\ 0 & 0 & m_S \end{pmatrix} \tag{19}$$

Where $m_\nu \approx \frac{\mu_S}{m_D^2} \frac{m_D^2}{m_D^2+m_R^2}$, and m_N, $m_S \approx \sqrt{\frac{m_D^2+m_R^2}{2}} \mp \frac{\mu_S}{2} \frac{m_R^2}{m_D^2+m_R^2}$. Unlike in the previous regime, here it is the smallness of the surviving Majorana mass term, μ_S, which gives the SM neutrinos their low mass. The mixing angle between the active neutrinos and the heavy sterile states is given by: $|U| \approx \frac{m_D}{m_R}$. This mixing angle quantifies the contribution of the heavy sterile component to the light neutrino state.

In searches, these particles are often called Dirac type HNLs, and through this thesis we will refer to them as such, where we are working with a single heavy Dirac HNL in the limit $\mu_S \to 0$. In reality, we might more accurately refer to them as *pseudo-Dirac* HNLs. This is because, if we set the Majorana mass equal to zero, admitting that m_N and m_S are degenerate, we also have set the mass of the SM neutrino equal to zero, which is the same problem we were initially trying to negate. Fortunately, this distinction does not significantly impact our experimental simulations.

Indeed, choosing that the SM neutrino has a Yukawa coupling of $\mathcal{O}(1)$ is somewhat arbitrary, since we recall that the electron has a Yukawa coupling of $\mathcal{O}(10^{-6})$. If we were to search for HNLs in the regime with Yukawa coupling of $\mathcal{O}(1)$, then we predict the mass of the HNL to be in the $\mathcal{O}(10^{14} - 10^{15})$ GeV. If, however, we take it to be of the same order as the electron, we can reasonably search for Dirac HNLs of $\mathcal{O}(10 - 100)$ GeV. The latter of these cases would be referred to as the *low scale inverse-seesaw mechanism* and produces precisely the type of object we will search for in this study.

2.4 Solutions to Problems Beyond the Standard Model

Physics beyond the SM is not just alluded to with the existence of non-zero neutrino masses. Direct cosmological observations give evidence for the existence of so-called *dark matter* and the prevalence of matter over antimatter in the universe, known as the baryon asymmetry of the universe (BAU) demands a mechanism through which it can be understood. Interestingly, our extension of right-handed neutrinos may provide solutions to both of these outstanding problems within the SM.

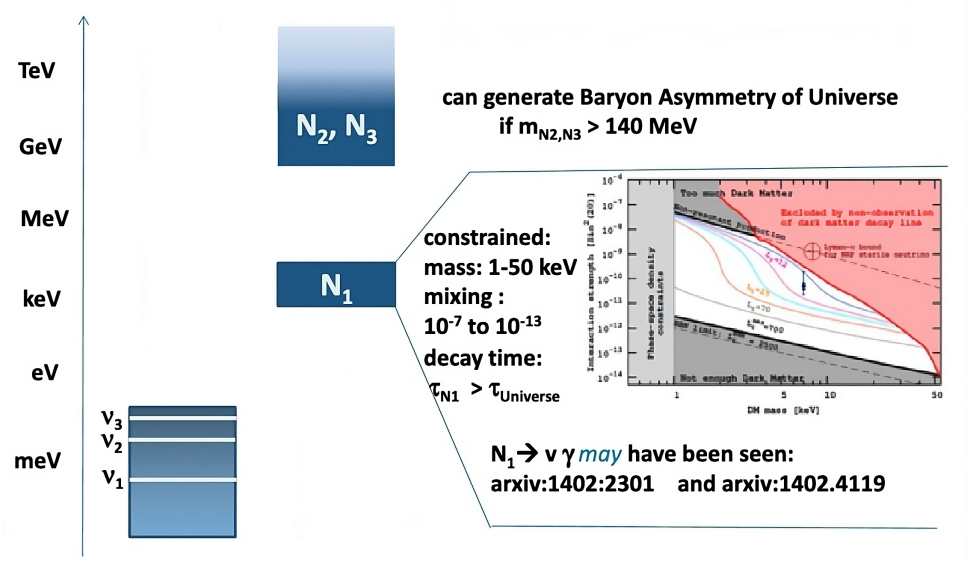


Figure 6: Possible scenarios for HNL explainations of dark matter and the BAU (Image: taken from [29]).

By introducing three generations of RH heavy neutrinos, $N_{1,2,3}$, the two heavier particles, $N_{2,3}$, generate the masses of active neutrinos via the Type I Seesaw mechanism discussed in Section 2.3. The same two right-handed neutrinos are also responsible for generating the BAU provided that their masses are close to each other, via *low-scale leptogenesis*. In leptogenesis, the decay of these heavy neutrinos in the early universe creates an asymmetry between leptons and anti-leptons due to CP-violating processes. Specifically, if the masses of N_2 and N_3 are sufficiently close to each other, this can enhance the CP-violation effects and the resulting lepton asymmetry due to the resonant enhancement, a phenomenon known as *resonant leptogenesis*.

The lightest sterile neutrino N_1 is the DM candidate. A viable DM candidate must be electrically neutral, cosmologically stable, and interact weakly with ordinary matter. The requirement to be a viable DM candidate forces the Yukawa couplings of N_1 to be very small, leaving the lightest active neutrino almost massless [28]. It must be minimal to avoid current direct detection while still ensuring N_1 was produced in the right abundance in the early universe, a balance achieved via mechanisms like "freeze-in" or "freeze-out". The lifetime of N_1 must also exceed the age of the universe, which is a convenient byproduct of the tiny Yukawa coupling [30].

The mass of N_1 is constrained to be in a range that makes it cold or warm DM, typically on the keV scale, to match the observed structure formation in the universe. Cold DM refers to particles that were non-relativistic at the time of galaxy formation, leading to the large-scale structure we observe today, while warm DM can slightly erase small-scale structures, providing a potential solution to discrepancies like the missing satellite

problem [31]. The signature for the DM candidate in this picture is the decay $N_1 \to \nu\gamma$, which may have already been seen [32, 33].

2.5 Present Searches for HNLs

As a foundation for the research in this thesis, we will explore present and future searches for HNL models, by looking at the existing and projected constraints on the possible HNL parameter space. HNLs are being searched for actively in experiments such as ATLAS at CERN, and will be pushed further by future experiments such as SHiP and the FCC. Understanding these limits in the context of the results set out in Section 6 will be critical for underscoring the need for the FCC machine. A summary of the limits in the HNL parameter space with respect to the electron mixing angle is given in Figure 7.

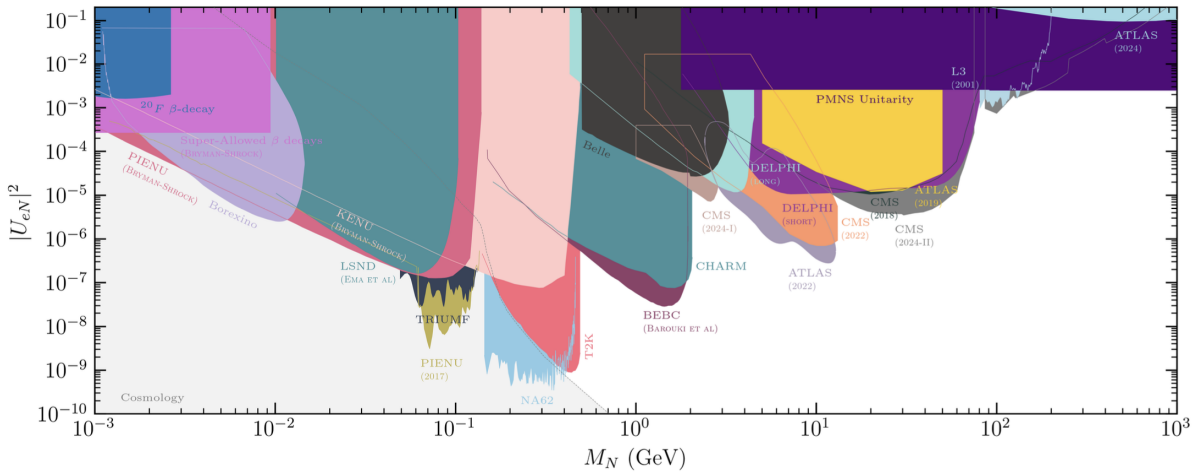


Figure 7: Existing and projected constraints on heavy neutrino mixing at 90% C.L., as a function of the HNL mass, M_N and electron neutrino mixing angle $|U_{eN}|^2$. The bounds set by the different experiments are displayed separately, but we only show those more relevant in each mass window. Single flavour dominance is assumed, with the limits on $|U_{eN}|^2$ shown [34].

The mass range considered in this thesis will be between 10-80 GeV, with mixing angles squared between 10^{-4} and 10^{-10}. To this end, we will focus on the best existing constraint in this range published by CMS in April of 2024, shown in Figure 8.

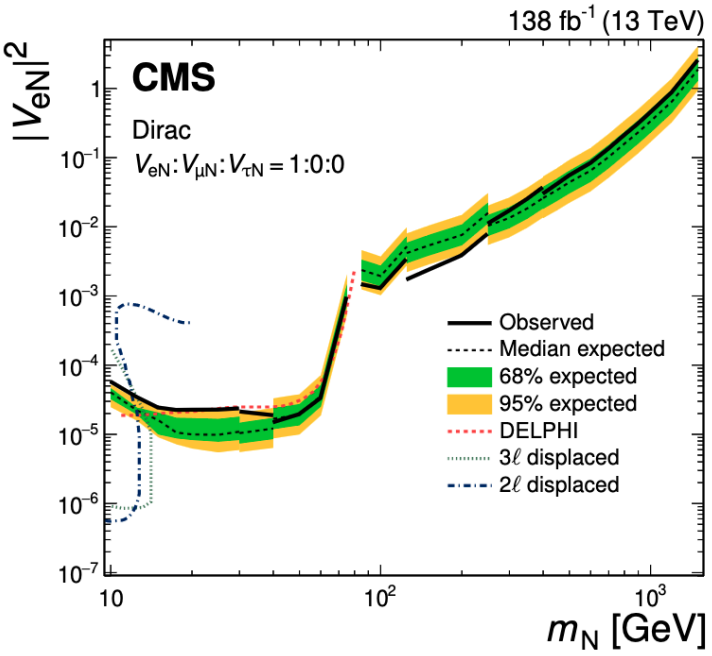


Figure 8: The 95% CL limits on $|V_{eN}|^2$ for Dirac HNLs. The area above the solid (dashed) black curve indicates the observed (expected) exclusion region. Previous results from the DELPHI Collaboration [35] are shown for reference. Previous CMS results "3ℓ displaced" [36] and "2ℓ displaced" [37] are shown to highlight the complementarity to other search strategies [38].

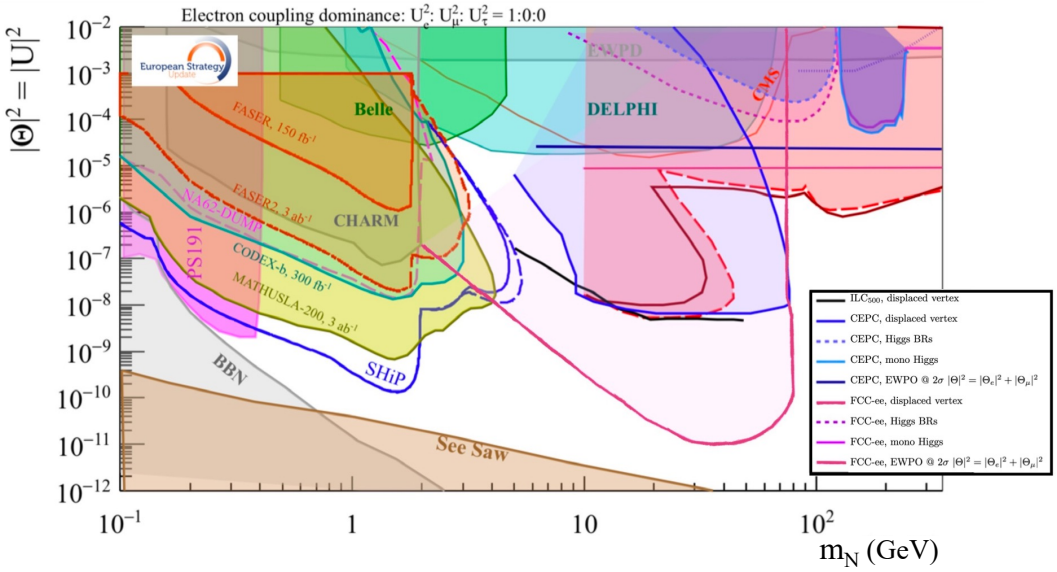


Figure 9: 90% CL exclusion limits for a HNL mixed with the electron neutrino, as presented in the European Strategy for Particle Physics Briefing Book [39]. The FCC-ee curves are in (overlined) dark purple—for FCC-ee, this is equivalent to a plot as a function of the sum of matrix elements squared $|U_N|^2$. The curve below the Z boson mass corresponds to the combined LLP and prompt analysis performed with 10^{12} Z bosons in [40]. The horizontal limit at high masses results from the effect of light-heavy neutrino mixing on the EW precision observables and remains valid up to $\mathcal{O}(1000)TeV$ [41].

On the left-side of the mass range considered in Figure 8, we see the existing limit within our mass range, with a reach on the mixing angle of approximately $\mathcal{O}(10^{-5})$. By contrast, the projected limit for HNLs at the FCC-ee is shown overlayed in Figure 9. This limit is obtained via simulation of the physics as well as detector reconstruction using a fast-simulation of one proposed detector design, IDEA.

Figure 9 shows a projected reach across all HNL final states of approximately $\mathcal{O}(10^{-11})$ in $|U_{eN}|^2$ in the mass range considered in this analysis. This is a consequence of the advantages of lepton colliders and their very "clean" detector environments compared to proton-proton collisions as is the case with the CMS result of Figure 8. In the next chapter, we introduce the FCC and explore further the topology of our considered HNL model in the detector.

3 The Future Circular Collider

The Large Hadron Collider (LHC) at CERN is the largest and most powerful particle accelerator ever constructed. It is located 100 meters underground at the Swiss-French border, with a circumference of 27 kilometers. The LHC achieves center-of-mass (CM) energies of up to 13.8 TeV in proton-proton collisions. Since its first collisions in 2008, the LHC has been a cornerstone of modern particle physics, leading to the discovery of the Higgs boson and providing record precision measurements of Standard Model parameters such as the masses of the W and Z bosons, as well as testing principles like lepton universality and many others.

As we look forward, the planned high luminosity upgrade of the LHC (HL-LHC) is projected to begin data taking at the earliest in 2029 [4]. Increasing instantaneous luminosities by a factor of 5 to 7.5 relative to the LHC nominal value, we will gain enormous statistics with which to battle the current tensions of the SM, and test our theories with hitherto unseen precision.

The Future Circular Collider (FCC) initiative is formulating plans for the future of CERN beyond the high luminosity upgrade. The FCC will enhance current capabilities in domains of both energy and intensity, targeting collision energies up to 100 TeV with the final stage of FCC-hh. It is expected that the FCC will serve both as a precision machine, and with higher CM energies available, also serve as a machine for outright NP discovery. Just as LEP set the stage for the LHC, the FCC-ee will be the first stage of operational collisions in the new collider, and will lay the groundwork for future electron-hadron (eh) collisions, and further still until hadron-hadron (hh) collisions. The projected schema of the LHC and FCC are overlaid in Figure 10.

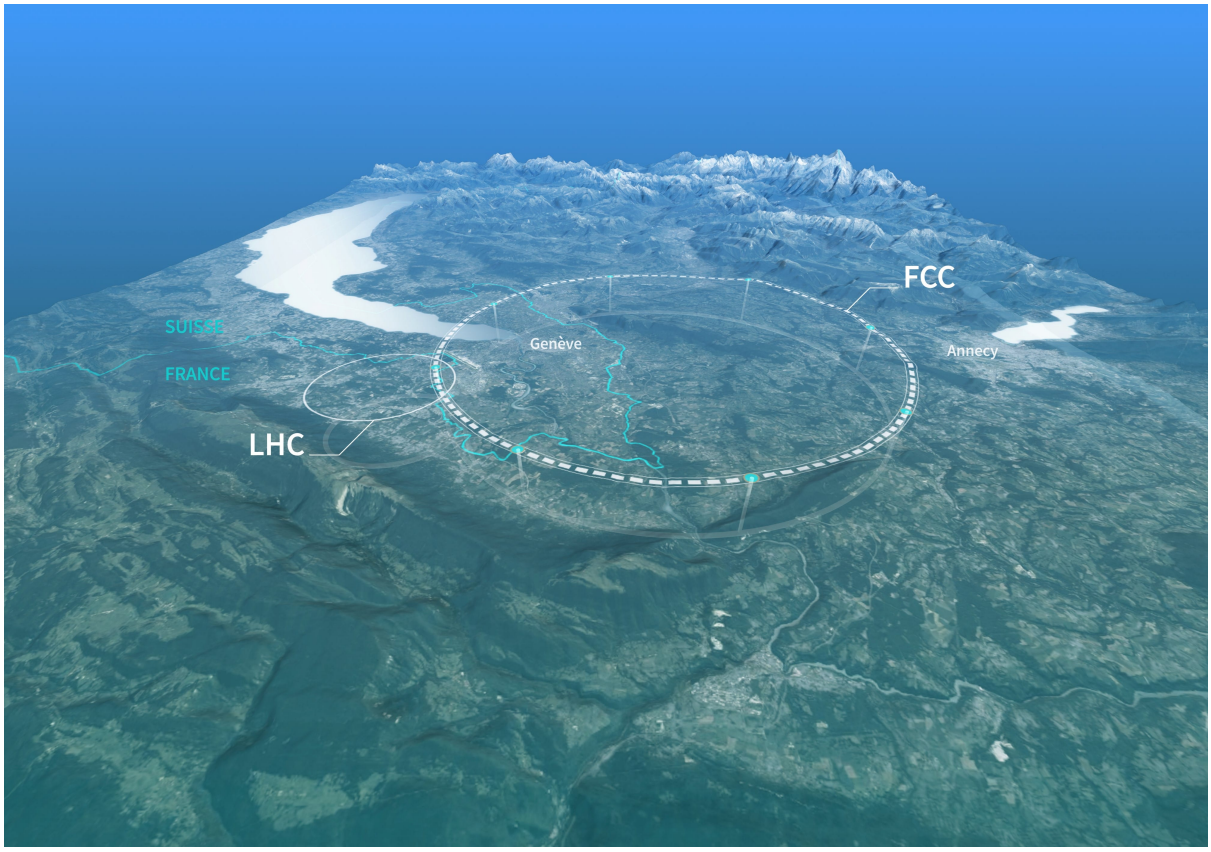


Figure 10: Projected schematic for possible placement and size of the FCC relative to the LHC (Image: CERN).

Figure 10 illustrates the magnitude of the FCC, highlighting the enormous feasibility challenge due to features of the terrain as well as the size of the machine. One natural question may be to wonder *why* a collider with a much bigger radius will allow us to reach such higher energies in the final hh collision stage. This is primarily due to synchrotron radiation losses, which are proportional to the fourth power of the particle energy and inversely proportional to the radius of the collider. In the FCC, the curvature of the particle trajectory is "gentler", resulting in significantly reduced synchrotron radiation losses. Consequently, particles can be accelerated to higher energies before the energy loss due to radiation becomes a limiting factor. This is one of the key reasons why the FCC, with its much larger radius compared to the LHC, can achieve higher collision energies for the same colliding particles.

In the following section, we will discuss the unfolding stages of the FCC, with a primary focus on the FCC-ee before discussing the proposed detector designs and the phenomenology of the HNL decays.

3.1 FCC-ee

The projected timeline of the FCC operations and their alignment with CERN operations is shown in Figure 11. The FCC is targeted to begin operations in circa 2048, and will take data for an estimated 15 years.

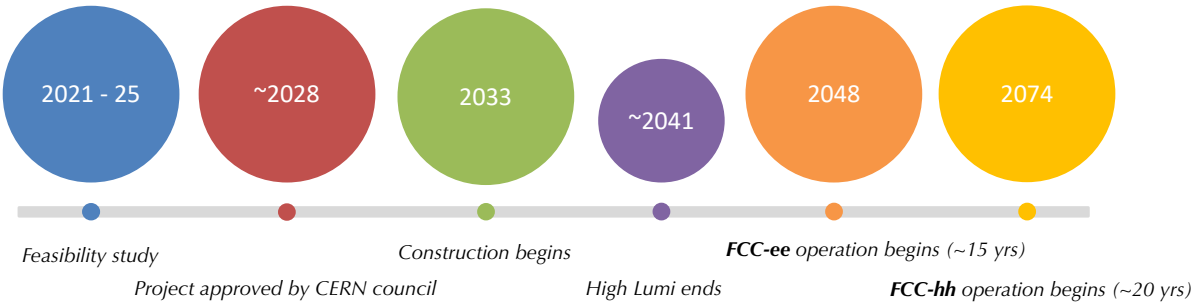


Figure 11: Projected timeline of the FCC project (Image: adapted from [42]).

The FCC-ee is set to unfold in phases, functioning sequentially as an electroweak, flavour, Higgs, and top factory. It will cover an energy spectrum starting from the Z pole and beyond to the WW threshold, extending through the peak Higgs production rate, and reaching beyond the $t\bar{t}$ threshold. The high luminosity available at the Z pole paves the way for exciting opportunities in the search for new particles. The stages of the FCC-ee runs, complete with their CM energies, luminosities and statistics are compiled in Table 1, and depicted further in Figure 12 [43].

Process	$\sqrt{s}$ (GeV)	$\mathcal{L}$ ($10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)	$\int \mathcal{L}\,\mathrm{dt}$ (ab^{-1})	Statistics
Z	91.2	230	150	3×10^{12} Z decays (visible)
WW	160	28	12	10^8 $\mathrm{W^+W^-}$ events
ZH	240	8.5	5	10^6 ZH events
$t\bar{t}$	350, 365	1.8, 1.55	0.2, 1.5	10^6 $t\bar{t}$ events

Table 1: Event statistics at various runs of the FCC-ee [44].

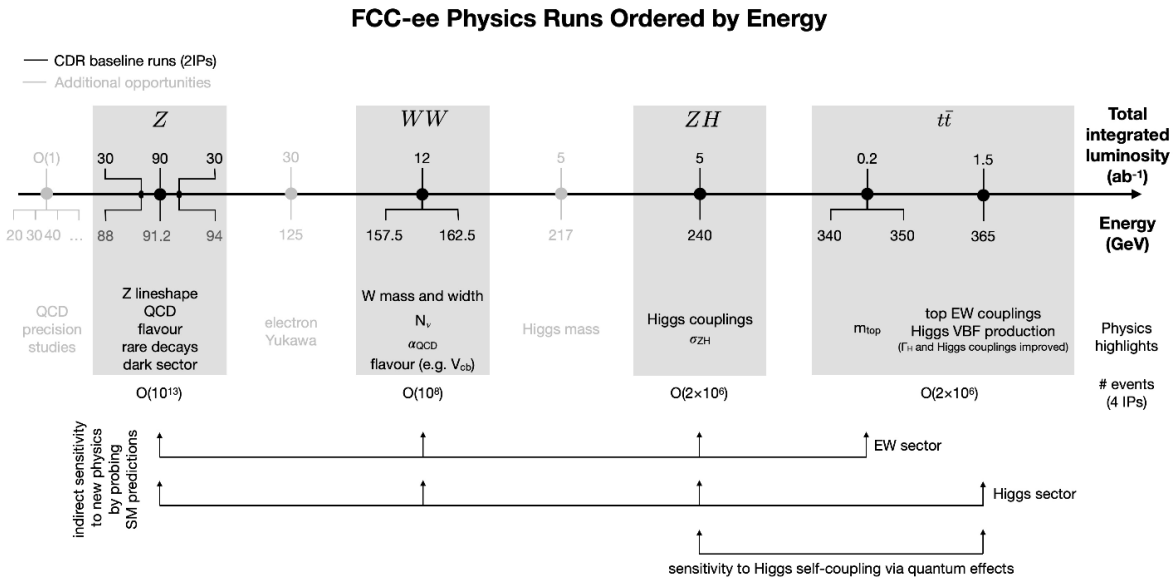


Figure 12: Projected FCC-ee runs ordered by energies (Image: FCC Collaboration).

The interaction environment in lepton-lepton collisions such as those at the FCC-ee are very 'clean'[9] compared to those of hadron colliders, even relative to linear e^+e^- colliders. The FCC-ee provides an excellent opportunity to probe new particles with masses between 1 and 100 GeV and their electroweak couplings, such as the benchmark heavy neutral lepton models we will discuss in Chapter 3 [45].

3.1.1 The IDEA Detector

There are two key detector designs described in the FCC-ee conceptual design report [46]: the CLIC-Like Detector (CLD) [47] and the Innovative Detector for Electron-positron Accelerators (IDEA) [48]. A third proposal is made for the Lepton coLlider Experiment with Granular calorimetry Read-Out (ALLEGRO), featuring a highly granular noble-liquid electromagnetic calorimeter [49].

For the purposes of this study, we will focus on the IDEA detector which can be seen below in Figure 13.

[9] In this context, 'clean' refers to the lower background noise in the collision environment, making it easier to analyse the results. Lepton-lepton collisions produce fewer secondary particles and have more straightforward initial states than hadron collisions, which involve complex interactions between quarks and gluons.

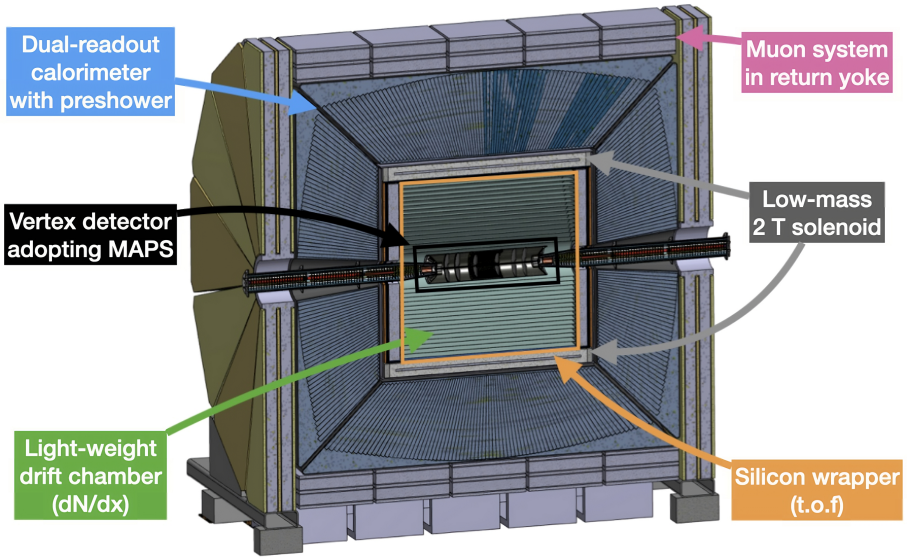


Figure 13: Schematic design of the proposed IDEA detector concept [50].

The detector requirements at the FCC-ee are stringent. With a heavyweight physics programme targeting immense luminosities, we must have detectors capable of matching systematics such that precision measurements and the discovery of weakly interacting particles are able to be correctly handled in the detector sub-systems.

In the IDEA detector, the silicon pixel vertex detector (VTX) sits at the centre, adjacent to the particle collision beam, for reconstruction of charged particle trajectories. Positioned near the beam, the VTX is essential for pinpointing the decay vertices of promptly decaying particles. Progressing outward, a short-drift wire chamber (DCH) sets out a large volume to ensure tracking of all charged particles. This arrangement facilitates the grouping of particles originating from a common decay, thereby aiding in the reconstruction of entire decay processes. The DCH employs an electric field generated by a superconducting solenoid coil to curve the paths of charged particles, enabling momentum measurements. Encasing the DCH is a layer of silicon micro-strip detectors, which contribute to refined track reconstruction.

For detailed energy measurements and particle identification, the IDEA detector includes a pre-shower detector preceding a dual-readout calorimeter. The pre-shower detector classifies particles based on the shower patterns observed in the tracking detectors, while the calorimeter stops almost all particles with dense material to measure energy deposition. The hadronic calorimeter is tasked with detecting hadrons that interact strongly, such as protons and neutrons, while the electromagnetic calorimeter is designated for electrons and photons. Decay products of the heavy tau lepton are recorded within the detector. Unlike tau leptons, muons are not halted by the electromagnetic calorimeter due to their comparatively long lifetime, but are instead measured using a magnetic yoke positioned

as the outermost sub-detector. Due to the clean environment of lepton-lepton collisions, the total collision momentum is completely known, unlike in hadron-hadron collisions where the linear momentum is unknown as a result of the underlying quark-quark interaction. Hence, neutrinos are inferred through the total missing energy in the event reconstruction, rather than the missing transverse momentum as at the LHC.

3.2 HNLs at the FCC-ee: Experimental Outlook

The Phenomenology of HNL production is discussed extensively in [29, 51, 52, 53, 54]. The search for HNLs in the Z resonance run of the FCC-ee demands the production of an SM neutrino with a heavy right handed counter part, as seen through the Feynman diagrams in Figures 14 and 15. In all cases considered, the HNL decays via an off-shell[10] W boson.

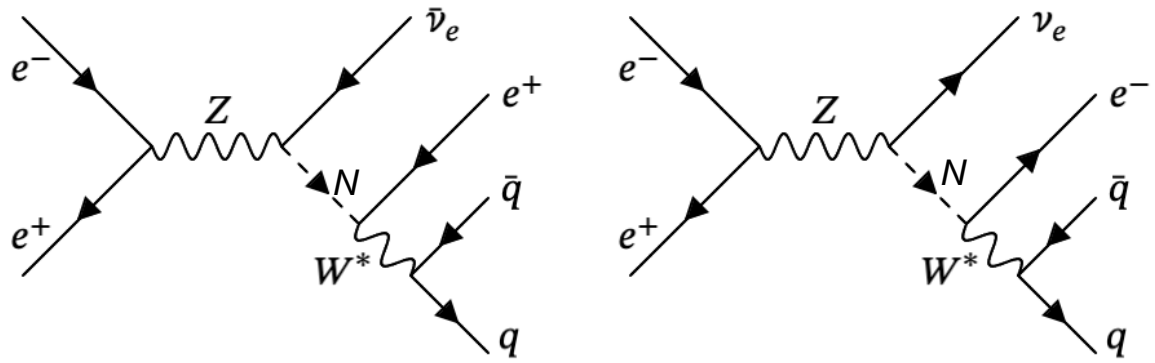


Figure 14: Feynman diagrams for the $e^+e^- \to Z \to N\bar{\nu}_e \to e^+q\bar{q}$ and $e^+e^- \to Z \to N\nu_e \to e^-q\bar{q}$. These are the lepton number violating processes corresponding only to a Majorana-type HNL, which has the ability violate letpon number conservation.

[10] An "off-shell" particle does not satisfy the usual energy-momentum relation for a real (on-shell) particle. In other words, it is a virtual particle that exists temporarily during the interaction and does not have a well-defined mass. This allows for different decay channels and intermediate states in particle interactions.

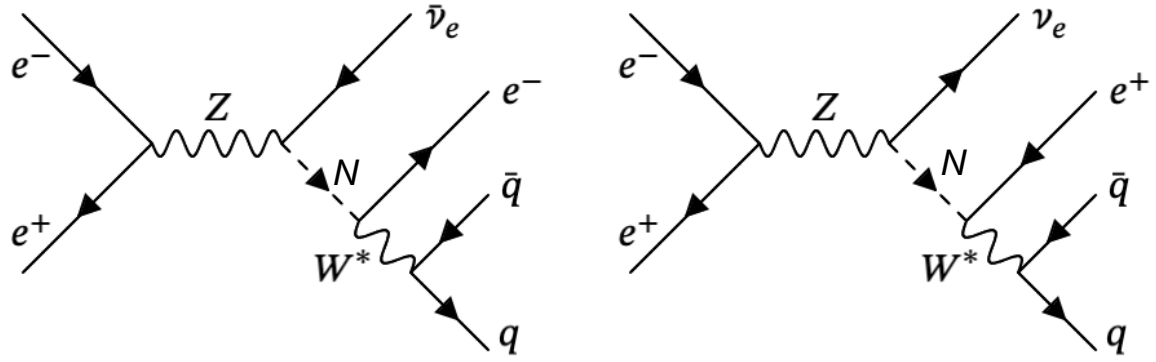


Figure 15: Feynman diagrams for the $e^+e^- \to Z \to N\nu_e \to e^+q\bar{q}$ and $e^+e^- \to Z \to N\bar{\nu_e} \to e^-q\bar{q}$. These are the lepton number conserving processes corresponding to a Dirac-type HNL, but are also allowed in the case of Majorana-type.

Figure 15 shows the electron dijet final state of the HNL decay via a virtual W boson following from its lepton number conserving production. The quarks produced in the fundemental interaction, are observed as jets, which are collimated sprays of hadrons resulting from the hadronisation of quarks. Figure 14 shows, on the other hand, a final state resulting from a lepton number violating HNL, one consistent with a Majorana type particle, in which neutrinoless double beta decay would also be possible.

HNLs produced via the Z resonance from e^+e^- collisions have been studied since the days of LEP with the DELPHI experiment [55]. The branching ratio for a the production of a neutrino anti-neutrino pair via a Z boson is known to be $BR\,(\mathrm{Z} \to \nu_l\bar{\nu}_l) \approx 0.063$ for any lepton flavour of $l = e, \mu, \tau$. The branching ratio for a Z boson decaying to a anti-neutrino HNL pair is given by Equation 20, where U_{lN} is the mixing between the standard model neutrino and the HNL. This branching ratio is equal to the branching ratio of $\mathrm{Z} \to \bar{N}\nu$.

$$BR\,(\mathrm{Z} \to N\bar{\nu}) = BR\,(\mathrm{Z} \to \nu\bar{\nu})\,|U_{lN}|^2 \left(1 - \frac{m_N^2}{{m_\mathrm{Z}}^2}\right)^2 \left(1 + \frac{1}{2}\frac{m_N^2}{{m_\mathrm{Z}}^2}\right) \tag{20}$$

Within the framework of the minimal Type I seesaw model, we consider the decay of a HNL via a virtual Z or W boson, shown in Equation 21. In our analysis, we focus on the electron dijet system in the final state, and hence are interested in the W^* decays of the HNL.

$$\begin{aligned} N \to \nu\mathrm{Z}^* &\longrightarrow \nu\bar{\nu}, \ell\bar{\ell}, \mathrm{q\bar{q}} \\ N \to \ell'\mathrm{W}^* &\longrightarrow \nu\bar{\ell}, \mathrm{q\bar{q}}' \end{aligned} \tag{21}$$

There are four distinct decay topologies ($\ell^+\ell^-\nu$, $\ell^\pm\nu q\bar{q}'$, $\ell\nu q\bar{q}$, and $\nu q\bar{q}$ with $l = e, \mu, \tau$,

$q = u, d, s, c, b$ and $q\bar{q} = u\bar{d}, c\bar{s}$) whose branching ratios are largely insensitive to the mass of the HNL. For masses below 50 GeV, the contribution from charged currents is about 75%, which increases to 80% for masses near 75 GeV. The branching ratios remain relatively stable across this mass range. The branching ratios are considered at length in [56], and Figure 16 summarises the behaviour of the branching ratio over the considered mass range for relevant final states.

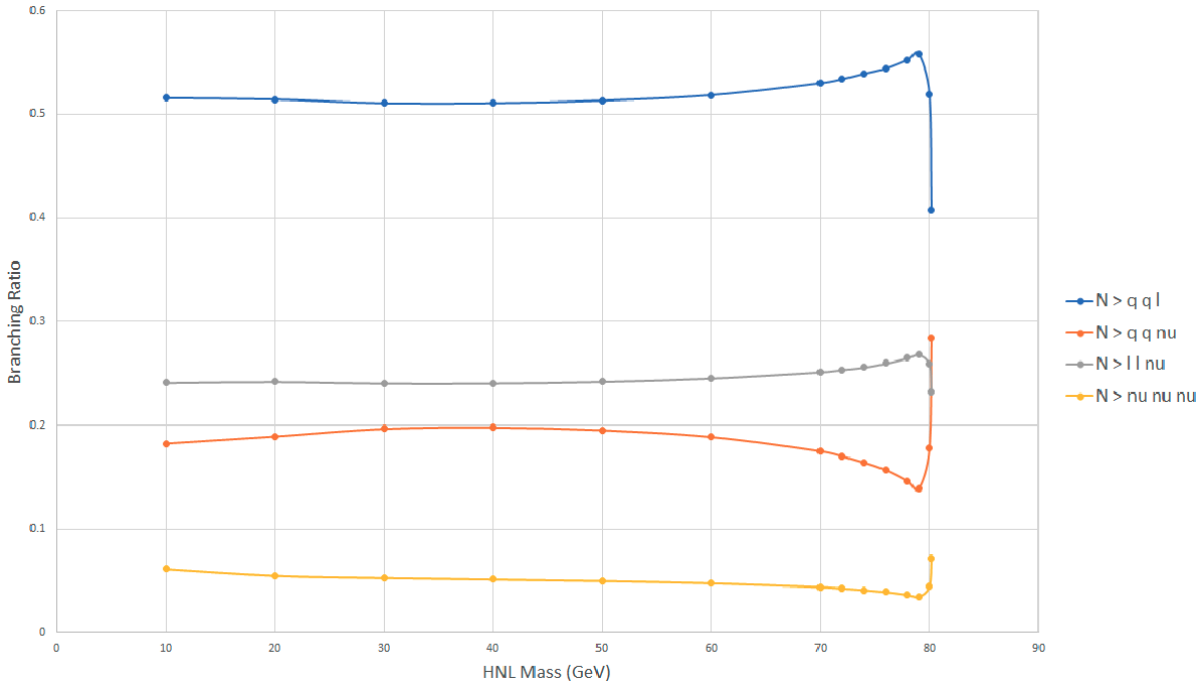


Figure 16: HNL branching ratios as a function of the HNL mass for fixed $|U|^2 = 10^{-3}$. It is helpful to note that the branching raito is independent of the mixing angle (Image: T. Sharma).

An interesting property of HNLs are the lifetimes and decay lengths. The decay length is proportional to the mixing angle, U_{lN}^{-2} and HNL mass, M^{-5}, meaning that for HNLs with $M < 30$ GeV with a mixing angle $U_{lN}^2 < 10^{-6}$ we have long-lived particles (LLPs) against a prompt background, creating a distinct and easily accessible signature for the HNL decay. The decay length for a given HNL is shown in Equation 22, and its lifetime in Equation 23.

$$L \approx 2.5 \times \left(\frac{10^{-6}}{U_{lN}}\right)^2 \times \left(\frac{100\,\text{GeV}}{m_N}\right)^5 [\text{cm}] \tag{22}$$

$$\tau_N = 5 \times 10^{26} \times \left(\frac{m_N}{1\text{keV}}\right)^{-5} \left(\frac{U_{lN}^2}{10^{-8}}\right)^{-1} [\text{s}] \tag{23}$$

Thanks to the consequential lifetime of some of the HNL models considered, the displaced

topology of the HNL decay from the interaction point displays a clear signature in the fiducial volume of a lepton collider such as the one proposed at the FCC-ee. The final state of the decays demonstrated in Figures 14 and 15 are shown topologically through the schematic of the FCC-ee event display shown in Figure 17. For some of the signal points in range, specifically those with small mass and small couplings, the secondary vertex can be some order of 1 m away from the interaction point, which would be a clear indication of a signal event. With a displaced vertex, substantial mass with no pileup in an e^+e^- machine, one would perhaps rightly expect a clear signature with promising discovery potential for the HNL model proposed.

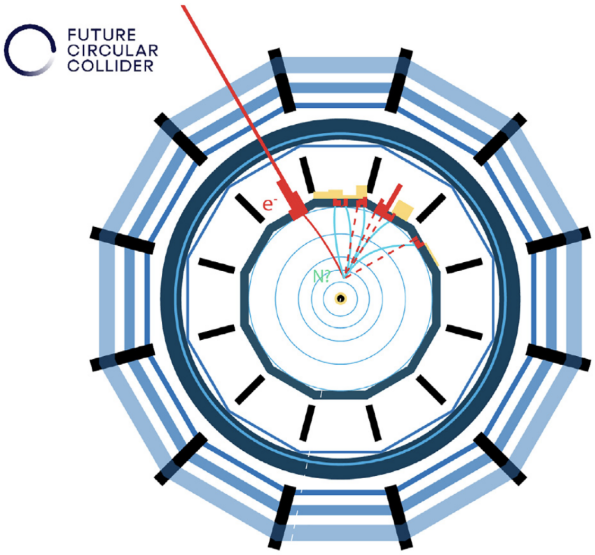


Figure 17: Representation of an event display at an FCC-ee detector of a HNL decay into an electron and a virtual W decaying hadronically (Image: FCC Collaboration).

The main background one might consider from such a machine when looking for our HNL signals is given by the limitation of the irreducible four-body decay of the Z boson, $Z \to W^*W \to l\nu qq$ which mimics the same final state as those considered in the case of the HNL. The Feynman diagram for one example of such a process is given below in Figure 18.

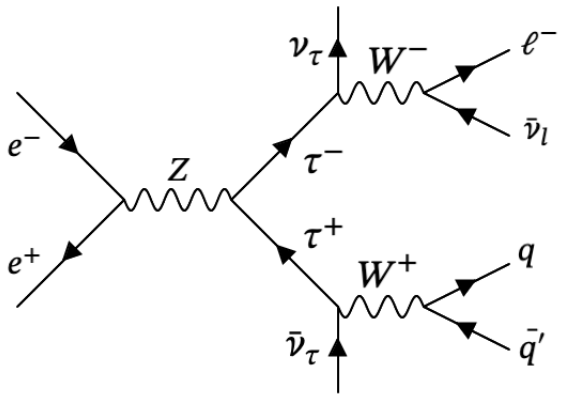


Figure 18: One example of the Feynman diagram resulting in a four body final state for $e^+e^- \to Z \to \tau\bar{\tau} \to l\nu q\bar{q}'$. This process represents an irreducible background for our signal topology.

Further considered background would come from the standard decay of the Z boson into pairs of charm quarks or bottom quarks, $Z \to c\bar{c}$ and $Z \to b\bar{b}$. These signatures would give considerably different topology to that of our signal, but we nevertheless consider them for the purposes of our analyses. Example Feynman diagrams for the $Z \to b\bar{b}$ and $Z \to c\bar{c}$ decays are given below in Figure 19.

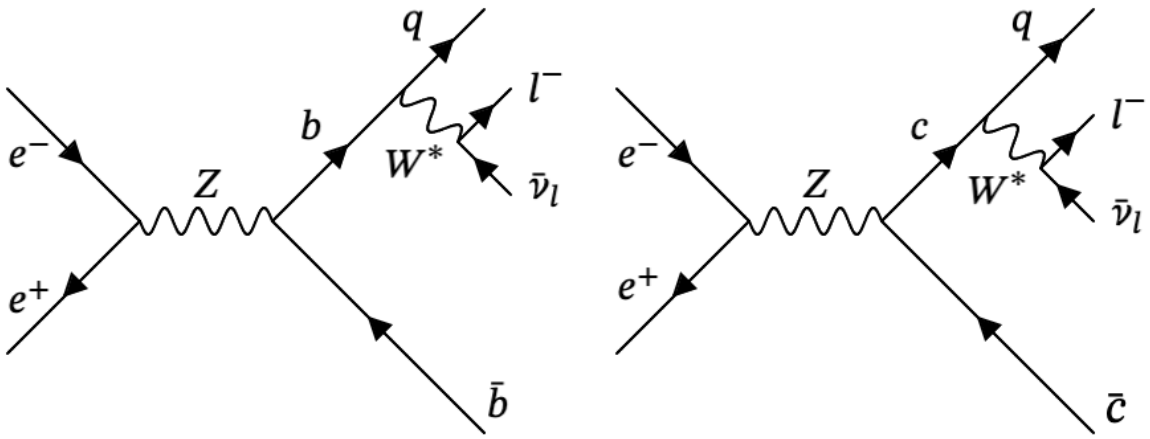


Figure 19: Possible Feynman diagrams for the $e^+e^- \to Z \to b\bar{b}$ (left) and $e^+e^- \to Z \to c\bar{c}$ (right) background processes.

In the next chapter, we discuss details of the FCC Software and the simulation details relevant for the analysis which will be discussed in Chapter 5.

4 Simulation

For the purposes of studying our benchmark HNLs, we worked within the FCCAnalyses framework. In this chapter, a summary of the methodology is composed.

4.1 Software setup

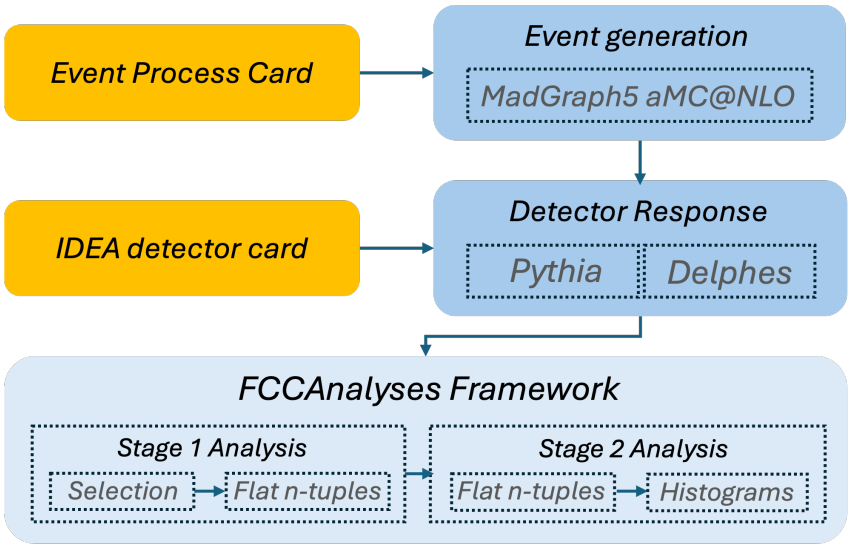


Figure 20: Flowchart of the `FCCAnalyses` framework complete with the event generation and detector simulation.

The Monte Carlo (MC) events are generated by parsing the event process card to `MadGraph`. For this analysis, the latest feature release of `MadGraph5_aMC@NLO` [57] at the time of access, `v3.5.3`, was used. For the HNL signals, the `HeavyN` model [58, 59] in the `Universal FeynRules Object` [60, 61, 62] is used to model the Dirac HNLs used for this study.

The IDEA detector card can be found in the repository linked in Reference [63]. The development of a comprehensive full detector simulation is a complex and resource-intensive task. Prior to its completion, a `Delphes` parametric fast simulation approach is adopted to approximate the interactions of particles with the detector. This method significantly streamlines the simulation process by applying predefined efficiency factors and smearing algorithms to the true momentum, energy, and time measurements of particles. These approximations are designed to mimic the expected performance of the detector, including its tracking efficiency, energy resolution in calorimeters, and timing resolution, without the need to simulate every interaction of particles with detector materials. The simulation of jet reconstruction, flavor tagging, and missing energy calculations are similarly treated, relying on smeared particle properties rather than detailed physical processes. While this approach does not capture the full complexity of particle interactions within

the detector, it provides a valuable tool for early physics analysis and strategy development. Significant progress is being made for a full detector simulation for the IDEA detector concept by the FCC collaboration [64].

The analysis framework used throughout this thesis is routed in the `Key4HEP` software, `EDM4HEP`, which is illustrated below in Figure 21.

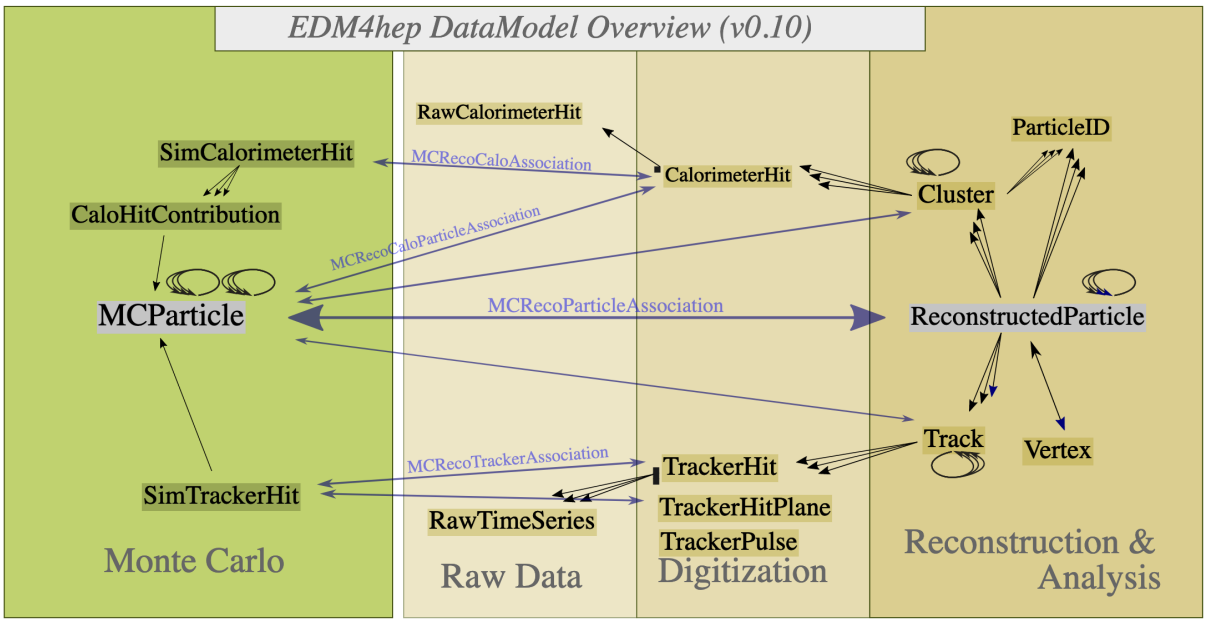


Figure 21: Graphic to illustrate the flow of objects within the key4HEP framework from Monte Carlo Level to detector reconstruction, including raw rata processing and digitization (Image: CERN [65]).

The `Delphes` IDEA card simulates the detector response and applies efficiency corrections to the raw MC events, effectively replicating the detector's performance. For example, if the final state decay products are highly collimated along the beamline, they do not interact with the detector, and hence remain undetected. Further, if a particle, for example a charged lepton, does not have sufficient enough energy (in the case of `Delphes` IDEA card, this is $E < 2$ GeV) then it will not be reconstructed. After pre-selection, the distributions are smeared to reflect inefficiencies in the detector subsystems.

The `FCCAnalyses Framework` [66] is the centralised analysis software used to perform data analysis for FCC studies, using `RDataFrame` [67, 68] in `ROOT` [69] to create flat n-tuples from reconstructed particle objects produced via `Delphes`. In stage 1 of the analysis, pre-selection is applied and various cuts are be made on the distribution. In the final stage of analysis, histograms are produced allowing for visual analysis of the distributions for variables such as electron P_T, which forms the basis for our sensitivity studies.

4.2 Jet algorithms

In this study, jet reconstruction was primarily conducted using the `FastJet` software, rather than the initial event generation phase with `Pythia`. This approach was chosen for the enhanced control and adaptability it provides when working directly with particle data from the `EDMHEP` files. The `Durham jet algorithm` [70] was used for the clustering jets.

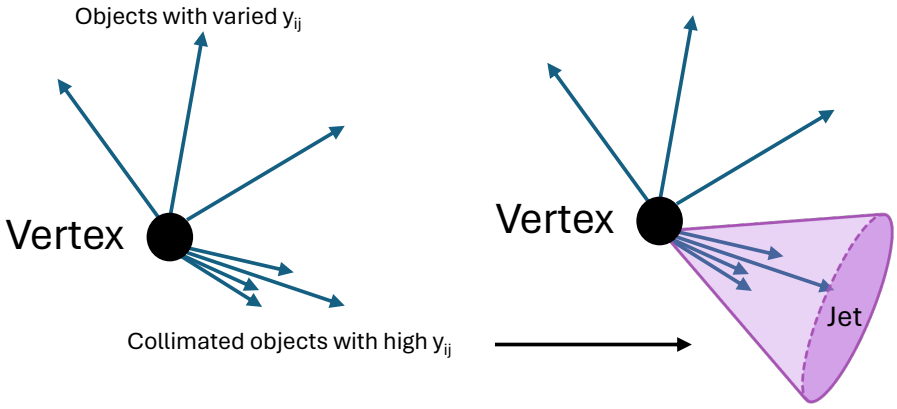


Figure 22: Illustration of jet clustering for highly collimated objects originating from a common vertex, where y_{ij} is the quantified measure of 'closeness' used by the clustering algorithm to find jets.

The essence of a clustering algorithm is to gather spatially proximate particles into a single object, known as a jet. Jets arise in the detector due to hadronisation processes from quarks produced in interactions immediately forming hadrons due to confinement. The criteria for proximity and the quantification of inter-particle distance, however, can vary across different algorithms. For the Durham algorithm used in this thesis, 'closeness' is quantified by comparing the squared invariant mass of a particle pair to the total squared invariant mass of the entire event. In Figure 22, closeness is defined with the metric $y_{ij} = 2\min\left(E_i^2, E_j^2\right)(1-\cos\theta_{ij})/E_{\text{vis}}^2$, where E_i and E_j are the energies of particles i and j, θ_{ij} is the angle between the momentum three vectors of the two particles, and E_{vis} is the total momentum of the constituents in the event. This metric is calculated for all particle pairs, and those with the minimum distance are combined to form a jet. The process continues iteratively, with distance recalculations and further merging, until a predefined condition is satisfied. For the particular analysis detailed in this thesis, the algorithm was configured to generate exactly two jets per event to align with the expected electron and dijet final state of the investigated process.

Analysis level jet reconstruction was decided over `Pythia` stage jet construction to exclude electrons or positrons from the particle groups considered for jet formation. Hence, the

final jet constructs do not incorporate the lepton identified in the event.

4.3 Sample Generation

4.3.1 Background Samples

In this thesis, we consider two instances of the centrally produced background, the $Z \to b\bar{b}$, and the $Z \to c\bar{c}$. The samples used correspond to those centrally generated in the `Winter2023` campaign by the FCC collaboration [71]. A third background considered is the 4-body final state discussed in Section 3.2, which we locally generate using the `MG5` Process Card found in Appendix B.2. The 4-body final state is expected to be the hardest to differentiate from our signal. It is generated in place of heavier jet final states, such as those originating from $Z \to \tau\bar{\tau}$, so that we only consider background with relevant topology relative to our signal process. Statistics from the backgrounds, including cross-section, non-normalised MC events[11] and production luminosity are given in Table 2.

Process	$\boldsymbol{\sigma}$(pb)	**Monte-Carlo events**	**Production** $\mathcal{L}$ (fb^{-1})
Z $\to b\bar{b}$	6.65×10^{3}	4.39×10^{8}	6.60×10^{1}
Z $\to c\bar{c}$	5.22×10^{3}	4.98×10^{8}	1.15×10^{2}
Z $\to$ 4body	1.40×10^{-2}	1.00×10^{5}	7.14×10^{3}

Table 2: Summary table to show the cross section, raw-events, and production luminosity for each of the background processes considered for this analysis.

For the analysis considered here, we will scale our events to a luminosity of 10 fb^{-1}, for which we have sufficient background statistics. In Chapter 6, results will be scaled to the expected luminosity of the FCC-ee at $\mathcal{L} = 150$ ab^{-1}, though it is critical to note that very small production luminosity in the central background processes impose a significant statistical challenge for making commentary on the discovery potential of HNLs at the true FCC-ee luminosity.

4.3.2 Signal Sample Generation

Signal samples, as described in the previous section, are generated using the `HeavyN` model in `UFO` and generated in `Madgraph`. We locally generate 56 samples with masses in the range: $10 \leq m_N \leq 80$ GeV with mixing angles squared between: $10^{-4} \leq |U_{eN}|^2 \leq 10^{-10}$. An example of the `MG5` Process Card for the mass point with $m_N = 10$ GeV, $|U_{eN}|^2 = 10^{-8}$

[11] Elsewhere, we use the term *raw* to mean the Monte-Carlo events which not normalised to any specific luminosity.

can be found in Appendix B.1. After several campaigns, it was possible to generate almost all of the parameter points in our range. Unfortunately, at very low cross sections, the production of the n-tuples[12] becomes highly susceptible to crashes, and for the signal point at $m_N = 30$ GeV, $|U_{eN}|^2 = 10^{-10}$ we were unable to produce a stage 1 flat n-tuple. This ultimately does not affect our analysis at 10 fb^{-1} where there is no such extension into this order of couplings for the 90% CL exclusion limit. For the scaling of results to 150 ab^{-1}, this point is estimated from the surrounding points, as leaving it without a value would skew the interpolated delimitation line. At this scale, statistical uncertainty arising from the luminosity is anyhow much larger than the uncertainty in interpolation of the considered point.

Information regarding the cross-section, raw events and the production luminosity for a 50 GeV HNL at a range of mixing angles is given in Table 3. A full tabulation featuring the cross sections of each signal point is found in Appendix C.1.

Process ($m_N = 50$ GeV)	$\boldsymbol{\sigma}$(pb)	**Monte-Carlo Events**	**Production** $\mathcal{L}$ (fb^{-1})
$\|U^2\| = 10^{-4}$	2.27×10^{-1}	1.00×10^5	4.41×10^5
$\|U^2\| = 10^{-5}$	2.27×10^{-2}	1.00×10^5	4.41×10^6
$\|U^2\| = 10^{-6}$	2.27×10^{-3}	1.00×10^5	4.41×10^7
$\|U^2\| = 10^{-7}$	2.27×10^{-4}	1.00×10^5	4.41×10^8
$\|U^2\| = 10^{-8}$	2.27×10^{-5}	1.00×10^5	4.41×10^9
$\|U^2\| = 10^{-9}$	2.27×10^{-6}	1.00×10^5	4.41×10^{10}
$\|U^2\| = 10^{-10}$	2.27×10^{-7}	1.00×10^5	4.41×10^{11}

Table 3: Summary table showing the cross section, raw-events, and production luminosity for the range of mixing angles considered at a fixed mass of 50 GeV.

4.3.3 Reconstructed Events

Within the `FCCAnalyses` framework, events can be distinguished as "truth" or "reconstructed". Truth data corresponds to the events generated using the physics processes in the process cards, whilst the reconstruction requires parsing the truth events through the `Delphes` IDEA card. For a small sample of the central $Z \to b\bar{b}$ background process, Figures 23, 24, 25 and 26 compare the distribution of truth and reconstructed events for both electron energy and electron η. These variables are chosen since these are the variables for which cuts are made on the truth-level. In Figures 24 and 26, a cut is made on the truth level events according to $E \geq 2.0$ GeV, $p_T \geq 0.1$ GeVand $|\eta| \leq 2.56$, as

[12] A flat n-tuple is used to store event information in a tabular format, where each row corresponds to an event and each column corresponds to a variable or feature of the event. In the case of this analysis, the n-tuples are stored as `ROOT` files.

described in the `Delphes` IDEA card.

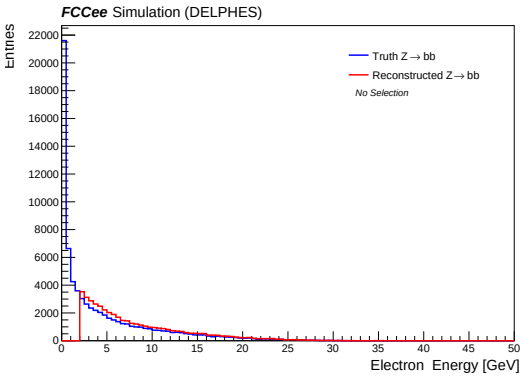


Figure 23: Distribution of an sample of Z $\rightarrow b\bar{b}$ for the lead electron energy at the level of truth and reconstruction, with no cut applied on the truth. We identify a clear 2 GeV cut on the reconstructed energy.

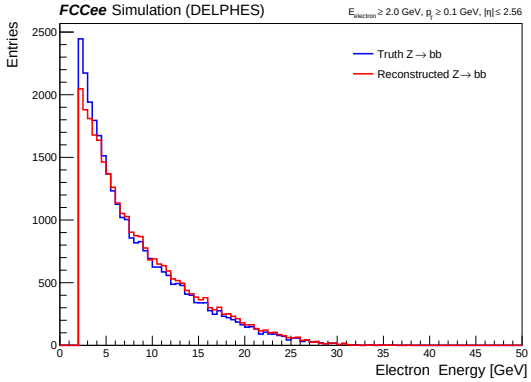


Figure 24: Distribution of an sample of Z $\rightarrow b\bar{b}$ for the lead electron energy at truth level and reconstructed level, with selection cuts: $E \geq 2.0$ GeV, $p_T \geq 0.1$ GeVand $|\eta| \leq 2.56$.

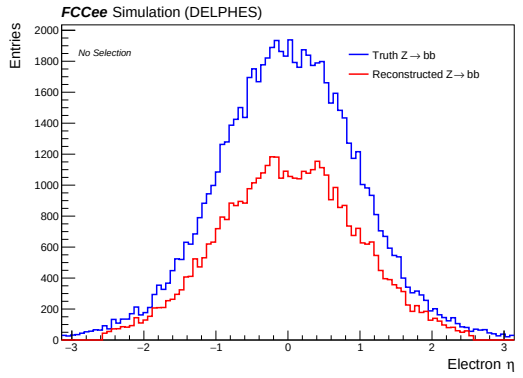


Figure 25: Distribution of an sample of Z $\rightarrow b\bar{b}$ for the lead electron η at the level of truth and reconstruction, with no cut applied on the truth. We identify a clear cut on reconstruction at $\eta = 2.56$.

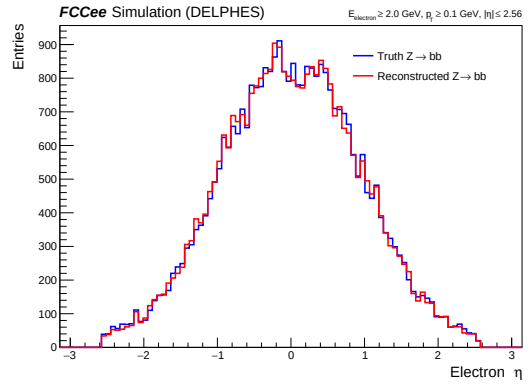


Figure 26: Distribution of an sample of Z $\rightarrow b\bar{b}$ for the lead electron η at truth level and reconstructed level, with selection cuts on truth matching $E \geq 2.0$ GeV, $p_T \geq 0.1$ GeVand $|\eta| \leq 2.56$.

After applying the pre-selection posited by the IDEA card, we see that the distributions of the truth and reconstructed samples indeed closely match. The small differences present after the cuts is due to the "smearing" which is applied when reconstructing the detector conditions to reflect the finite resolution of the sub-detector systems. These plots serve to give some idea of the action of the detector reconstruction on the samples generated both centrally and locally.

5 Analysis

Three analyses strategies were used to optimise the sensitivity in the electron dijet channel to our benchmark HNL models. In Section 5.4, a cut and count study was done to replicate the work found in [72]. This work mostly focuses on finding variable cuts satisfying criteria on three main mass points: 20 GeV, 50 GeV and 70 GeV for a fixed mixing angle of $|U^2| = 10^{-6}$. In Section 5.5, an extension of this work is composed by using boosted decision trees (BDTs) to optimise the sensitivity by training a model for each signal point. Finally, a deep neural network (DNN) is trained on each of the signal points to compare different machine learning techniques and their capacity to optimise the sensitivity of HNL searches at the FCC-ee.

5.1 Normalisation

The flat n-tuples produced through the `FCCAnalyses` framework are categorised as being central samples (i.e. centrally produced by the FCC collaboration), such as $Z \to b\bar{b}$ or $Z \to c\bar{c}$, or locally generated, such as our 4-body background and HNLs. In any case, there are a given number of raw MC events in each n-tuple. This number of events is irrespective of target luminosities and hence we must apply a normalising factor in order to properly make an assessment of the FCC-ee conditions.

The number of events for a given integrated luminosity is given by the relation below in Equation 24, where σ is the event cross section and $\mathcal{L}$ is the integrated luminosity. The luminosity L is a measure of the number of particles passing through a unit of area per second, and hence the integrated luminosity, $\mathcal{L} = \int L \, \mathrm{dt}$, is a measure of the number of events expected per unit of cross section.

$$n = \sigma \times \mathcal{L} \tag{24}$$

The number of events produced for the background n-tuples is tabulated below, alongside three of the HNL signals, in Table 4. This table is similar in nature to Tables 2 and 3, but this time with a focus on the expected number of events at two luminosities, at 10 fb^{-1} and 150 ab^{-1}.

Process	$\boldsymbol{\sigma}$ (pb)	**Monte-Carlo events**	**Events at $\mathcal{L}$ =**	
			10 fb^{-1}	150 ab^{-1}
Z $\rightarrow b\bar{b}$	6.65×10^{3}	4.39×10^{8}	6.65×10^{7}	9.98×10^{11}
Z $\rightarrow c\bar{c}$	5.22×10^{3}	4.98×10^{8}	5.22×10^{7}	7.82×10^{11}
Z $\rightarrow$ 4body	1.40×10^{-2}	1.00×10^{5}	1.40×10^{2}	2.10×10^{6}
20 GeV, $\lvert U^2 \rvert = 10^{-6}$	3.77×10^{-3}	1.00×10^{5}	3.80×10^{1}	5.66×10^{5}
50 GeV, $\lvert U^2 \rvert = 10^{-6}$	2.27×10^{-3}	1.00×10^{5}	2.30×10^{1}	3.40×10^{5}
70 GeV, $\lvert U^2 \rvert = 10^{-6}$	9.06×10^{-4}	1.00×10^{5}	9.00×10^{0}	1.36×10^{5}

Table 4: Processes used for this analyses alongside the cross section information, the produced number of events, and the expected number of events at 10 fb^{-1} and 150 ab^{-1}. Event numbers are given to the nearest integer.

Table 4 alludes to a key problem for studies such as these in the early stages. Such an enormous luminosity for the FCC-ee Z run leads to immense pressure on the Monte-Carlo statistics generation to match those expected at the collider. At this stage, we simply do not possess adequate statistics to make realistic commentary on the discovery potential of our HNLs at the true FCC-ee integrated luminosity since our central samples do not contain a sufficiently high number of events. Therefore, in this study, we normalise to a luminosity of 10 fb^{-1}, for which we have realistic simulated samples.

In order to normalise our MC events, we use the simple formula below in Equation 25, where N is the normalising factor.

$$N = \frac{\mathcal{L}_{target}}{\mathcal{L}_{sample}} = \frac{\mathcal{L}_{target} \times \sigma}{n_{sample}} \tag{25}$$

This equation, however, must also be robust enough to account for selection efficiencies. For example, if we do not observe the full number of events in the sample, n_{MC}, but instead apply pre-selection to the events, resulting in only a fraction surviving, n_{sel}, we must adjust accordingly. We can write the normalizing factor for events surviving a given selection in terms of the selection efficiency, $\xi = \frac{n_{sel}}{n_{MC}}$, as shown in Equation 26 below.

$$N = \frac{\mathcal{L}_{target} \times \sigma}{n_{sample}} \times \xi \tag{26}$$

5.2 Sensitivity

Statistical significance quantifies the probability (p-value) of observing a given excess to the expected background. In particle physics, the 'gold standard' for claiming discovery

is $Z = 5\sigma$. The probability that the background-only hypothesis can describe an observed excess of events in a given region corresponding to this value is 3×10^{-7}, meaning this would only occur as a random fluctuation around 1 in every 3.5 million cases [73].

$Z = 2\sigma$ on the other hand represents a probability of 0.05, which is not robust enough to claim discovery but is a good indication of a statistically significant observation, often called the 95% confidence level (CL) limit. For our sensitivity plot, we will use $Z = 2\sigma$ to and attempt to increase the parameter space covered by the 95% CL limit.

To quantify the significance of observing n total events with an expected number of background events, b, we use the definition given in the ATLAS note found in Reference [74], written in Equation 27 below. In Equation 27, σ denotes the uncertainty on the background events, not to be confused for where it is otherwise used for the cross section or standard deviation. For the purposes of this study, we employ a 10% uncertainty, $\sigma = 0.1$.

$$Z = \sqrt{2\left(n \cdot \ln\left[\frac{n\,(b+\sigma^2)}{b^2+n\sigma^2}\right] - \frac{b^2}{\sigma^2}\ln\left[1+\frac{\sigma^2(n-b)}{b\,(b+\sigma^2)}\right]\right)} \tag{27}$$

To visualise the distribution of each discriminating variable, we produce a histogram divided into a given number of bins. For each bin, the number of background and signal events are counted, wherein significance is then calculated by comparing the quantity of background to the total number of events.

In order to quantify the sensitivity of a given region of interest, we cumulatively count the significance of how ever many bins fall within this region. This can be done from either the left to the right (LR) or the right to the left (RL).

In a LR regime, we start counting from the right most bin, and accumulate the significance all the way to the left. We then make a cut to the optimal threshold, and keep the events to the right of the cut. In the RL, we count from the first bin to the last, and take events to the left of the cut. The difference between these two regimes in shown more concretely in Figures 27 and 28.

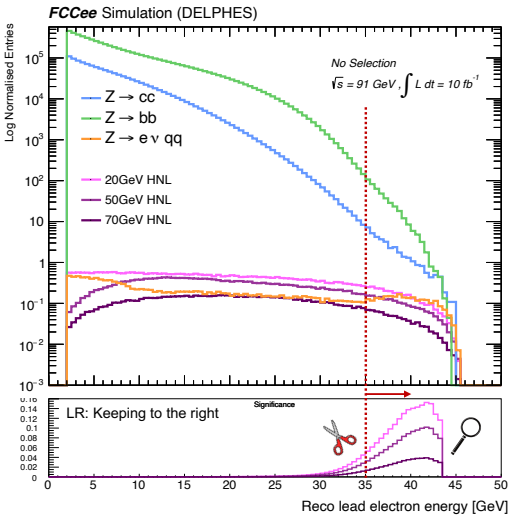


Figure 27: Calculation of the significance in a LR regime, making a cut on the distribution by keeping everything to the right of the cut made, e.g. $E > 35$ GeV.

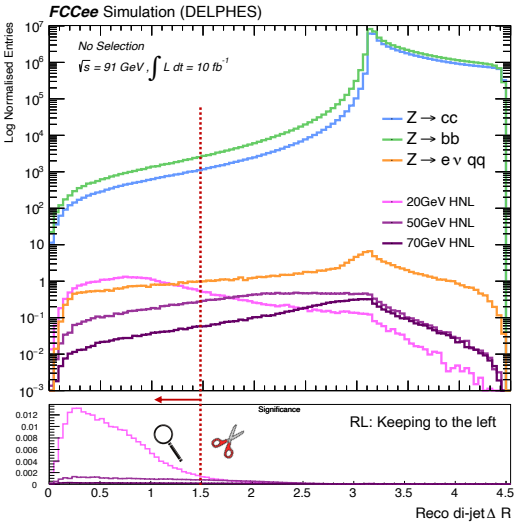


Figure 28: Calculation of the significance in a RL regime, making a cut on the distribution by keeping everything to the left of the cut made, e.g. $\Delta R < 1.5$.

In Figures 27 and 28, the dashed line represents the position where a possible cut might be made, and the arrow orthogonal to it represents the region kept after making this cut, reinforced by the illustrative scissors and magnifying class. One region is kept for analysis and the significance is counted, whilst the other is cut away. This is the essence of a cut and count analysis.

Of course, whilst maximising significance is important, the optimal cut is not completely clear from these distributions alone since it is also necessary to account for the yield of a cut, i.e. how much of the absolute signal and background remains. Additionally, some cuts become a little more difficult to find when we consider the entire signal parameter space, since for many variables, the signal behaves differently depending on its properties. For instance, a signal point with a low mass and a very small mixing angle would be considered long lived, and if we plot the decay length of that signal, it would be difficult to make a global cut across all points without losing discrimination unnecessarily in some regions of the parameter space. We will return to the distributions and optimising the cut and count analysis later in this chapter.

5.3 Discriminating Variables

The essence of any sensitivity study is to identify kinematic differences between signal and background regions for a given variable. To this end, we will set out the variables of interest and show their distributions for our three background processes and additionally for three chosen signal points at 20 GeV, 50 GeV and 70 GeV for a mixing angle of

$|U^2| = 10^{-6}$ in all cases, since this forms the basis of the cut and count study. Then, optimisation will be performed using techniques involving boosted decision trees and deep neural networks. The findings of this analysis will be explored throughout Section 5, and summarised in Section 6.

5.3.1 Angular Variables

Many of the variables of interest are related to the angle at which objects enter the sub-detector systems, typically the tracker. Different particles will enter the detector at different angles relative to the interaction point, for each specific interaction. The main discriminating variables of this class that we will discuss include the pseudorapidity, (η), the angles ϕ and θ, the 3D angle Ψ and a composite quantity, ΔR formed from η and ϕ. The definitions of these angles can by understood by the schematic in Figure 29.

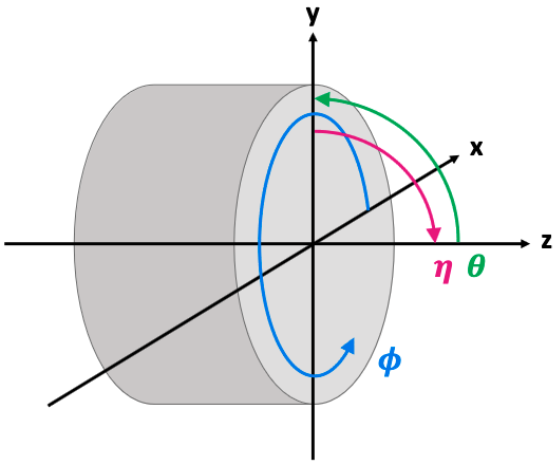


Figure 29: Definitions of angular variables η, ϕ and θ relative to the z-axis beamline (Image: taken from [75]).

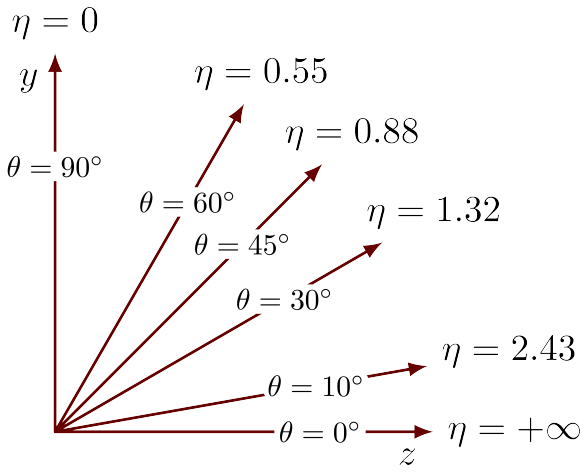


Figure 30: Relationship between angle relative to the Z axis, θ and the pseudorapidity η.

The relationship between η and θ is shown in Figure 30, and formalised in Equation 28. Both angles describe the position in the yz plane, but η is Lorentz invariant, and is hence chosen typically instead of θ.

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{28}$$

The azimuthal angle in the xy plane is bound between 0 and 2π radians, and we expect a flat distribution since the collision occurs along the z axis and hence the initial transverse momentum is zero. The result is this is that there is spherically symmetric choice of final position for decay products to interact with the detector volume, with no motivation for any one azimuthal angle over another. Indeed, the ϕ distribution is shown for our HNL

signals and our background processes in Figure 31.

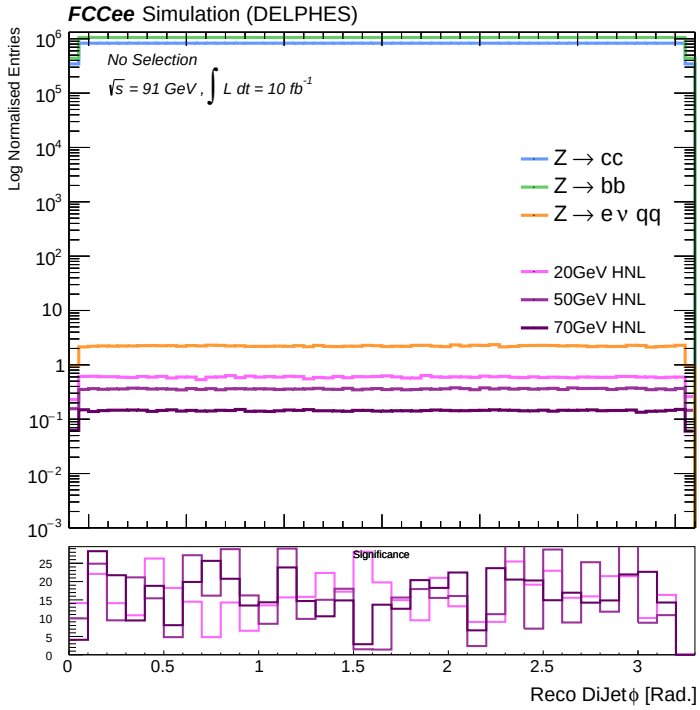


Figure 31: Distribution of the azimuthal angle, ϕ, for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

When trying to maximise the discrimination between the signal and background, we must consider variables which exhibit distinct behavioural differences between our HNLs and the SM background. Angular distance is indeed a promising candidate for a category of variable, since the background processes $Z \to b\bar{b}$ and $Z \to c\bar{c}$ will produce many back to back jets with a high angular separation, whereas the signal will produce more collimated jets since it decays via an off-shell W boson.

The angular distance between objects in the detector is taken to be Lorentz invariant along the longitudinal plane by using the azimuthal angle ϕ, and the pseudorapidity, η. This distance is characterised as in Equation 29. In the case of our study, we use two instances of ΔR, once between the two jets in the di-jet final state, and again between the dijet system and the lead electron. These will later be denoted as the electron dijet ΔR, or ΔR_{ejj}, and the dijet ΔR, or ΔR_{jj}. For a small ΔR, two objects are said to be collimated. Conversely, for large ΔR, two objects are isolated from one another. The distributions for ΔR_{ejj} and ΔR_{jj} are given in Figures 32 and 33 respectively.

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{29}$$

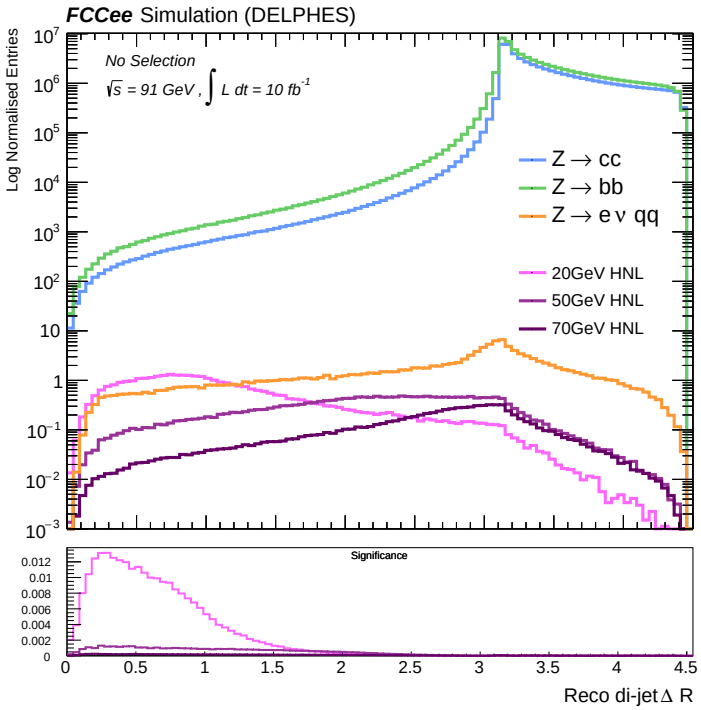


Figure 32: Distribution of the dijet ΔR for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

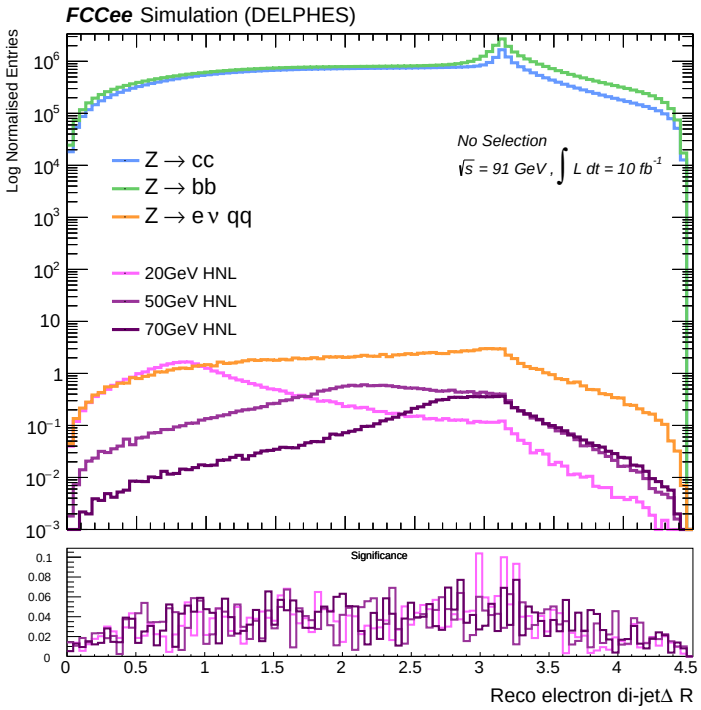


Figure 33: Distribution of the electron dijet ΔR for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

Another metric composed to study the angular separation of the jets is the 3D angle

between the jet four vectors, j_1 and j_2. The definition of Ψ is given in Equation 30, and in practice differs only slightly from the ΔR_{jj}, encoding much of the same information. Small kinematic differences between these two variables exist, and can be seen by comparing the two plots, which leads to differences in their discriminating power as seen in the lower panel of both Figure 32 and 34; hence, both are kept for the cut and count consideration.

$$\Psi = \cos^{-1}\left(\frac{j_1 \cdot j_2}{|j_1|\,|j_2|}\right) \tag{30}$$

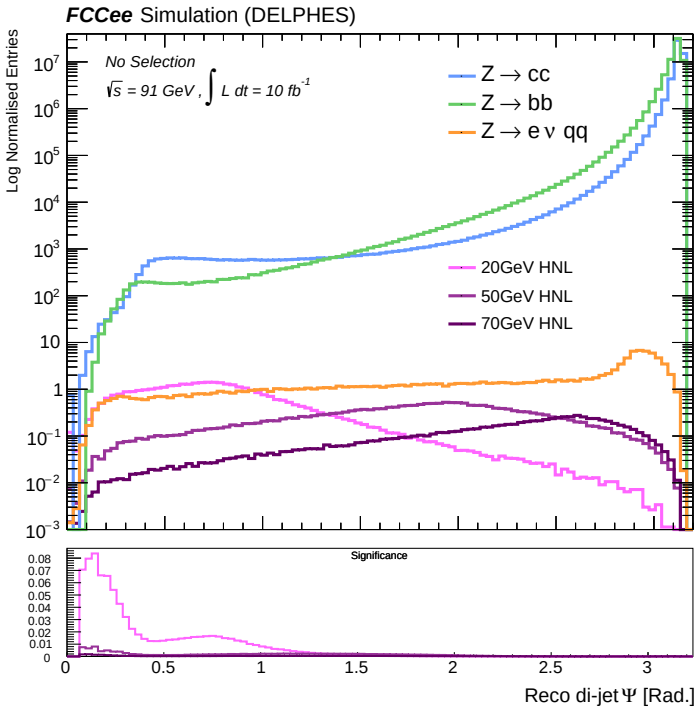


Figure 34: Distribution of the dijet 3D angle, Ψ, for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

The final angular variable we consider throughout this analysis is given by the θ of the missing energy. Unlike in a hadron collider, lepton collisions involve point-like interactions where the total energy is known beforehand. In hadron-hadron collisions, we are subjected to the probability distribution functions within the sea of quarks and gluons, and can only infer neutrinos through missing transverse momentum. The result of this difference is that we can infer neutrinos in our final state through total missing energy at the FCC-ee, and the direction of this can be expressed through θ. We expect a somewhat flat distribution since the neutrino is a product arising from the secondary decay and there is no a priori reason this should happen in any one place of the detector relative to the interaction point. The distribution is shown in Figure 35.

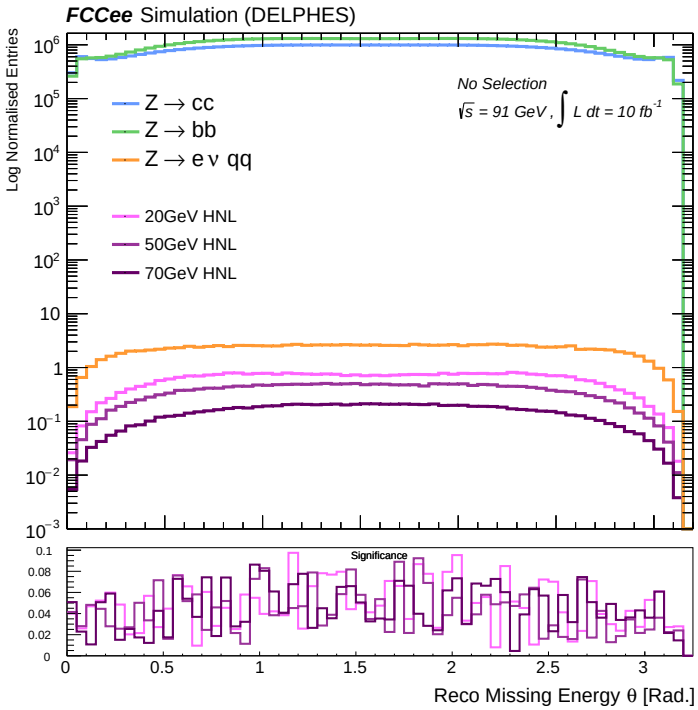


Figure 35: Distribution of the missing energy θ for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

Though Figure 35 does not initially appear to show good discrimination, we cannot discount its value, as its discriminating power may become evident when used in conjunction with other variables. Such non-linear combinations of cuts are a key reason why machine learning techniques may be more suitable for high-dimensional data compared to typical cut-and-count approaches.

5.3.2 Energy Variables

The second class of variables important for our consideration is those concerning the energy of the decay products. In the Z run of the FCC-ee, the centre of mass energy is equal to 91 GeV, meaning the invariant mass of the decay products will sum to this energy. This puts constraints on the relative energies of the decay products, and hence we should see quite distinct properties of the decay products for high mass compared to low mass HNLs in terms of their allowed energies. This could offer a clear ground for further discrimination.

The first of these variables is the lead electron energy produced from the secondary vertex in the decay of the HNL, or indeed originating from the background through one of b or c jets. In the case of the HNL, the electron will have a higher mean energy until we begin to approach the mass of Z boson, where the electron energy will be constained by the conservation of energy since the HNL will be produced with low momentum. The plot of

the lead electron energy is shown in Figure 36.

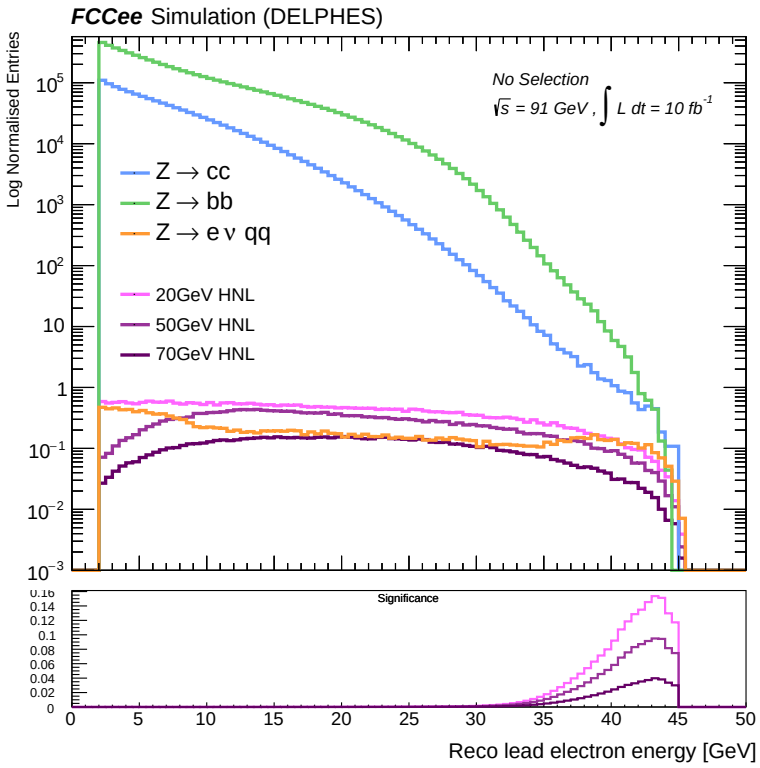


Figure 36: Distribution of the lead electron energy for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

In Figure 36 we indeed see the earlier described `Delphes` cut at E > 2 GeV, meaning that any later cut we make on the lead electron energy will have a higher yield than other possible cuts, since we already have some form of selection efficiency present with this "hidden" truth cut. We see good discrimination in the higher energies between approximately 35 and 45 GeV, where the background drops of against a relatively constant signal.

The second consideration is made for the energy of the neutrino produced alongside the HNL in the Z decay, characterised by the total missing energy. The distribution of the missing energies is given in Figure 37.

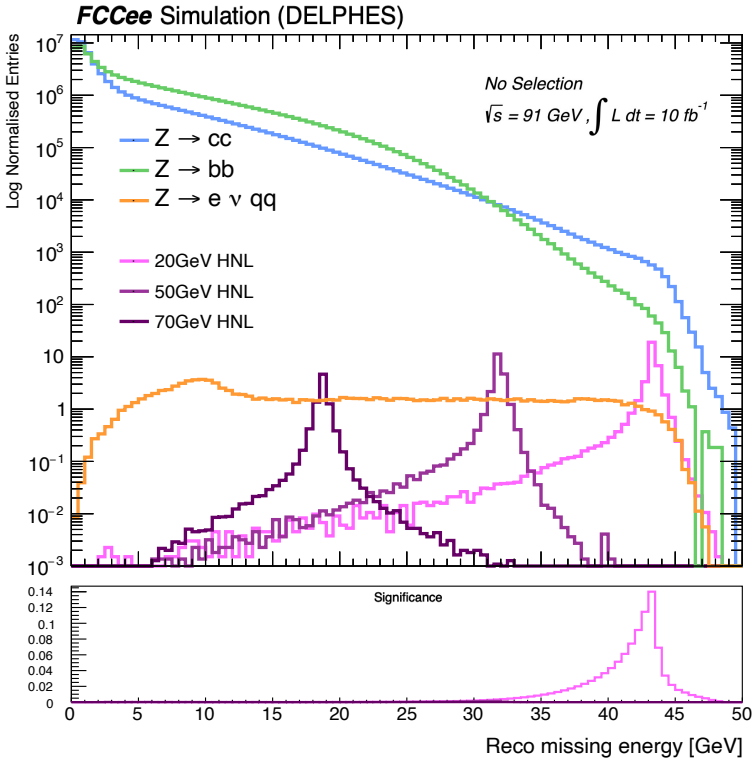


Figure 37: Distribution of the missing energy for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}.

Here we see good discrimination for the low mass point, which is produced alongside a high momentum neutrino. Conversely, the higher mass signal points have a lower discrimination. Once again this is since we approach the Z mass and constrain strongly the energy with which we can produce the neutrino.

5.3.3 Vertex Variables

The last class of variables considered, mostly for the machine learning optimisation, is the vertex based variables. For different signals and backgrounds, we would see a varied vertex topology. We first consider the tracks by assessing the total number of tracks, the number of primary tracks, and the χ^2 of the primary vertex. In this context, the total number of tracks counts all tracks originating from both the primary vertex, and any secondary or tertiary interaction vertices. Primary tracks aim to include only those originating from the initial interaction. The χ^2 of the primary vertex is a measure of how well the reconstructed vertex fits the tracks associated with it, with lower values indicating a better fit [76]. The distribution for the number of total tracks is shown in Figure 38, and is shown for the primary tracks in Figure 39.

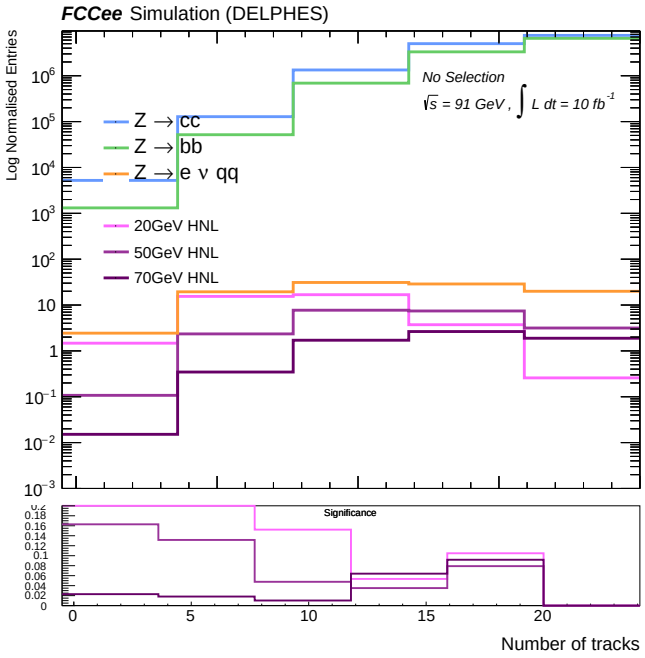


Figure 38: Distribution of the total number of tracks for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}

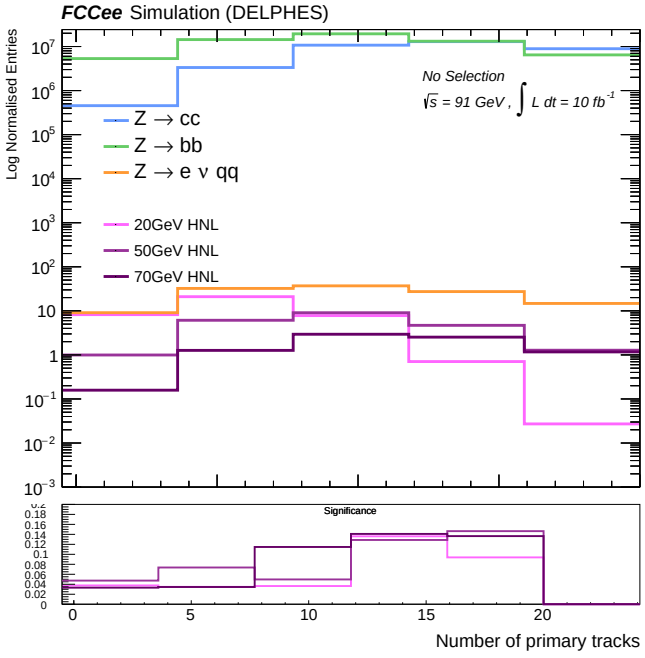


Figure 39: Distribution of the number of primary tracks for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}

The distribution for the χ^2 of the primary vertex is shown in Figure 40 below. We see a clear peak for most signal and background for a χ^2 of 1, which implies good track reconstruction. For the $b\bar{b}$ final state, we see less events for low values of χ^2. We could

speculate that this behaviour is expected since the b-jets are less promptly decaying than for the $c\bar{c}$, and have high track multiplicity. As a result, reconstruction of the primary vertex becomes very challenging since they are defined via the secondary vertex, and hence there is a larger associated χ^2 value [77].

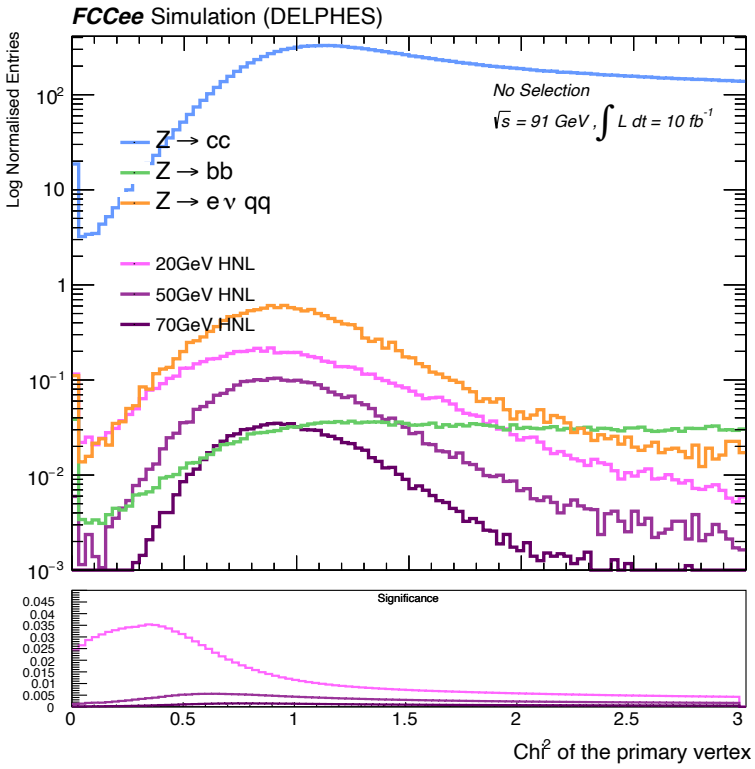


Figure 40: Distribution of the primary vertex χ^2 for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}

We consider displaced signatures by searching for any discrimination in the lifetime of the particle, mostly through investigating the transverse impact parameter, d_0. The transverse impact parameter is an indirect measurement of the lifetime, and is a projection in the xz plane of the displacement between the primary and secondary vertex. This is shown in Figure 41.

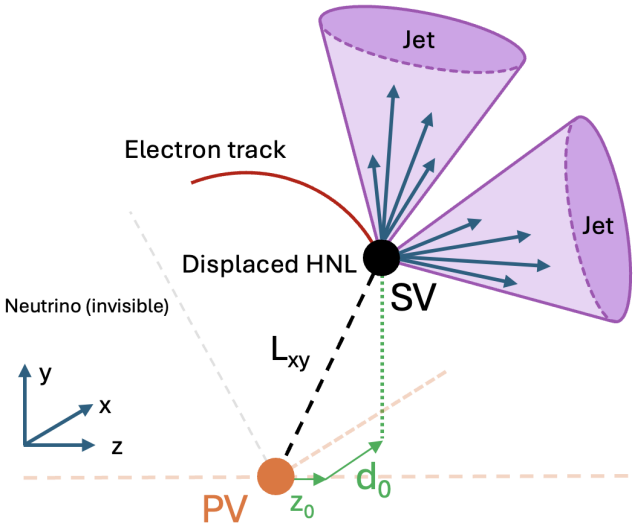


Figure 41: Characteristics of an electron di-jet system initiated by a displaced HNL decay: secondary vertex (SV), as well as a longitudinal (z_0) and transverse impact parameter (d_0) of individual tracks from their point of closest approach to the primary vertex (PV).

The distribution of the absolute d_0 for our three signal points against the three SM background processses is shown in Figure 42 below.

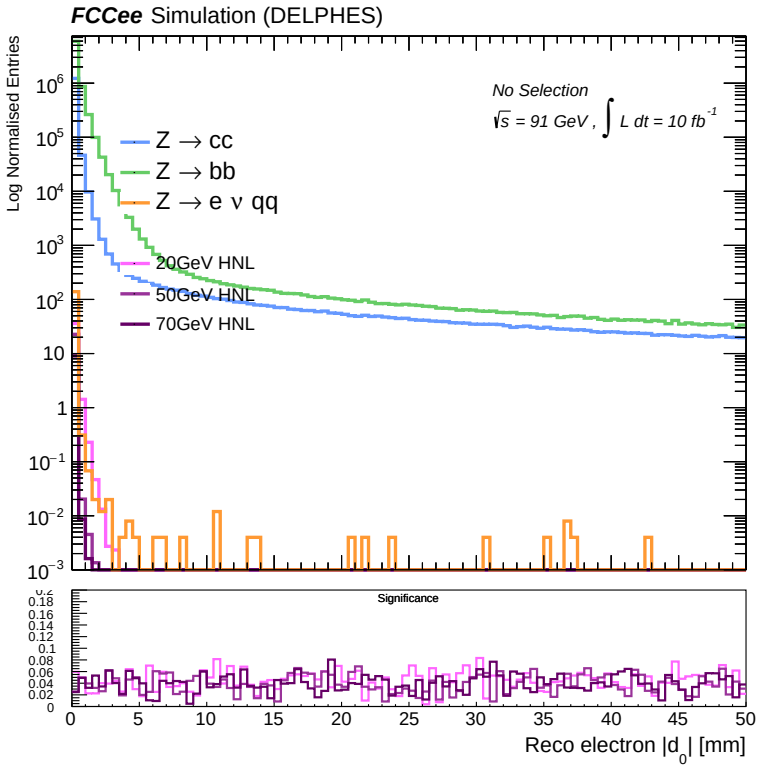


Figure 42: Distribution of the transverse impact parameter for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}

The significance of the transverse impact parameter is an additionally useful measure, shown in Figure 43. This variable represents the ratio between the track impact parameter and its uncertainty [78]. Comparing this distribution to the absolute d_0 shown in Figure 42, we see a clearer separation of the signal points, and as such, for future consideration of the lifetime for separating the study into prompt and displaced signals, we use the d_0 significance as the relevant parameter.

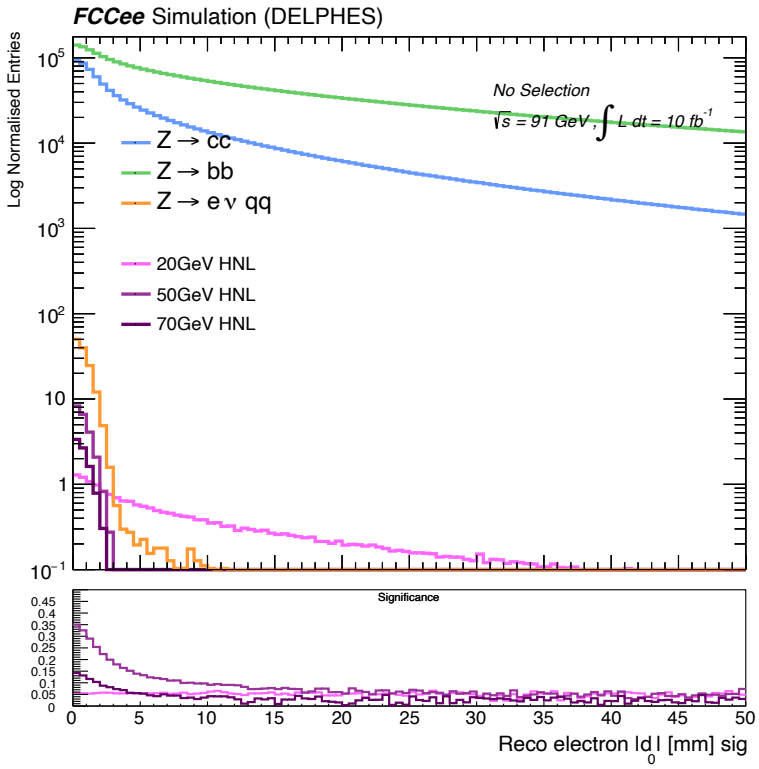


Figure 43: Distribution of the transverse impact parameter significance for three benchmark HNL mass points and the three background processes at an integrated luminosity of 10 fb^{-1}

5.4 Cut and Count

In order to use the variables set out in the previous section to conduct a cut and count study, regions with maximised sensitivity were taken as the optimal kinematic cuts. To perform the analysis, three mass points were considered and a series of sequential cuts were applied to the signal points. For the full cut and count study, it is beneficial to refer to [72]. Here we will summarise our replication of the main results as a foundation for the BDT and DNN optimisation. A summary of the selections applied is found in Table 5.

Variable	Selection
Missing energy	$> 12\,\text{GeV}$
Leading electron energy	$> 35\,\text{GeV}$
3D di-jet Angle	$< 2.4\,\text{rad}$
Di-jet – Electron ΔR	< 3

Table 5: Summary of the selections chosen for the cut and count study

Each cut is made sequentially, and the number of events surviving each selection is tabulated in the cut-flow shown in Table 6.

Selection	20 GeV	50 GeV	70 GeV	4-body	$Z \to c\bar{c}$	$Z \to b\bar{b}$
No selection	10^5	10^5	10^5	10^5	4.9×10^8	4.4×10^8
$E_{\text{miss}} > 12$	9.9×10^4	9.9×10^4	9.9×10^4	7.8×10^4	3.3×10^7	5.6×10^7
$E_{\text{miss}} > 12 \wedge E_{e^-} > 35$	8079	8090	8541	5206	101	817
$E_{\text{miss}} > 12 \wedge E_{e^-} > 35 \wedge \psi < 2.4$	7780	7290	8333	4853	60	46
$E_{\text{miss}} > 12 \wedge E_{e^-} > 35 \wedge \psi < 2.4 \wedge \Delta R < 3$	7478	5035	3017	3184	2	1

Table 6: Summary of the cut-flow for the three mass points considered as well as the standard model background processes.

It is worth noting that the selections made are especially powerful on the $b\bar{b}$ and $c\bar{c}$ background processes, which we expected as a consequence of our 'clean' HNL signatures. We see just one and two events survive respectively, with the lead electron energy selection having the largest impact of the selections, partly since this removes events already hidden from the lead electron energy spectrum below the `Delphes` cut. One issue with such an impressive cut at this level is that scaling to the full luminosity becomes an even greater challenge, illustrated below in Table 7.

Events	Selection	20 GeV	50 GeV	70 GeV	4-body	Z $\to c\bar{c}$	Z $\to b\bar{b}$
MC events	All cuts	7478	5035	3017	3184	2	1
10 fb^{-1} events	All cuts	2.991	1.007	0.302	4.204	0.213	0.151
150 ab^{-1} events	All cuts	44,868	15,105	4,525	63,055	3,191	2,265

Table 7: Raw MC cutflow compared to normalised cut flow at 10 fb^{-1} and 150 ab^{-1}

In Table 7, we see that we are, in effect, attempting to model a distribution of $\mathcal{O}(10^3)$ events in the case of $b\bar{b}$ and $c\bar{c}$ with just 1 and 2 MC events respectively. In Figure 44, which shows the background and signal distribution combined after all cuts, we see

the effects of this especially for our high mass signal point will lower yield under these selections, where to either side of the signal peak we see erroneous peaks from background noise.

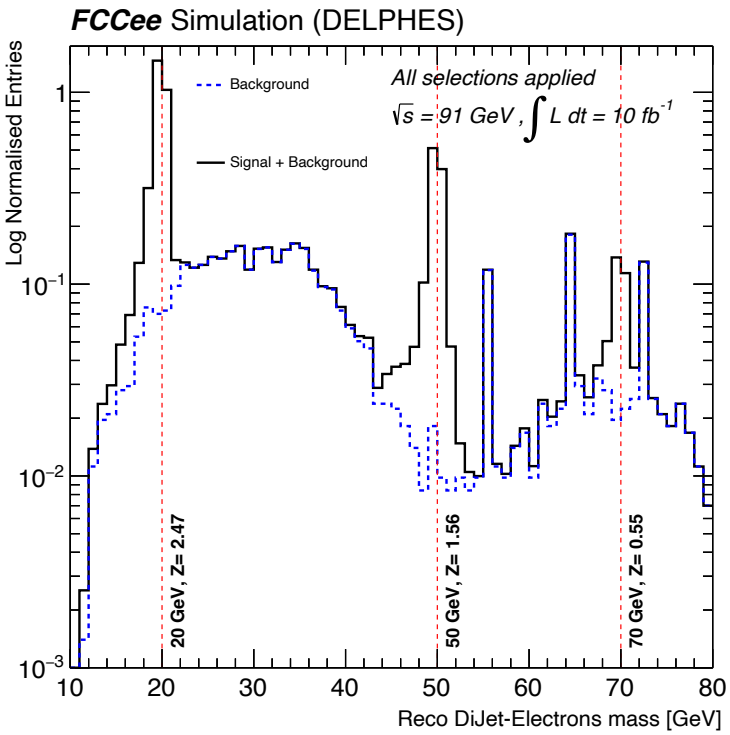


Figure 44: Distributions of the three signal points after all cuts made.

Performing our selections globally across all signal mass points, we can calculate the significance of the signal to form a 2σ limit plot based on these selections. This is shown in Figure 45.

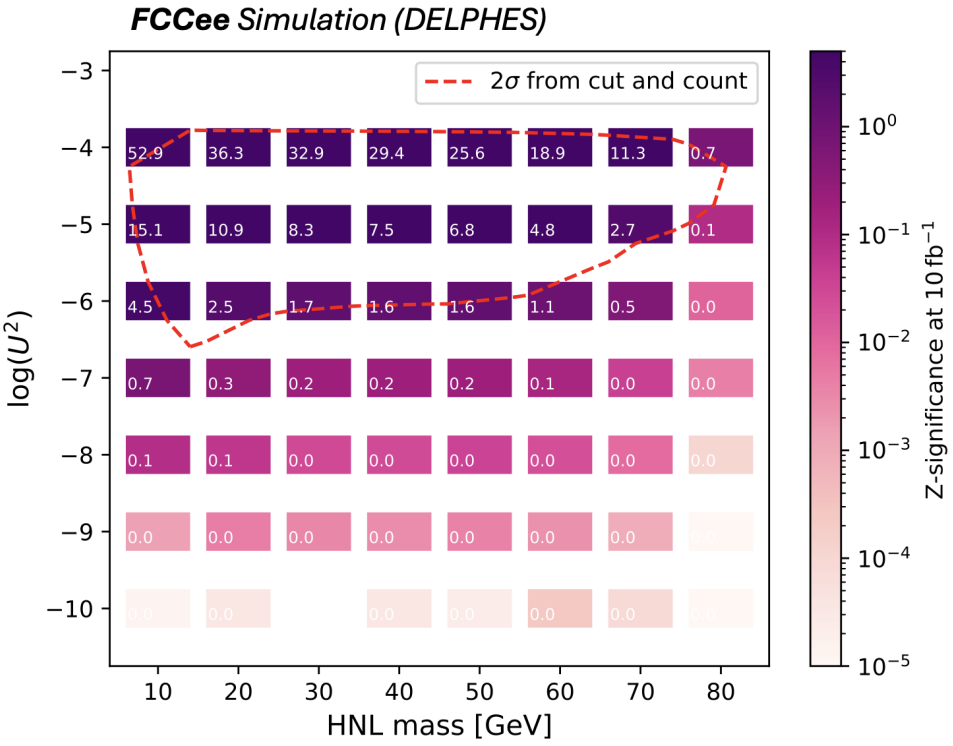


Figure 45: Sensitivity plot after cut and count analysis for HNL masses between 10 and 80 GeV for mixing angles between $|U|^2 = 10^{-4}$ and 10^{-10}. This plot is normalised to an integrated luminosity of $\mathcal{L} = 10$ fb^{-1}. The red dashed line delimits the extrapolation of 2σ significance.

Figure 45 illustrates the 2σ delimitation of the discovery potential for our HNL models using the global cuts identified in the cut and count study. Employing global selections in this manner leaves significant room for optimization, as more suitable selections can be identified for each model, especially as the topology changes between LLPs and prompt decays. A suitable approach for this optimization is to use machine learning (ML) techniques. To this end, we explore the optimization of the results shown in Figure 45 using BDTs and DNNs.

5.5 Boosted Decision Trees

Boosted Decision Trees (BDTs) are a family of supervised[13] machine learning techniques which are commonly used for complex classification problems. Though they have been used for some decades, BDTs have become increasingly popular in HEP applications for both classification and regression, since they are very easy to use 'out of the box', and are well-equipped to handle features of physics data with little fine-tuning, such as class imbalance and missing values. Decision-tree based algorithms have long-standing

[13]Supervised techniques are those which involve training on labelled data, meaning the model learns based on an expected outcome, rather than finding arbitrary trends in data distributions.

influence in computing and data-analysis, but a single decision tree algorithm is itself referred to as a *weak learner*. Boosting refers to producing a multivariate algorithm by creating an ensemble of weak learners. The main idea of boosting is to add new models to the ensemble sequentially. At each iteration, a new weak, base-learner model is trained with respect to the error of the whole ensemble learnt so far; where previously a weak learner might have misclassified a signal or background event, the weights of the model are adjusted and used in the next weak learner. Boosting algorithms construct new base learners (trees) maximally inversely correlated with the gradient of the loss function of the entire ensemble of learners [79]. A schematic for the tree structure of a BDT model is given in Figure 46.

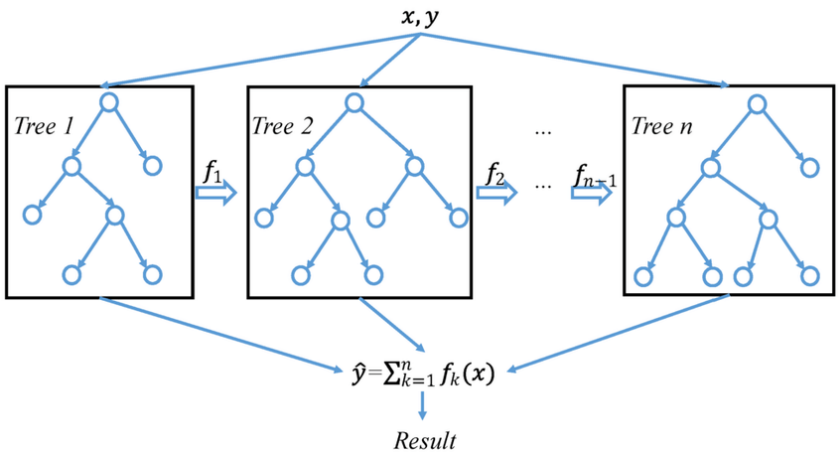


Figure 46: A graphical interpretation of BDT arcitechtures, with data x and associated labels y. The model uses input features f_n and combines weak learning trees to form a boosted ensemble for classification or regression problems (Image: [80]).

`XGBoost` [81] is amongst the most popular of the ensemble boosted decision tree algorithms available. For this analysis, as the first extension of the cut and count study, we trained an `XGBoost` model to classify our MC produced samples of the background and signal, where the truth labels were set to 0 for background, and 1 for signal.

The architecture of the model, such as the maximum depth and the number of estimators is found via cross-validation techniques using `GridSearchCV` imported into `Python` from the `Skikit-Learn` package [82]. The training is performed on raw data based on the training-testing split defined in the data preparation. The raw MC events used for testing are classified by the model, then weighted to a chosen luminosity using Equation 26, wherein significance is calculated and an optimal threshold in the BDT output is chosen to preserve a region with a favourable signal to background ratio.

5.5.1 Data Preparation

Careful data preparation is required in order to best understand and optimise ML frameworks. The flat n-tuples produced via the `FCCAnalyses Framework` are used to store our raw events data for both signal and background. Using `RDataFrame`, the events are processed into dataframe objects and filtered according to a chosen criteria. The aim of this filter is to select a signal rich region of interest in order to reduce computational time spend by the model in regions of the parameter space with very little signal and high levels of background. The lead electron energy shown in Figure 36 was used as the criteria for filtering due to its powerful efficiency in the cut and count procedure. A cut of 15 GeV on the lead electron energy was chosen for this analysis, as it was very efficient at cutting the main background volume of $b\bar{b}$ and $c\bar{c}$, whilst maintaining good efficiency on the signal region and sufficient enough background statistics for training. Table 8 summarises the efficiencies of the cut on the background processes and three benchmark signal points. A full tabulation of the cut efficiencies is appended to Appendix C.2.

Process	No Filter	Lead Electron Energy > 15 GeV	Efficiency (%)
Background			
$Z \to b\bar{b}$	438,538,637	13,074,288	2.98
$Z \to c\bar{c}$	498,091,935	3,875,708	0.78
$Z \to$ 4body	100,000	17,127	17.13
Signal			
20 GeV, $\|U^2\| = 10^{-6}$	100,000	52,508	52.51
50 GeV, $\|U^2\| = 10^{-6}$	100,000	59,982	59.98
70 GeV, $\|U^2\| = 10^{-6}$	100,000	64,387	64.39

Table 8: BDT filter efficiencies for background and signal processes as part of the data pre-processing, where the filter conditions require the lead electron energy > 15 GeV, additionally, the number of final state reconstructed electrons is required to be exactly one.

Table 8 has two interesting features; the first of which was already clear from the cut and count study, which is that the four body final state of the Z decay remains the most irreducible of the three SM background processes, with $\approx 17\%$ efficiency after the filter, whereas the $b\bar{b}$ and $c\bar{c}$ backgrounds are very effectively mitigated by such a cut on the electron energy. The second interesting feature is that the signal efficiency reduces as the mass of the HNL decreases; this is because lower mass HNLs produce a final state electron with lower energy, hence a E > 15 GeV cut is more strict on these mass points.

Of course, once we begin to approach the Z mass, the efficiency begins to drop again for the 80 GeV HNLs, since there is less available energy with which to produce the final state electron.

The features input to the BDT correspond to those found in Table 9. These features were chosen after investigations into the number of events and with feature engineering in the prepossessing stage to assess which variables were correlated and uncorrelated.

Object	Variables
Leading electron	$E, \phi, d_0, \sigma_{d_0}, \Delta R_{ejj}$
Neutrino	E_{miss}, θ
Di-jet system	$\Delta R_{jj}, \phi$
Vertex and tracks	$n_{\text{tracks}}, n_{\text{primary tracks}}, \chi^2_{\text{vertex}}$

Table 9: Variables used as input features for the `XGBoost` model, where ΔR_{jj} represents the dijet ΔR and ΔR_{ejj} represents the electron-dijet ΔR.

In order to understand which features are appropriate for training, a correlation matrix is plotted between pairs of features to parameterise the strength and direction of the relationship between variables. To do this, the Pearson correlation coefficient is calculated and plotted; the diagonal entries are all unity since we are comparing variables with themselves, and the figures are mirrored along the diagonal. A score of 1 indicates perfect correlation, -1 implies perfect anti-correlation, and a score of 0 means there is no measurable correlation between two variables. All features with a correlation score of $\pm$ 0.9 relative to any other feature were investigated as possible candidates for removal for training, and the performance was noted with and without these variables. Ultimately, for the BDT, it was found generally that higher feature dimentionality gave better results, even with high correlation scorses such a those between the dijet ψ and the ΔR_{jj}. The feature correlation matrix for a given signal point is shown in Figure 47.

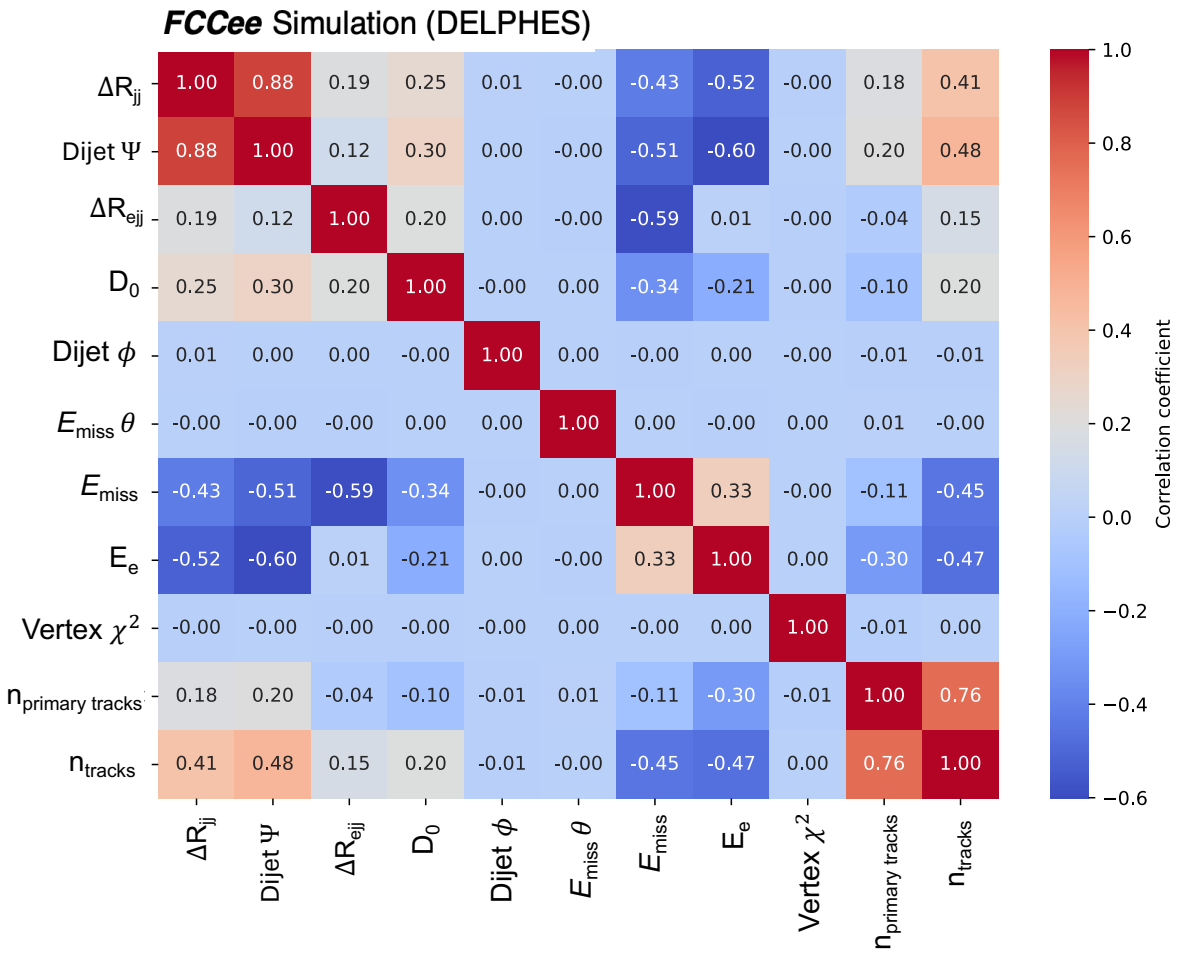


Figure 47: BDT feature correlation matrix for signal point at 70 GeV, $|U|^2 = 10^{-6}$.

Overall, we see good independence between variables through Figure 47. Though, beyond just this signal point, it is noted that as the mass of the HNL becomes smaller the correlation between the electron-dijet ΔR and the diijet ΔR become more correlated. This is understood as low energy electrons produced by low mass HNLs being collimated along the axis of the dijet system, meaning as the dijet system and the electrons are produced with lower energy $\Delta R_{ejj} \to \Delta R_{jj}$ and the correlation score increases.

5.5.2 Training and Testing

A unique BDT is trained for each of the 56 signal points. Each model is trained on a subset of the data surviving the electron energy filter, with a predefined split between training and testing such that once the training of the model is complete, we can assess its generalising power on an unseen set of the data. Each model was trained on 1/2 of the signal sample and 1/3 of the background sample surviving the filter. These values were chosen to maintain a good amount of background in the training relative to the signal since the filter was much more efficient on the background. The raw MC events used for training are booked in Table 10 below.

Process	Training Events	Testing Events
Total Background	5,655,708	11,311,415
20 GeV, $\|U^2\| = 10^{-6}$	26,254	26,254
50 GeV, $\|U^2\| = 10^{-6}$	29,991	29,991
70 GeV, $\|U^2\| = 10^{-6}$	32,194	32,193

Table 10: Distribution of training and testing events for different processes used for the BDT models.

To mitigate the imbalance between the majority (background) and the minority (signal) class, weights were applied to each signal point for each training relative to the imbalance between the classes. This weight is roughly similar to oversampling the minority class, but instead simply tells the BDT to place much more care in misclassifying the signal, and the loss function[14] is penalised more intensely for false-negatives[15].

As mentioned, the specific architecture of the BDT was found using a grid search of the optimal parameters for the maximum depth of the trees and the number of estimators. The *maximum depth* of the trees refers to the maximum number of levels in each decision tree. A deeper tree can model more complex patterns but might lead to overfitting if too deep. Conversely, a shallower tree might be too simplistic and underfit. The *number of estimators* refers to the total number of trees in the forest. Increasing the number of estimators can improve the model's accuracy up to a certain point, beyond which improvements may plateau while computational costs continue to rise. Additionally, going beyond plateau, it is possible the model loses generalising power and will yield worse performance on unseen data. A visual representation of overfitting and underfitting is shown in Figure 48.

[14]The loss function is used in machine learning to measure the difference between the predicted values and the actual values. It quantifies how well a model's predictions match the true outcomes, with the goal of minimising this difference during the training process.

[15]False-negatives occur when a signal sample is falsely labelled as a background event, within a defined threshold. Usually, this means the signal with a truth label of 1 was given a BDT score < 0.5.

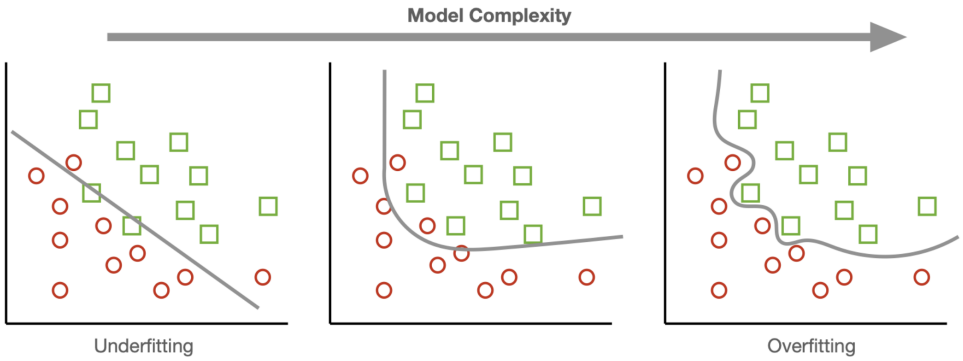
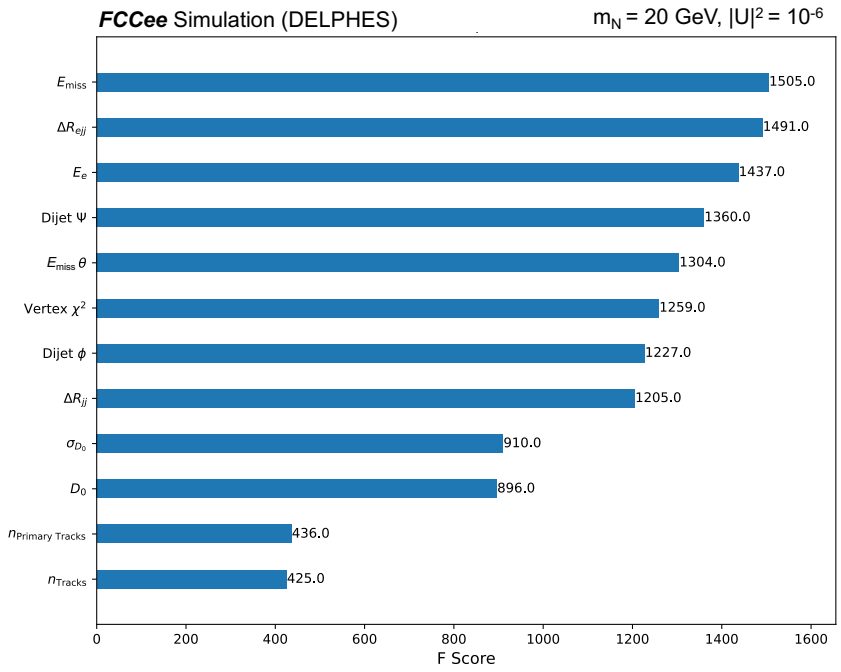
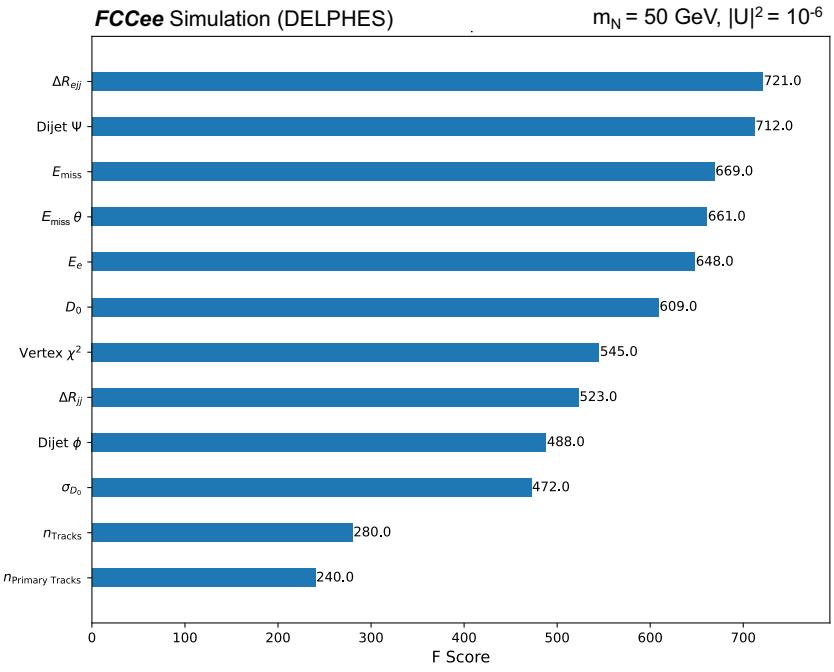


Figure 48: A dataset with two classes shown. Underfitting, on the left, is where the model fails to identify complex trends in data. Overfitting, on the right, is where the model finds highly nuanced patterns in the data and will generalise poorly to new problems. The middle plot is a good attempt at properly fitting the data without capturing too much noise (Image: [83]).

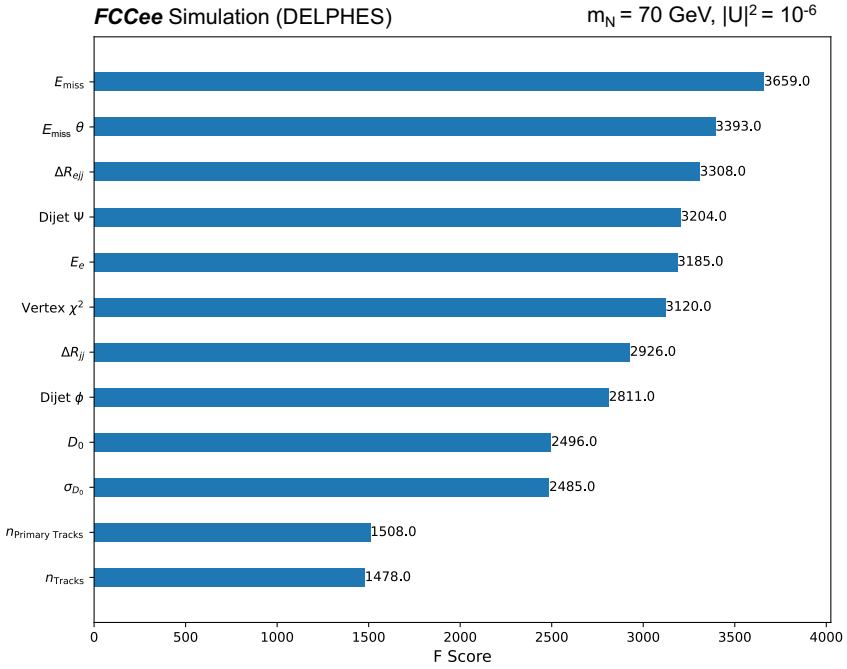
Assessing the importance of each feature input to the model is a good means of assessing how well the model has learnt, i.e. using our knowledge of each signals kinematics, we can understand if the model is sensibly using each feature. It is also useful to see if we can simplify the model any further by removing unused variables, possibly reducing the chance of overfitting and computational load of training the model. To visualise feature importance, an inbuilt method of `XGBoost` was used to plot the features of the BDT against their use by the model. This can naively be interpreted simply as how often the decision trees are split via a criteria using a given feature, or indeed how significant a given feature is to the output of the model [84].



(a) BDT feature importance for mass point 20 GeV, $|U|^2 = 10^{-6}$



(b) BDT feature importance for mass point 50 GeV, $|U|^2 = 10^{-6}$



(c) BDT feature importance for mass point 70 GeV, $|U|^2 = 10^{-6}$

Figure 49: Feature importance for three different mass points from the signal points trained where F score is a measure of how 'useful' a variable is to the model for (a) 20 GeV, $|U|^2 = 10^{-6}$, (b) 50 GeV, $|U|^2 = 10^{-6}$, (c) 70 GeV, $|U|^2 = 10^{-6}$.

These plots show that the BDT model, in each case, is using all of the input variables for the three mass points we are discussing here. Of course, for easier to identify mass point such as 10 GeV, $|U|^2 = 10^{-3}$, the model may be much simpler and require many less of the variables to have the same performance. A summary of the most important feature for each model trained in the signal parameter space is given in Figure 50.

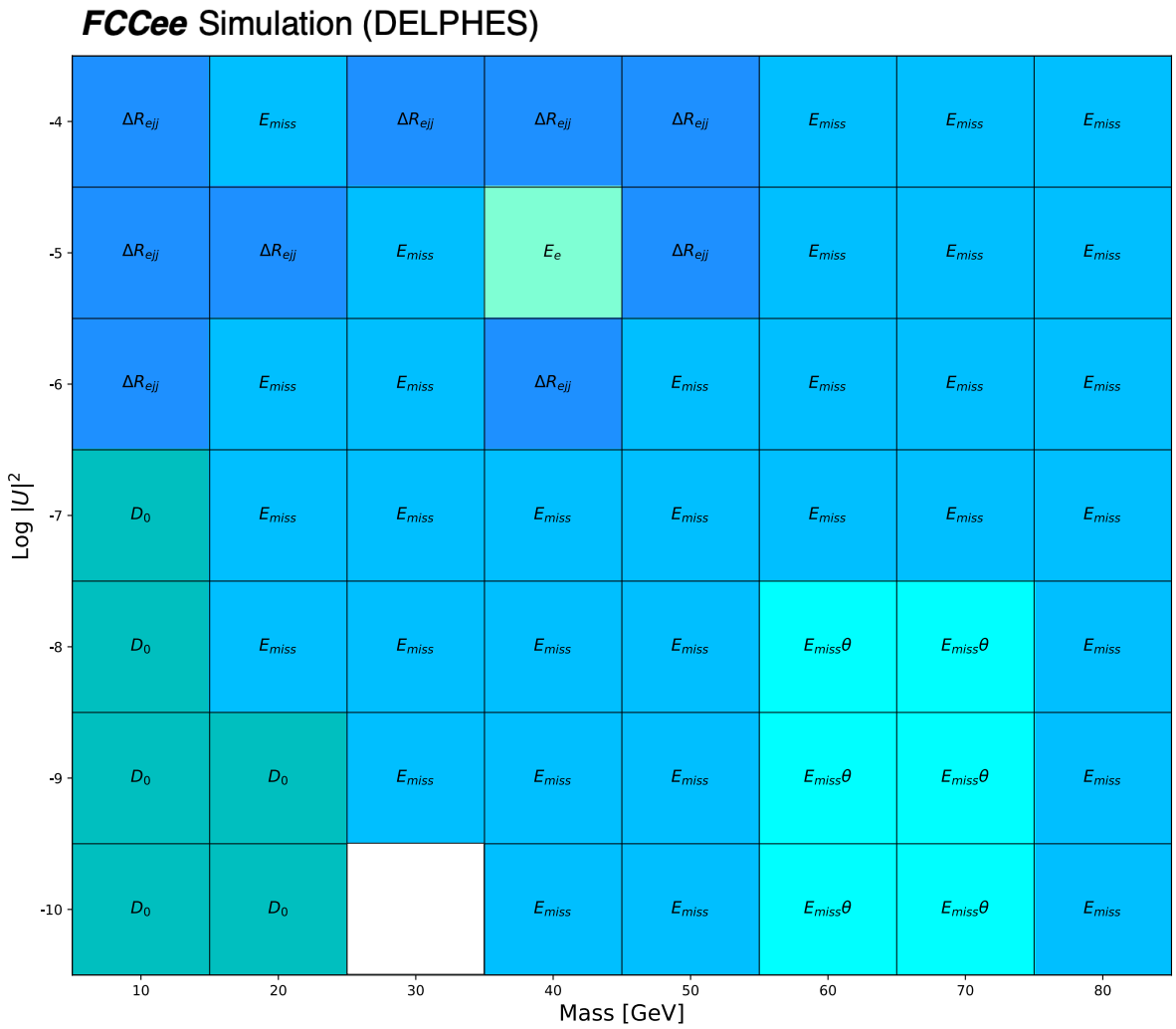


Figure 50: Grid summarising the feature with the highest 'F Score' for each signal point BDT.

In general, as seen in Figure 50, the most important feature for a BDT model trained was the reconstructed missing energy. We see in the region as we approach very prompt signatures, towards low mass and high mixing angle (since the decay length $\propto U^{-2}M^{-5}$) the transverse impact parameter becomes most important. This validates that the model is indeed utilising the most sensible variable for each mass point and naturally separating long lived decays from prompt signals.

The region in which the θ of missing energy becomes most dominant is interesting, since it is somewhat unexpected based upon the distribution of Figure 35. One possible explanation is simply that in conjunction with the other variables, the BDT is able to use variables which initially do not show much discriminating power and find patterns which we cannot trivially find via cut and count or a simple check of the distribution. For complex signal points such as those at high mass and small mixing angle, the model becomes sufficiently large in size that many of the variables are used very frequently. In

the case of the region with emphasis on θ of missing energy, the missing energy, E_{miss}, was consistently the second most leading variable by a very small margin of difference.

We might infer a more clear understanding of the region which demonstrates a reliance on the electron-dijet ΔR, since we can directly interpret the distribution of Figure 33. For the low mass HNLs, the distribution peak sits at relatively low ΔR_{ejj}, offering better discrimination compared to the high mass points for which the peak is around $\Delta R_{ejj} = \pi$, in the same place as for the three backgrounds. Hence, for the lower mass points, it is very encouraging that the BDT models use this variable to identify a difference in the kinematics between the signal and background.

After training, the model is saved as a `ROOT` file using `TMVA` [85], and applied to the unseen test data set aside during data preparation. The BDT model makes predictions of the label of each event between 0 and 1, where 0 indicates the model is certain that this is background event, and 1 indicates certainty in the event being signal. After normalising the MC events in the BDT output response to a chosen luminosity, we compute the significance and select a cut on the BDT output to preserve some region with a favourable signal to background ratio. This part of the procedure is rather similar to the essence of a cut and count study, but instead of making individual variable cuts, the result is some 'super cut' on the BDT response. An example of a decision tree is given in Figure 51 for the 10 GeV, $|U|^2 = 10^{-4}$ mass point. This point was chosen since we get very good discrimination for this point and it produces comparatively "simple" tree diagrams.

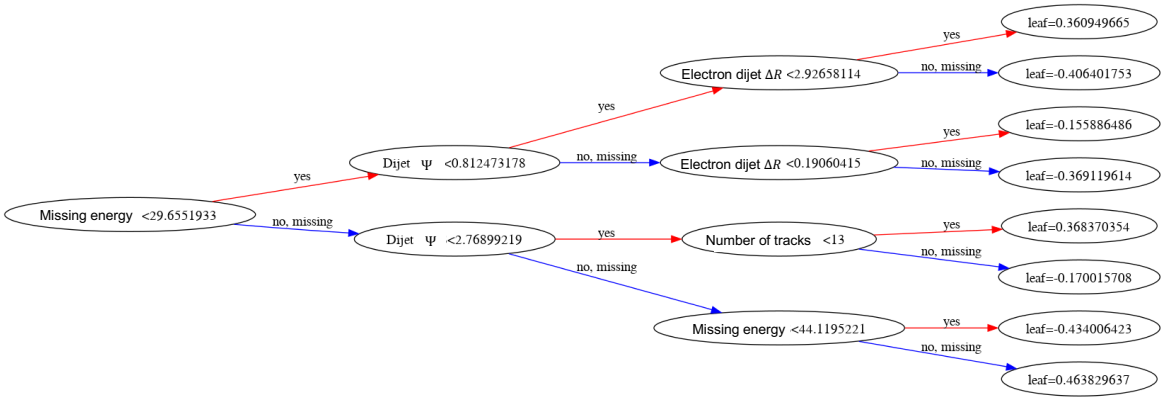


Figure 51: Example of the BDT structure coming from the training of $m_N = 10$ GeV, $|U|^2 = 10^{-4}$. The leaf outputs represent estimates from the base-learners, and in the case of binary classifcation are converted to probabilities using the sigmoid function [86].

The tree shown in Figure 51 gives some feel for the basic structure of decision trees. Indeed, the final model is built upon an ensemble of many of these trees, where each is built with efforts to reduce the errors (residuals) of the previous trees. In this context, *residuals* refer to the difference between the actual target values and the predictions made

by the model at each stage of training.

During the training process, `XGBoost` minimises a loss function by optimizing the leaf scores[16]. For classification tasks, this typically involves calculating the gradient and hessian of the loss function with respect to the current prediction. The optimal leaf score for each leaf node j is given by Equation 31.

$$w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{31}$$

where:

- I_j is the set of instances in leaf j, i.e. how many elements of the training set fall on a given leaf.
- g_i is the gradient of the loss function for instance i, which can be crudely thought of us the difference between the model prediction and the truth.
- h_i is the hessian of the loss function for instance i, and essentially measures the rate of the change of the gradient of the loss function.
- λ is the regularization term to prevent overfitting, by penalising overly complex models which learn the training set too well.

For binary classification, the leaf scores from all trees are summed to obtain the final prediction score. This score is then transformed into a probability using the logistic (sigmoid) function in Equation 32.

$$p = \frac{1}{1 + e^{-y}} \tag{32}$$

In Equation 32, y is the sum of the leaf scores from all trees for a given instance. This transformation ensures that the output is a probability value between 0 and 1, representing the likelihood of the positive class. For example, the leaf with the score ≈ 0.3609 in Figure 51 can be input to Equation 32 to yield a probability ≈ 0.5893.

In the following section, we will explore the results of our best BDT models and parameterise the output of the model in terms of sensitivity such that a 2σ delimitation can be drawn as a comparison to typical cut and count methods.

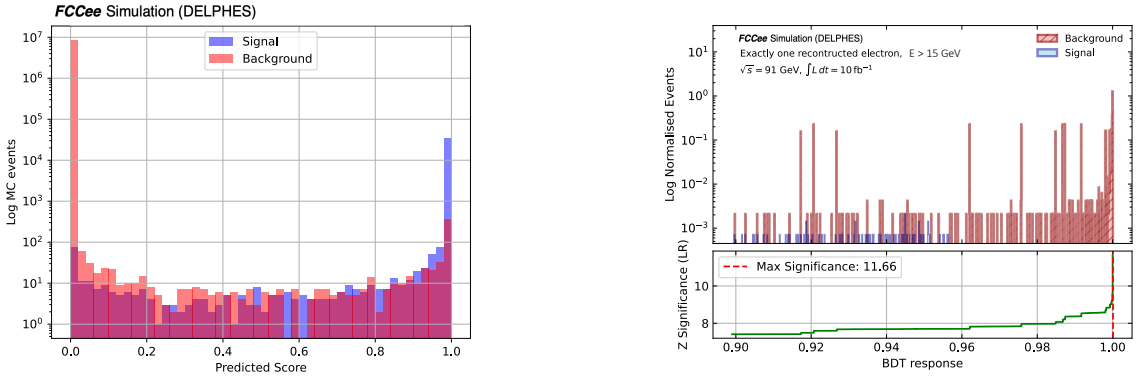
[16]For a comprehensive overview of the mathematics of boosted decision trees, References [87, 88] can be used.

5.5.3 Results

After the model is trained for each signal point on the training set, it is loaded from `TMVA` and applied to an unseen test set. The model assigns probabilities to each of the elements in the test set, and through taking a cut in the signal region on the BDT output, we can keep a high amount of signal events whilst discounting much of the background. This is principally similar to the essence of a cut and count study, but for each signal we must only make one cut rather than a sequence of 'crude' cuts on given variables.

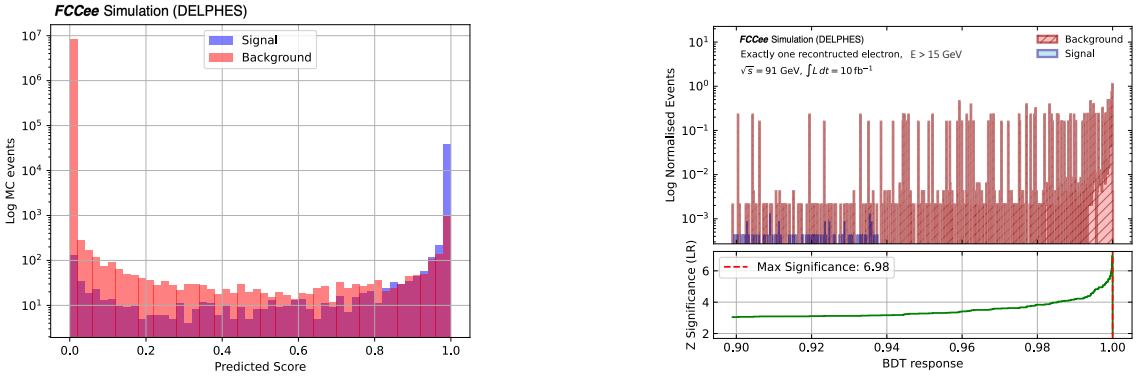
After the events in the test set are classified, with their assigned scores between 0 and 1, we can compare to the truth label and see a trend of how the model predicts the events as a benchmark for its performance. The raw performance of the BDT classifier is shown at $|U|^2 = 10^{-6}$ for the 20, 50 and 70 GeV signal points in Figures 52a, 53a and 54a respectively. It is very important to note that this information is simply an insight into the performance of the model on each of the test set points, and has not yet been normalised to any integrated luminosity, and hence cannot be used to infer much information regarding the ultimate sensitivity.

The BDT scores in Figures 52a, 53a and 54a all exhibit "healthy" behaviour for a classifier in the sense that we see a spike of the background at the truth label for these events of 0, and similarly for the signal events at 1. By re-binning the histograms in the signal region (taken between 0.9 and 1.0, containing most of the signal events), we can choose an optimal cut on the BDT output, and calculate the cumulative significance. Importantly, the significance at this stage is always calculated in the LR regime, since we always expect with the performance of our model that we yield a signal peak at 1. The BDT output, normalised to $\mathcal{L} = 10\ fb^{-1}$ is shown in the signal region for three mass points in Figures 52b, 53b and 54b below.



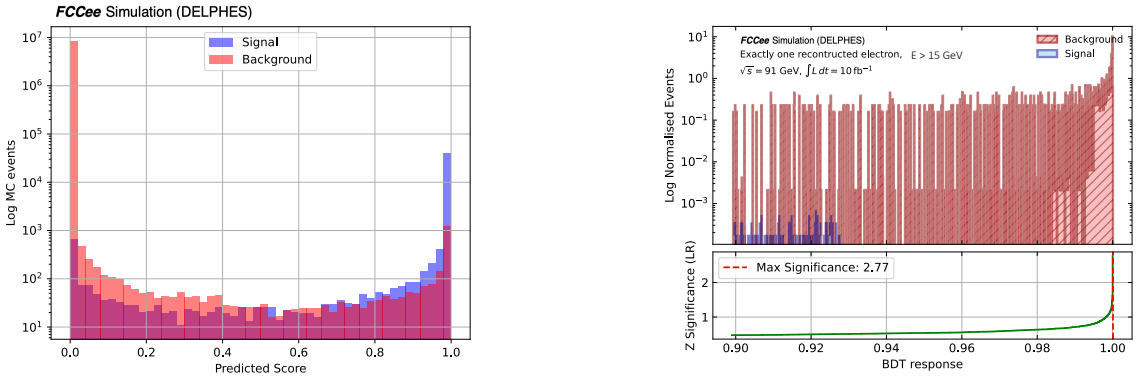
(a) (b)

Figure 52: BDT scores for mass point 20 GeV, $|U|^2 = 10^{-6}$: (a) Raw BDT classification, (b) Scaled BDT classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a BDT threshold of 0.999 is chosen, with 23.6 signal events, and 0.485 background events.



(a) (b)

Figure 53: BDT scores for mass point 50 GeV, $|U|^2 = 10^{-6}$: (a) Raw BDT classification, (b) Scaled BDT classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a BDT threshold of 0.999 is chosen, with 13.5 signal events and 1.12 background events.



(a) (b)

Figure 54: BDT scores for mass point 70 GeV, $|U|^2 = 10^{-6}$: (a) Raw BDT classification, (b) Scaled BDT classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a BDT threshold of 0.999 is chosen, with 3.29 signal events, and 0.627 background events.

Performing this analysis for each of the mass points, we can make a 2D contour plot as we did for the cut and count method. One point of consideration is that for the cut on the BDT output made, there is no requirement for the minimum yield, this means that we may yield non-integer number of events in our background events which can be seen in the Figures above. This is, of course, nonphysical and susceptible to high statistical uncertainty. Since we are normalising to a very small fraction of the expected luminosity of the FCC-ee, we continue with this approach to highlight the power of ML methods as compared to traditional methods and expect that with sufficient MC events to simulate the true luminosity, any requirements we place on the number of surviving events would not hugely effect our result. The delimiting contour plot for each of the BDT models is compared to the cut and count in Figure 55 below.

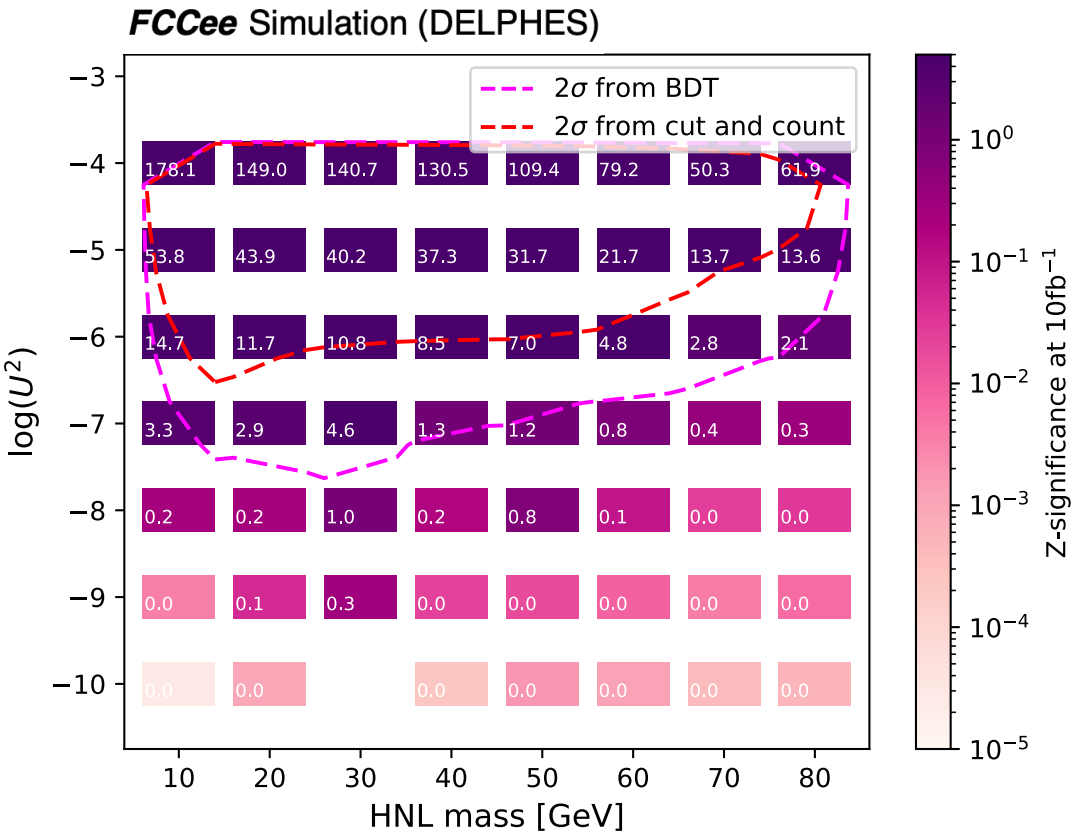


Figure 55: Sensitivity plot after BDT analysis for HNL masses between 10 and 80 GeV for mixing angles between $|U|^2 = 10^{-4}$ and 10^{-10}, normalised to an integrated luminosity of $\mathcal{L} = 10$ fb^{-1}. The magenta line delimits the 2σ contour for the BDT analysis, whilst the red does so for the cut and count study.

Figure 55 shows a clear improvement exceeding an order of magnitude in the couplings in the 95% CL limit compared to the cut and count study. This result is once again normalised to $\mathcal{L} = 10$ fb^{-1} to match the available MC statistics in the background, and scaling to the full luminosity will be done when comparison is drawn to the expected reach of the FCC-ee.

It has so far been shown that BDTs offer a powerful alternative to typical cut and count analysis techniques. In the following section, we will explore the use of deep neural networks in this domain and draw comparison between the three techniques in an attempt to maximise our sensitivity.

5.6 Deep Neural Network

Deep Neural Networks (DNNs) are a type of ML algorithm characterised by their layered architecture of interconnected neurons. These networks are capable of learning patterns within data, making them powerful tools for both supervised and unsupervised learning

tasks. Each layer in a DNN transforms the input data through weighted connections and activation functions, allowing the network to capture hierarchical features. Training a DNN involves optimising the weights using back-propagation[17] and gradient-based optimisation algorithms, such as stochastic gradient descent (SGD) or Adam, to minimise the loss function. Compared to BDTs, which are easier to interpret and robust with smaller datasets, DNNs function well with large datasets and can automatically extract features, but have many more fine tuning possibilities and can hence be very sensitive to small changes in data preparation and model architecture. A schematic of a simple DNN architecture is shown in Figure 56.

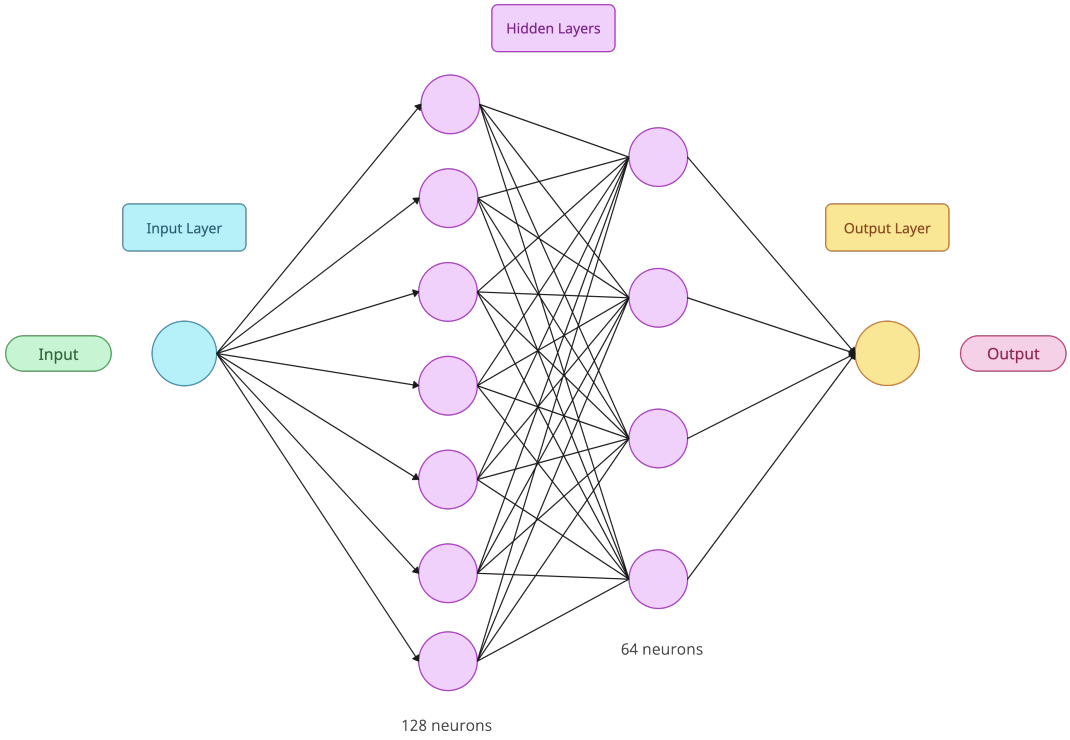


Figure 56: DNN schematic showing the model architecture used periodically throughout this analysis for testing data splits. The final model was found using `RandomSearch` to optimise the hyperparameters.

5.6.1 Data Preparation

As with the BDT, and perhaps even more so, precise data preparation is a key aspect of good DNN performance; it requires careful feature engineering, a sensible training / testing split and normalisation of the input features. The features used for the DNN are the same as those described in Table 9, with the exception of the 3D dijet angle, ψ, which was removed due to its high correlation score. Conversely to the BDT, it was found that

[17] Back-propagation is a method used in artificial neural networks to calculate the gradient of the loss function with respect to each weight by the chain rule, working backward from the output to the input [89].

removing highly correlated features improved performance. One example of a correlation matrix for the input DNN features is shown below in Figure 57.

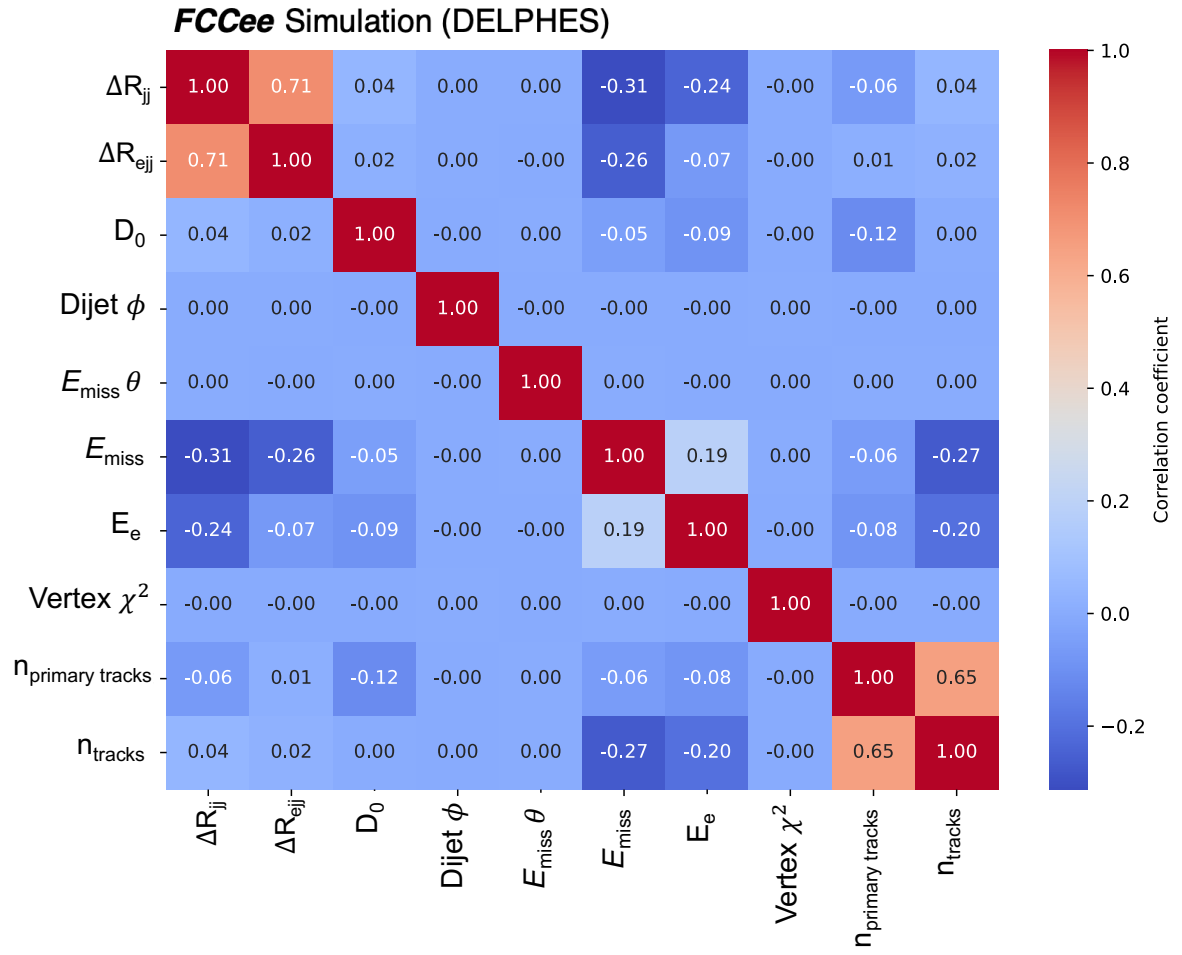


Figure 57: Correlation matrix for DNN input features for mass point 70 GeV, $|U|^2 = 10^{-6}$

Experimenting with different filters on the lead electron energy, as we did with the BDT, revealed that the performance of the model was better with a more favourable ratio of the signal and background, since the DNN was incredibly sensitive to the under-represented minority class. As a result, despite wanting to maintain a good amount of total events since this is known to generally improve DNN performance [90], we take a lead electron energy filter of 20 GeV to increase the ratio between signal and background. The efficiency of this filter is found for the three background processes and the benchmark signals in Table 11.

Process	No Filter	Lead Electron Energy > 20 GeV	Efficiency (%)
Background			
$Z \to b\bar{b}$	438,538,637	4,680,898	1.07
$Z \to c\bar{c}$	498,091,935	889,924	0.18
$Z \to$ 4body	100,000	13,376	13.38
Signal			
20 GeV, $\|U^2\| = 10^{-6}$	100,000	39,201	39.20
50 GeV, $\|U^2\| = 10^{-6}$	100,000	42,942	42.94
70 GeV, $\|U^2\| = 10^{-6}$	100,000	47,902	47.90

Table 11: BDT filter efficiencies for background and signal processes as part of the data pre-processing, where the filter conditions require the lead electron energy > 20 GeV, additionally, the number of final state reconstructed electrons is required to be exactly one.

After the filter, surviving events are concatenated into a two separate `NumPy` [91] dataframe objects, for training and testing. The chosen split between the training and testing commanded a significant amount of the total time spent working on optimising the DNN. It was found that even with oversampling techniques and class weights, the DNN structure used was incredibly sensitive to the balance of signal and background in the training and test sets.

Ultimately, investigations into the typical 80/20 split, as well as other splits such as training on exactly the same number of events as we expect in the 10 fb^{-1} conditions, lead to the conclusion that the best split was an exact 50% split between training and testing, whilst ensuring that each of the three sub background processes were split in a representative way. The statistics for the training and testing events are tabulated in Table 12.

Process	Training Events	Testing Events
Total Background	2,792,099	2,792,099
20 GeV, $\|U^2\| = 10^{-6}$	19,601	19,600
50 GeV, $\|U^2\| = 10^{-6}$	21,471	21,471
70 GeV, $\|U^2\| = 10^{-6}$	23,951	23,951

Table 12: Distribution of training and testing events for different processes used in the analysis.

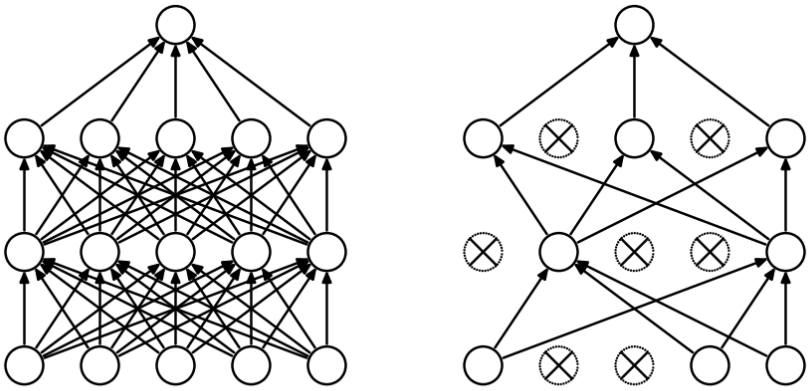
5.6.2 Training and validation

The data saved during pre-processing is then parsed to the model, which is implemented using `Keras` in `TensorFlow`. In the training of a deep neural network, the parameters of interest, usually referred to as hyperparameters, can take many forms. For our model, we use `RandomSearch` from `KerasTuner` to attempt combinations of hyperparameters over 20 trials for 100 epochs, optimising for the parameters shown in Table 13. As opposed to the grid search approach used with the BDT, we perform a random search here since there are many parameters and values they may take, and a comprehensive grid search was beyond the computational constraints of this study.

Hyperparameter	**Range**	**Step**
Units in Input Layer	32 to 512	32
Number of Hidden Layers	1 to 5	1
Units in Hidden Layers	32 to 512	32
Learning Rate	1×10^{-5} to 1×10^{-2}	Log scale
Dropout Rate	0.2	Fixed
Activation Function	ReLU	Fixed
Output Activation Function	Sigmoid	Fixed
Optimizer	Adam	Fixed
Loss Function	Binary Crossentropy	Fixed
Metrics	Accuracy, Precision, Recall, AUC	Fixed

Table 13: DNN hyperparameters and their values searched during random search.

The use of dropout inside of the model is designed to prevent overfitting, as described previously by Figure 48. By randomly dropping units (and their connections) during training, dropout helps to ensure that the model does not become overly reliant on any particular set of neurons. This process forces the network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons. As a result, the model generalizes better to unseen data. A schematic of dropout is shown in Figure 58 below.



(a) Standard Neural Net (b) After applying dropout.

Figure 58: Schematic representation of Dropout used to avoid overfitting in neural networks.

For the training of the DNN, the validation loss was monitored to further avoid overfitting. The validation set was composed of a 20% subset of the training data, meaning that at each epoch, the model could test the weights it had learnt to an unseen slice of data, and monitoring the performance on the validation set provides a less biased indication during training of how the model is truly performing. The metrics used to measure this, as seen in Table 13, include the Accuracy, Precision and Recall, which are defined in Equations 33, 34 and 35 respectively. In these equations, TP represents true positive cases, TN represents true negative, FP are false positives and FN are false negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{33}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{34}$$

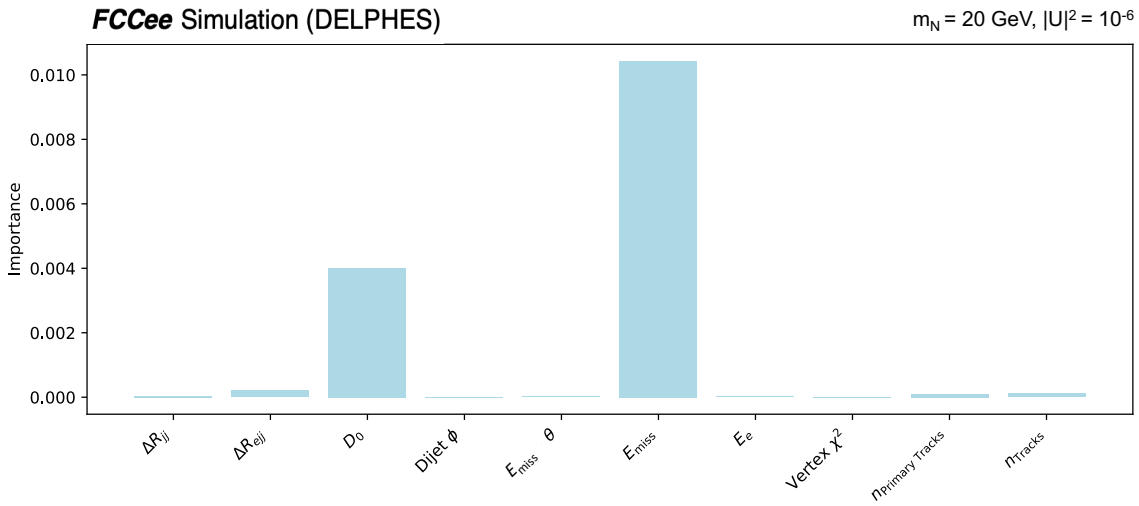
$$\text{Recall} = \frac{TP}{TP + FN} \tag{35}$$

As discussed, the ratio of signal to background for a classification task of this kind is highly important, since our minority class (in this case, the signal) is by construction necessarily underrepresented since we are searching for very low cross section signal events in many of the cases. The problem with under-representing one of the classes in a task such as this with a DNN is that the model is very likely to become biased to the majority class. On the test set, the result of this is very low recall, since the model puts high emphasis on assuming that ambiguous events are likely to fall within the majority class, as background events. Hence, the rate of FN is very high, and the rate of FP is very low so we see high

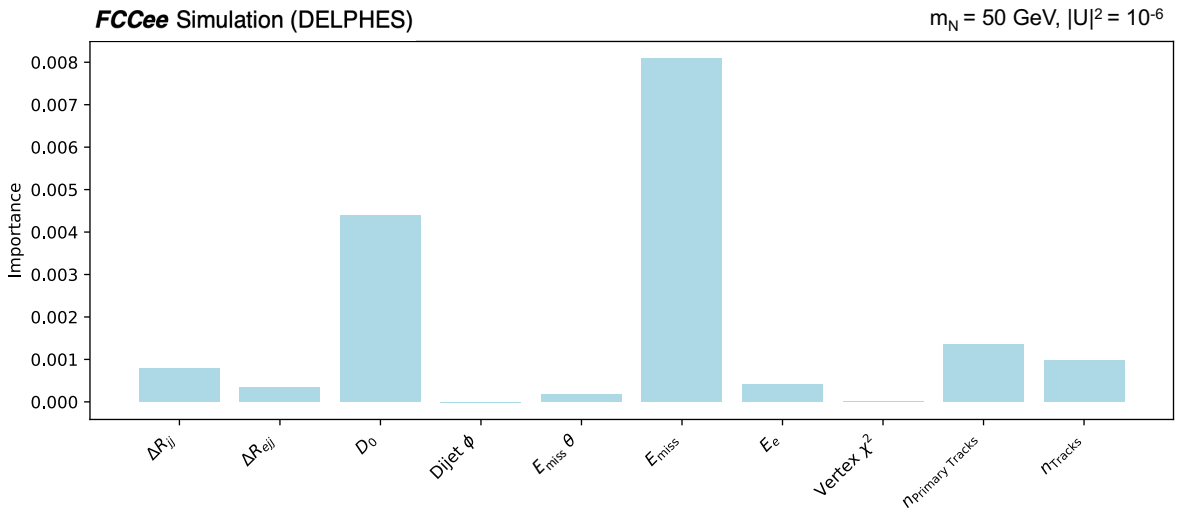
precision and accuracy and low recall. In essence, the model is not able to understand how the signal looks.

Conversely, attempts to over-sample the minority class using class weights or techniques such as `SMOTE` [92], lead to the very high recall, and correct identification of almost all signal events, but very poor precision, since the FP rate increases dramatically, with many background cases being predicted as signal. In both cases, good significance is hard to achieve with a signal region either low in signal, or high in background. Ultimately, we found that the best technique was to apply class weights to the training set that matched the cross-section weights that we normalise to in the final luminosity.

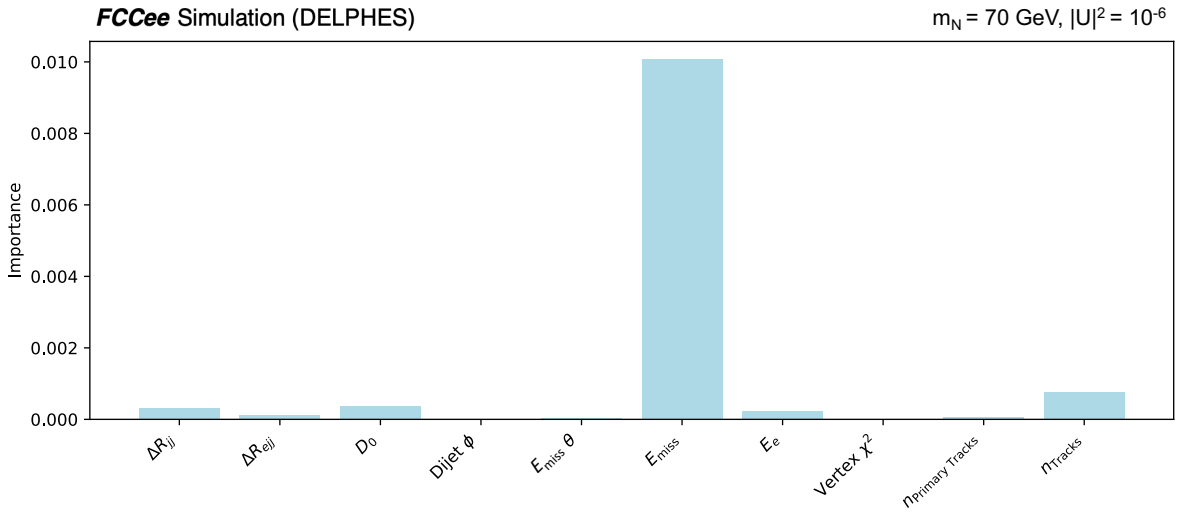
To visualize feature importance, `SHAP` [93] was used to plot the features of the DNN against their use by the model. Here, importance is measured as the *permutation importance*, which indicates how much the model's predictive power relies on each feature by observing the change in performance when the feature's values are randomly shuffled. This method provides an intuitive understanding of feature relevance by directly measuring the impact of each feature on the model's performance. Unlike the F-score for the BDT, which measures how often a feature is used to split the data, permutation importance evaluates the actual decrease in model accuracy, making a direct comparison between the two challenging. In both cases, however, we are making some measure of how important a feature is to the model in question.



(a) DNN feature importance for mass point 20 GeV, $|U|^2 = 10^{-6}$



(b) DNN feature importance for mass point 50 GeV, $|U|^2 = 10^{-6}$



(c) DNN feature importance for mass point 70 GeV, $|U|^2 = 10^{-6}$

Figure 59: DNN permutation feature importance for three different mass points from the signal points trained where F score is a measure of how 'useful' a variable is to the model for (a) 20 GeV, $|U|^2 = 10^{-6}$, (b) 50 GeV, $|U|^2 = 10^{-6}$, (c) 70 GeV, $|U|^2 = 10^{-6}$

In all three cases shown in Figure 59, the missing energy E_{miss} is found to be the most important feature. Unlike the BDT however, there is very rarely any compensation from other features such as the ΔR_{ejj}, which could be an indication that the final model used is not as optimised as the BDT. For the lower mass points, D_0 is the second most important variable, which is a clear indication that the performance of the model is sensible, even in the case where it is not fully optimised. The variable with the highest permutation feature score for each signal point is shown in Figure 60 below.

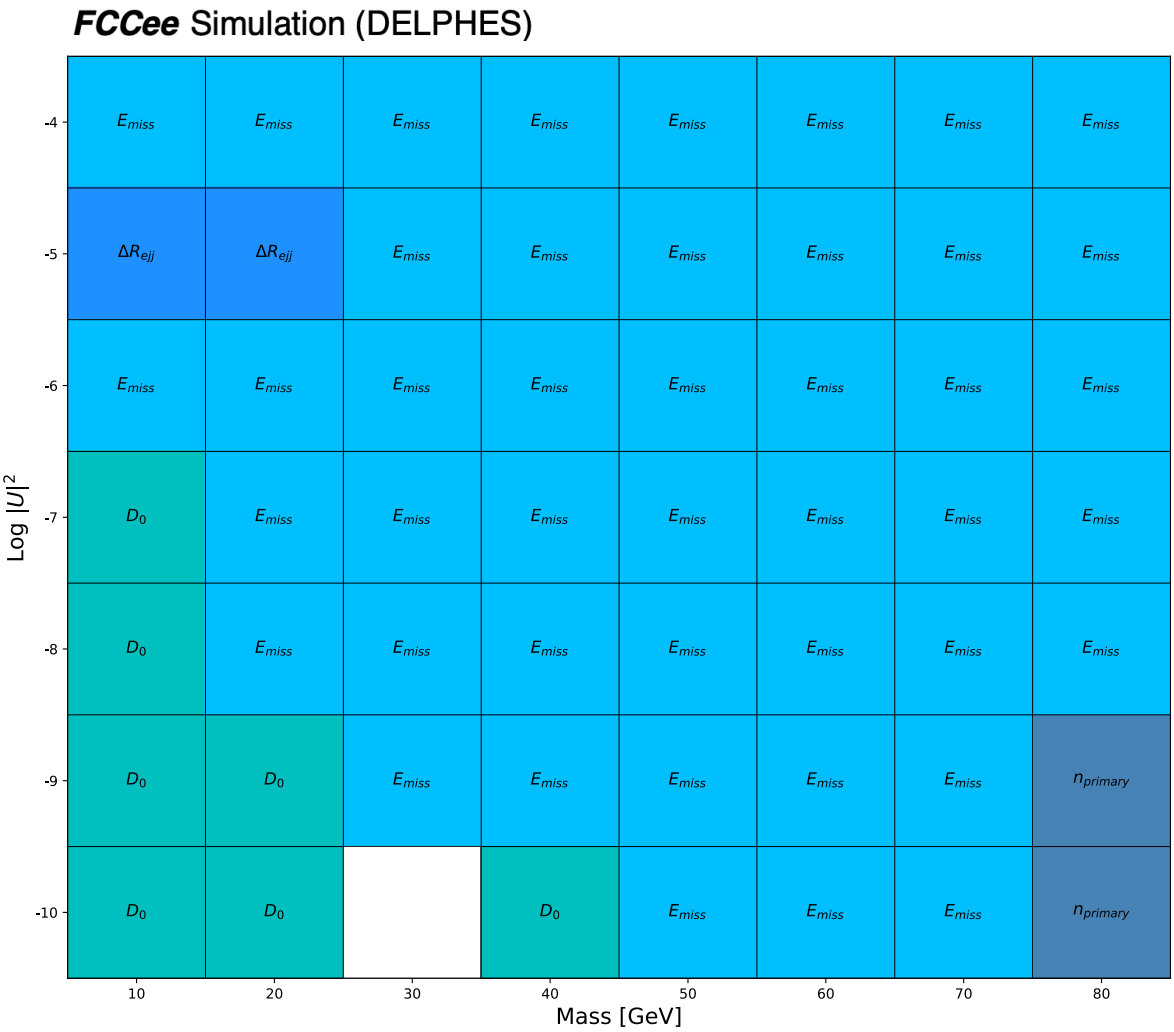


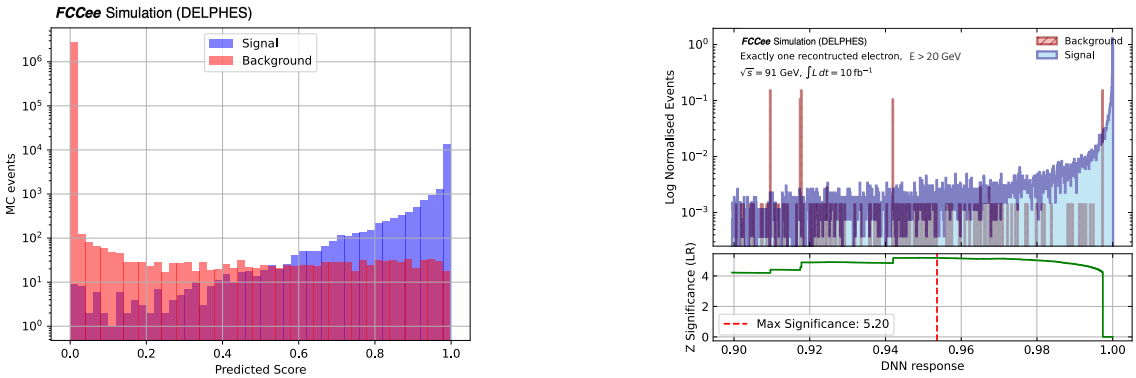
Figure 60: Matrix of the most important feature for each DNN model trained.

The interesting difference here between the BDT feature importance grid shown in Figure 50 and the DNN features in Figure 60 is the bottom right importance of the number of primary tracks, and the overall greater dominance of the missing energy. We see from our benchmark HNLs that E_{miss} is much more dominant than for the BDT feature importance, and indeed it is then expected that for the mass points on the BDT where the lead electron energy or the ΔR_{ejj} dominated slightly above the missing energy, these points are now almost all governed by the missing energy in the DNN models. We do however, still see a clear region where the prompt nature of the decay allows D_0 to take importance, which is once again a nice indication that the model uses input features accurately. Additionally, the two mass points shown where ΔR_{ejj} is most important is interesting as it coincides with the region of the BDT parameter space where this is also true. The reason, perhaps, that this trend is not as dominant as for the BDT is that for some mass points, as a result of the random search performed rather than a grid search, the model is not as optimised. For permutation feature scores, it is known that *"features*

that are deemed of low importance for a bad model (low cross-validation score) could be very important for a good model." [94]. The importance of the number of tracks for the two smallest mixing angles at 80 GeV could have a physical basis, suggesting that the number of tracks increases as the available energy of the produced jets increases. This is particularly plausible because, although these are primary tracks, this mass for the signal points exhibits the most prompt decays. Consequently, reconstructing the primary tracks from the total number of tracks becomes challenging due to the lack of an identifiable secondary vertex.

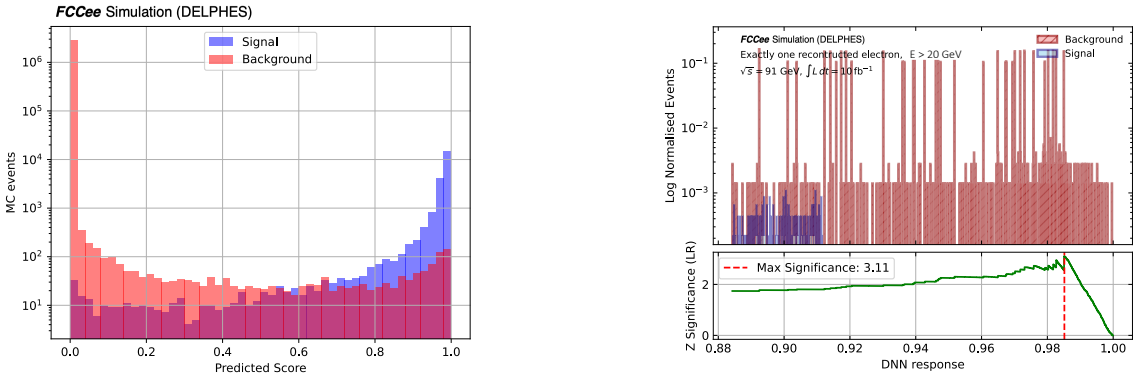
5.6.3 Results

After training, the results are obtained via applying the model to an unseen test set. Once again, the significance is calculated by making predictions on raw MC events and normalising these events relative to the number of events we expect at 10 fb^{-1}. Both the raw DNN classification, and the normalised signal region are shown in Figures 61, 62 and 63, for the 20, 50 and 70 GeV mass points at $|U|^2 = 10^{-6}$.



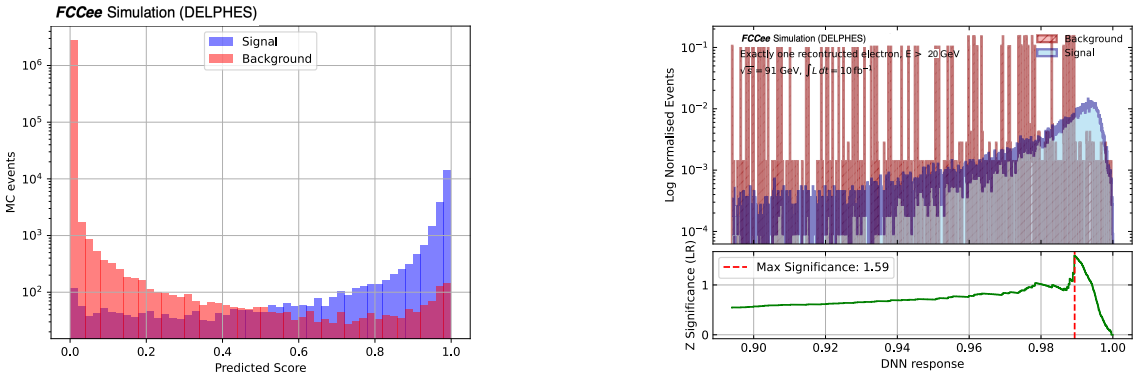
(a) (b)

Figure 61: DNN scores for mass point 20 GeV, $|U|^2 = 10^{-6}$: (a) Raw DNN classification, (b) Scaled DNN classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a DNN threshold of 0.957 is chosen, with 4.07 signal events, and 0.0014 background events.



(a) (b)

Figure 62: DNN scores for mass point 50 GeV, $|U|^2 = 10^{-6}$: (a) Raw DNN classification, (b) Scaled DNN classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a DNN threshold of 0.986 is chosen, with 3.11 signal events and 0.129 background events.



(a) (b)

Figure 63: DNN scores for mass point 70 GeV, $|U|^2 = 10^{-6}$: (a) Raw DNN classification, (b) Scaled DNN classification to $\mathcal{L} = 10\ fb^{-1}$. For this signal, a DNN threshold of 0.989 is chosen, with 1.27 signal events, and 0.423 background events.

The results in Figures 61, 62 and 63 allow us to draw a direct comparison between the BDT and DNN for each mass point. One trend is clear, that in the case of the 50 and 70 GeV raw classification, the background rejection in the signal region is fairly comparable, and even better in the case of the 70 GeV HNL, hence in some cases we see greater precision with the DNN. However, unanimously we see a worse recall around the threshold where we would wish to make the cut on the predicted scores. For example, in the case of all three benchmark signals for the BDT, we see a very identifiable peak at a BDT score of 1, meaning we can make a very rigid cut and allow for high sensitivity. On the other hand, even with good raw classification, the DNN systematically fails to group sufficiently many of the signal events into a unifed peak, meaning that we must take looser cuts and hence accept a higher amount of background than for the BDT,

yielding lower overall significance.

Indeed, since we are using the cross section weights to train the model, this could be an indication that for the heavier mass points, where the cross section is lower, the precision is increasing and the recall is decreasing. This is because, for small cross section signals, we are essentially instructing the loss function to be much more penalised for misclassifying the background than the signal. Rather, this performance is inverted for the 20 GeV mass point seen in Figure 61, since it has the highest of the three cross sections in the figures shown, and the background is misclassified more intensely whilst a less smeared spike exists at an output of 1 for the signal.

Summarising the sensitivity across all signal points in the parameter space, we can interpolate the 2σ delimitation and overlay it with the highest performing classifier scores, in this case the BDT. This allows us to see the three distinct lines: the cut and count (red), the BDTs (magenta) and the DNNs (blue) and make direct comparison between the three channels. This can be seen in Figure 64 below.

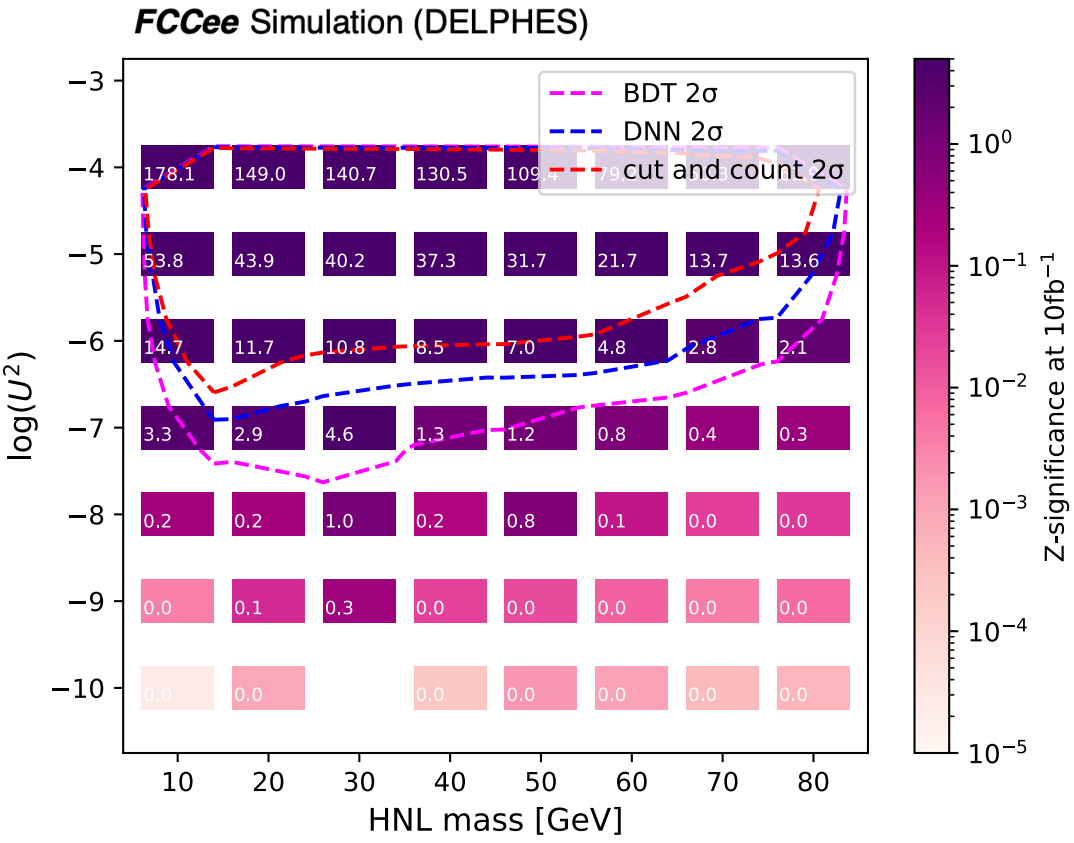


Figure 64: Sensitivity plot after BDT analysis for HNL masses between 10 and 80 GeV for mixing angles between $|U|^2 = 10^{-4}$ and 10^{-10}, normalised to an integrated luminosity of $\mathcal{L} = 10$ fb^{-1}. The magenta line delimits the 2σ contour for the BDT analysis, whilst the red does so for the cut and count study. The line in blue shows the final result for the DNN.

In Figure 64 we see an improvement of the DNN when compared to the cut and count, but see that the BDT yields overall a better performance than the DNN by almost an order of magnitude in the couplings at the points of greatest difference.

Importantly, we have shown that in both cases, ML is more effective than cut and count methods for maximising the sensitivity in an analysis such as this. As to exploring the difference between the BDT and DNN performance, we must do so with a fair amount of speculation since many would characterise the DNN to be something of a "*black box*", and where we can easily understand the tree structure of the BDT, the process of the DNN learning is much more abstract. One potential reason for the BDT outperforming the DNN in this specific task could be the nature of the dataset. BDTs tend to handle imbalanced datasets more effectively than DNNs, particularly in cases where the signal-to-background ratio is extremely low, as is the case here, shown in Table 4.

BDTs are inherently robust to class imbalance due to their iterative boosting process, which places more emphasis on correctly classifying the minority class after each iteration. This allows BDTs to focus on hard-to-classify instances in the minority class, potentially leading to better performance in scenarios with a low signal yield [95, 96, 97]. On the other hand, DNNs often require extensive tuning of hyperparameters, data augmentation, and sophisticated balancing techniques to achieve similar performance in such imbalanced settings.

It is reported elsewhere [98, 99] that for *tabular* data, such as the data used in this study, decision trees can vastly outperform neural networks. In [98], the empirical study conducted stipulates that a key reason is related to the irregular nature of tabular data, and a decision tree's robustness in ignoring uninformative features compared to a neural network. The conclusion the authors draw may be applied here, where we found that the emphasis on DNN optimisation was heavily lead by experimenting with different data preparation and feature engineering techniques, as opposed to simply the model architecture.

5.7 Long Lived vs Prompt HNLs

The last section of analysis conducted as part of this thesis research was done to exploit the topological differences in long lived and prompt signatures to try to maximise the sensitivity. The lifetime of the particles, in this case parameterised by D_0 or any such variant of D_0, such as the D_0 significance. In our parameter space, there is a wide range of values that the decay length can take, as we can see below in Figure 65, which shows the mean values of the D_0 [mm], and highlights the relationship between the mass, the coupling and the particle lifetime.

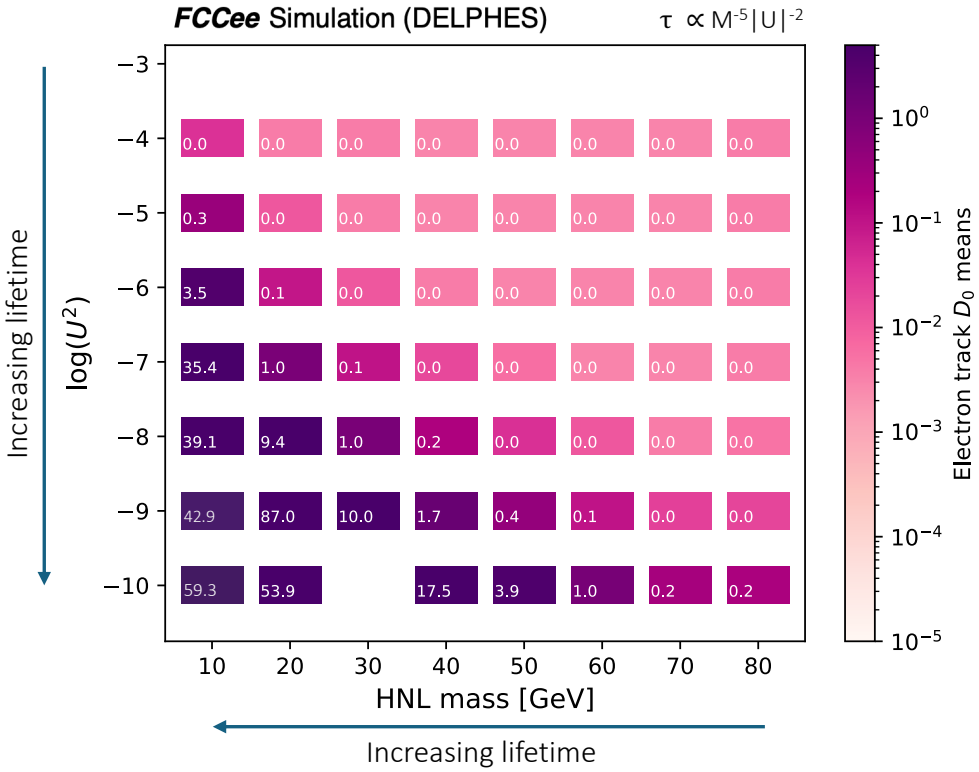


Figure 65: A grid of the mean value of D_0 [mm] for each HNL signal point in the parameter space considered.

Using our most effective classifier, the BDT, we included the D_0 significance parameter, σ_{D_0}, inside the pre-process filter, in order to try to produce trainings specifically for long lived or for prompt decays. The σ_{D_0} was chosen for the filter as opposed to simply D_0 since the distributions posited a higher discriminating power. Both were tried, and it was found that σ_{D_0} was easier to find a threshold for the filter and ultimately therefore yielded a better result. The mean value of σ_{D_0} for each signal point in the parameter space can be found in Figure 66 below.

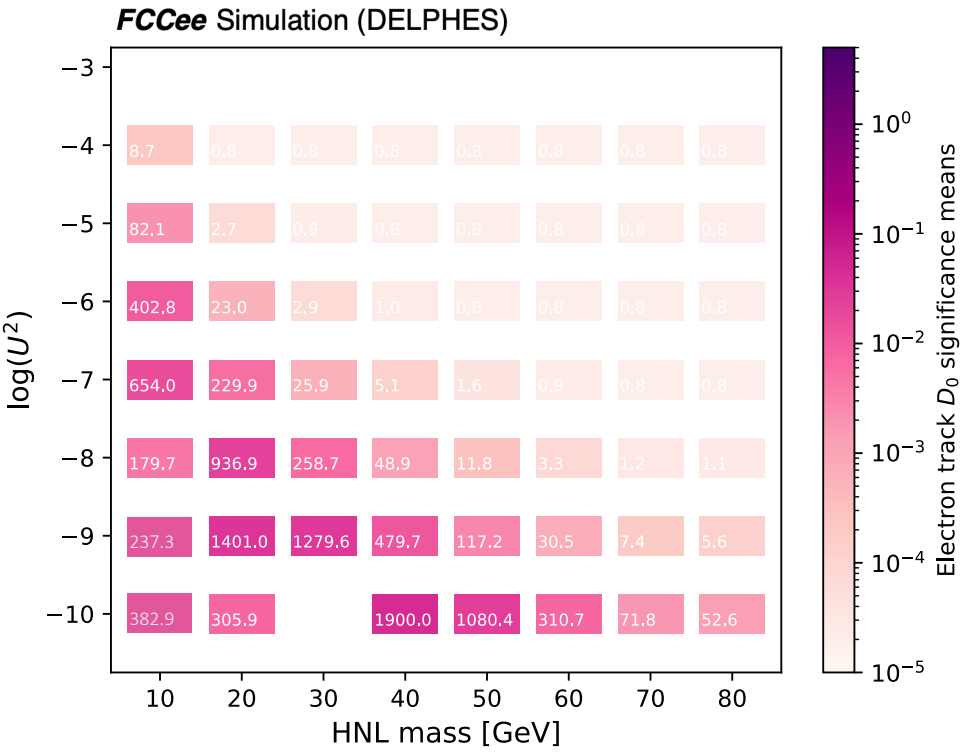


Figure 66: A grid of the mean value of σ_{D_0} for each HNL signal point in the parameter space considered.

The idea behind this was such that, although a unique BDT is trained for each mass point, by feeding the model a region richer in signal relative to the background, we may help the BDT to better optimise the classification between signal and background.

As a result a combined filter of lead electron energy > 15 GeV and $\sigma_{d_0} < 5$ was used for promptly decaying HNLs, and $\sigma_{d_0} > 5$ was used LLP signatures. This filter was such that, in both cases, specifically in the case of LLP signatures, many of the signal points had a yield of zero. This means that the filter for LLPs allowed us to only train models in the region with larger σ_{d_0}, and vice versa for the prompt decay filter.

A summary of the results using the BDT model described in Section 5.5 is found in Figure 67. In this plot, the blue contour represents long lived particles at low mass and low mixing angles

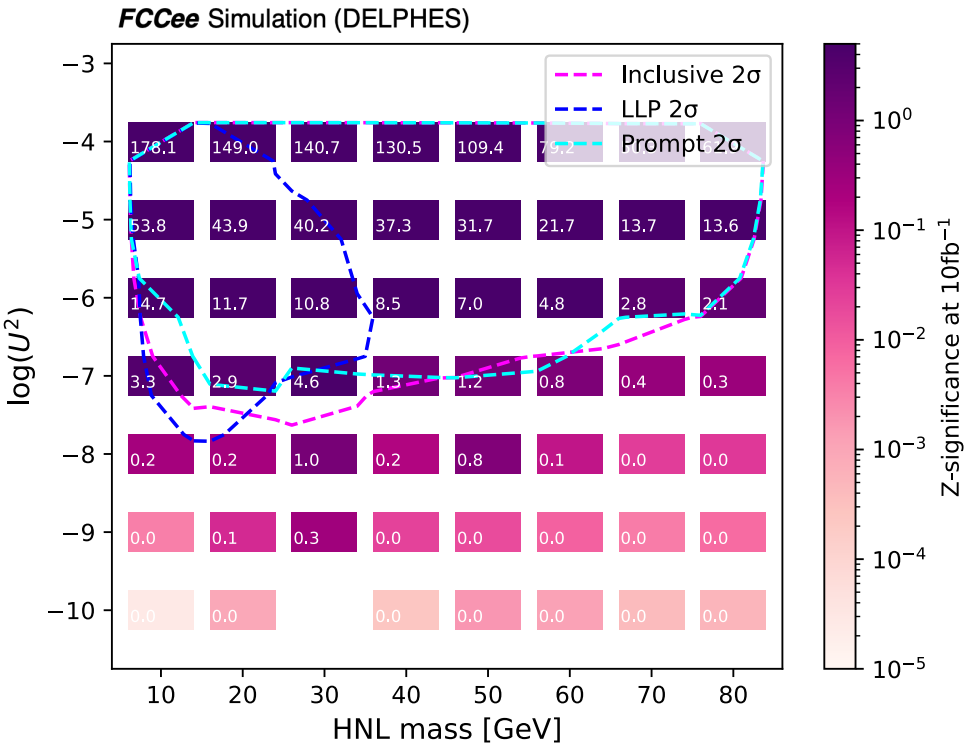


Figure 67: Sensitivity plot after BDT analysis for HNL masses between 10 and 80 GeV for mixing angles between $|U|^2 = 10^{-4}$ and 10^{-10}. This plot is normalised to an integrated luminosity of $\mathcal{L} = 10$ fb^{-1}. The magenta, blue and cyan dashed lines show the extrapolation of 2σ significance respectively for: the BDT with filter stipulating lead electron energy > 15 GeV, the long lived particles with lead electron energy > 15 GeV and $\sigma_{d_0} > 5$, and the prompt decay with lead electron energy > 15 GeV and $\sigma_{d_0} < 5$.

In Figure 67, we see when using the filter requiring $\sigma_{d_0} > 5$ to isolate the LLPs (blue), we have a slight increase in the performance of the BDT with just the lead electron energy (magenta), though this improvement is very limited. We might use this result in conjunction with the feature importance grid in Figure 50 to see that, at the bottom of this contour is the region where D_0 already becomes the most important parameter to the model, an indication that the model is naturally able to identify the strength in using this parameter to separate the long lived HNL from the background. In addition, the prompt decay contour (cyan) also yields a similar shape in the prompt region to that of the BDT without the D_0 filter. While this investigation is valuable and could be expanded further, it appears that when employing machine learning methods and training a unique model for each signal point, the distinction between LLPs and prompt decaying signatures is inherently addressed, provided the model has sufficiently good predictive power.

6 Results and Discussion

6.1 Sensitivity study

The aim of this thesis was to improve upon the existing 95% CL limit in the HNL parameter space for masses 10 GeV $\leq m_N \leq$ 80 GeV with mixing angles $10^{-4} \leq |V_{eN}|^2 \leq 10^{-10}$. An overall improvement on the parameter space for the electron dijet final state has been found to exceed 2 orders of magnitude in the coupling relative to the cut and count study.

To contextualise our findings, the 95% CL limit in the HNL parameter space is plotted in Figure 68 for the latest result published for the CMS Experiment at CERN [38], as well as the expected reach of the FCC-ee to all HNL decay modes [41]. The theoretical limit of the Type I seesaw model [100] is also included in the figure.

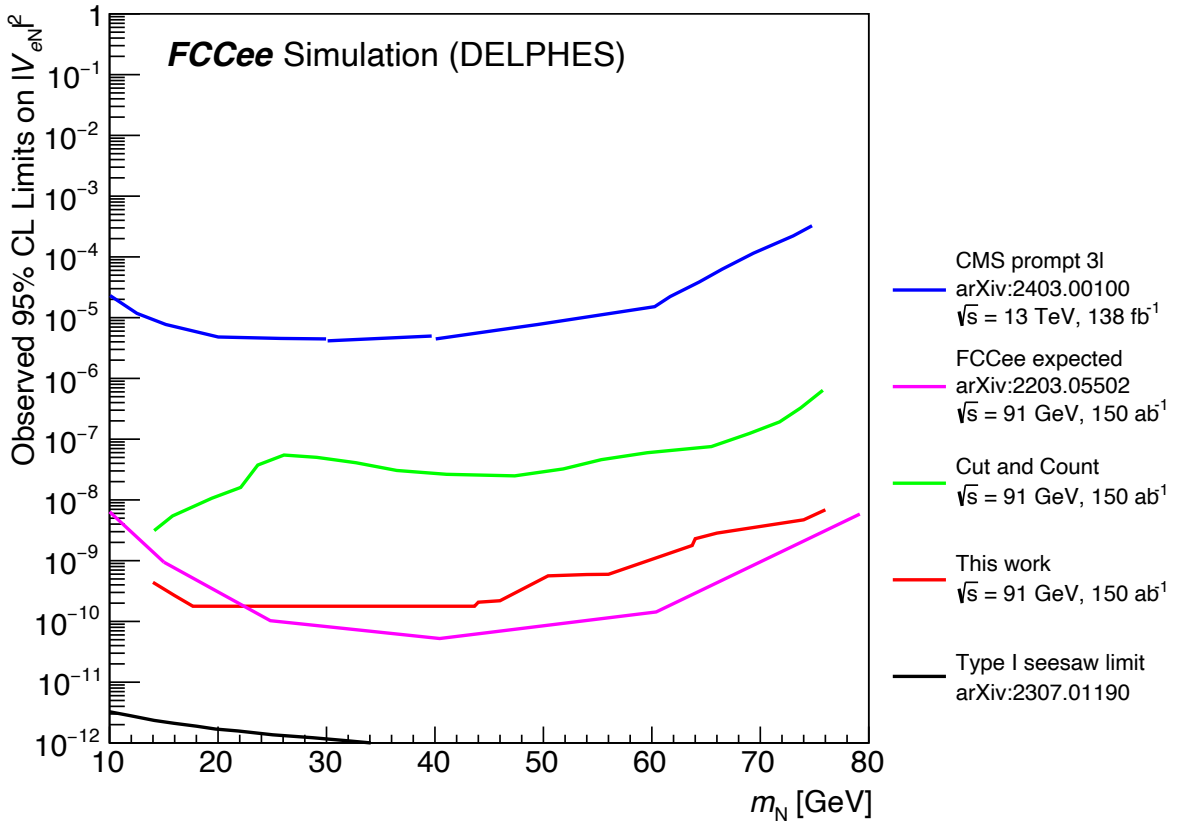


Figure 68: Observed 95% CL Limits compared for this thesis and for the latest HNL search at CMS, the full reach of the FCC-ee to HNLs and the Type I seesaw model limit.

Figure 68 summarises our results scaled to the full expected luminosity of 150 ab^{-1}, and contextualises them with existing work in this region of the HNL parameter space. Whilst this figure works well to show the power of the ML techniques used, specifically with the BDTs, care should be taken to interpret them when quoted without statistical

uncertainties. Indeed, as previously mentioned, when scaling to the full luminosity, the scale factors involved as a result of insufficient MC statistics make accurate commentary very difficult, and this should simply be used as an indication of the approximate region of the parameter space probed.

6.2 Outlook

This thesis demonstrates an example of using ML techniques in low signal yield scenarios to identify new physics in future particle physics experiments. We have shown its relative power compared to typical cut and count studies, and drawn comparison between BDT and DNN techniques for this genre of classification task.

Whilst the ML models defined here are powerful, optimisation of these models could yield even stronger results, given adequate time to continue fine tuning of the hyper-parameters. Further, a key issue for this study is found in gaining reliable insight into the full luminosity of the FCC-ee machine due to the lack of statistics available for the main background processes. It is imperative to continue this line of work by investigating filters at the truth level which allow us to have more background statistics in the signal region. Additionally, further work could be done on the research laid out in this thesis to properly assess and parameterise the systematic and statistical uncertainties to be able to truly understand the final result at 150 ab^{-1}.

One point of consideration for further study may also include the use of similar style BDT and DNN models to look for kinematic differences between Dirac and Majorana HNLs. Work on this type of discrimination can be found in [72], and ML techniques may prove to yield even stronger results.

Lastly, combining the limits of the various HNL final states to find an improved limit for the total sensitivity of the FCC-ee to all HNL final states is now highly important as we move closer to the feasibility study submission for the FCC, in order to continue building upon an increasingly convincing physics case for the machine to be built. Work in the other HNL final states can be found: in the $ee\nu$ final state [75], in the μjj final state [101] and in the $\mu\mu\nu$ channel [102].

7 Conclusion

In this thesis, we have investigated ML-based techniques to increase the projected sensitivity of HNL discovery in the electron dijet final state at the FCC-ee. We have demonstrated the power of these methods over conventional cut-and-count procedures by reproducing our own cut-and-count analysis in alignment with existing studies in this HNL final state, and optimizing this result with BDTs and DNNs. Further attempts to optimize were trialed by separating signatures into prompt and long-lived HNL decays.

It has been shown that a BDT trained for each signal mass point is capable of outperforming a DNN. However, any conclusion to this end is unlikely to be robust, as there are vastly many ways one could further optimize the DNN models presented here. We speculate that the observed performance discrepancy could be due to the nature of the BDT's ability to handle the highly imbalanced datasets typical in HNL searches, where the signal is exceedingly rare compared to the background. BDTs, with their ensemble learning approach, might be better suited to picking out the rare signal events from a large background.

Future work should explore advanced techniques such as hyperparameter optimization as well as careful feature engineering to fully exploit the potential of DNN use in this domain. Additionally, increasing statistics in the background processes, especially in signal-rich regions, is vital to make robust claims on the discovery potential of HNLs at the FCC-ee's true luminosity. A full detector simulation is also required to ensure careful consideration of the efficiencies and resolutions of the IDEA design as it becomes more concrete.

We make the case in this thesis that, beyond its self-evident claim of being a precision machine, the FCC-ee presents a unique opportunity to hugely improve our limits in the parameter space of the HNLs, with many orders of magnitude better delimitation in the HNL couplings than limits set by studies at CMS or beyond. It will be of immense interest to look forward over the next few decades to the completion of HL-LHC, the submission of the feasibility study of the FCC machine, and the landscape of particle physics in general. It is a deeply exciting time to be involved with the community, and it is my hope that this thesis serves the purpose of encouraging further studies of this nature by exploring the boundaries of existing limits in this space.

References

[1] A. Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, pp. 1–29, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037026931200857X

[2] C. Collaboration, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Physics Letters B*, vol. 716, pp. 30–61, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269312008581

[3] "The higgs at last," *SA Special Editions*, vol. 22, no. 2s, p. 4, May 2013.

[4] B. Bockelman, P. Elmer, and G. Watts, "Iris-hep strategic plan for the next phase of software upgrades for hl-lhc physics," 2023.

[5] D. Hanneke, S. Fogwell, and G. Gabrielse, "New measurement of the electron magnetic moment and the fine structure constant," *Physical Review Letters*, vol. 100, no. 12, p. 120801, 2008. [Online]. Available: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.100.120801

[6] Y. F. et al. (Super-Kamiokande Collaboration), "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, no. 8, pp. 1562–1567, 1998.

[7] Q. A. et al. (SNO Collaboration), "Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}b solar neutrinos at the sudbury neutrino observatory," *Physical Review Letters*, vol. 87, no. 7, p. 071301, 2001.

[8] K. E. et al. (KamLAND Collaboration), "First results from kamland: Evidence for reactor antineutrino disappearance," *Physical Review Letters*, vol. 90, no. 2, p. 021802, 2003.

[9] P. A. Z. et al. (Particle Data Group), "Status and perspectives of neutrino physics," *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 8, p. 083C01, 2020.

[10] MissMJ, "Standard model of elementary particles," 2019, licensed under Creative Commons Attribution 3.0 Unported license. [Online]. Available: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg

[11] C. Quigg, "The electroweak theory," 2002.

[12] Y. Fukuda and others (Super-Kamiokande Collaboration), "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998.

[13] Q. R. Ahmad and others (SNO Collaboration), "Measurement of the rate of νe + d

→ p + p + e- interactions produced by 8b solar neutrinos at the sudbury neutrino observatory," *Phys. Rev. Lett.*, vol. 87, p. 071301, 2001.

[14] T. Araki and others (KamLAND Collaboration), "First results from kamland: Evidence for reactor antineutrino disappearance," *Phys. Rev. Lett.*, vol. 90, p. 021802, 2003.

[15] P. F. de Salas, S. Gariazzo, O. Mena, C. A. Ternes, and M. Tórtola, "Neutrino mass ordering from oscillations and beyond: 2018 status and future prospects," *Frontiers in Astronomy and Space Sciences*, vol. 5, oct 2018. [Online]. Available: http://dx.doi.org/10.3389/fspas.2018.00036

[16] A. Loureiro, A. Cuceu, F. B. Abdalla, B. Moraes, L. Whiteway, M. McLeod, S. T. Balan, O. Lahav, A. Benoit-Lévy, M. Manera, R. P. Rollins, and H. S. Xavier, "Upper bound of neutrino masses from combined cosmological observations and particle physics experiments," *Physical Review Letters*, vol. 123, no. 8, aug 2019. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.123.081301

[17] V. Brdar, A. J. Helmboldt, S. Iwamoto, and K. Schmitz, "Type i seesaw mechanism as the common origin of neutrino mass, baryon asymmetry, and the electroweak scale," *Phys. Rev. D*, vol. 100, p. 075029, Oct 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.100.075029

[18] B. J. P. Jones, "The physics of neutrinoless double beta decay: A primer," 2022.

[19] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.

[20] P. Minkowski, "$\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?" *Phys. Lett. B*, vol. 67, pp. 421–428, 1977.

[21] T. Yanagida, "Horizontal symmetry and masses of neutrinos," in *Proceedings of the Workshop on the Unified Theory and the Baryon Number in the Universe*, 1979, pp. 95–99, conf.Proc. C7902131.

[22] M. Gell-Mann, P. Ramond, and R. Slansky, "Complex spinors and unified theories," in *Supergravity Workshop at Stony Brook*, 1979, pp. 315–321, conf.Proc. C790927.

[23] R. N. Mohapatra and G. Senjanovic, "Neutrino mass and spontaneous parity violation," *Phys. Rev. Lett.*, vol. 44, p. 912, 1980.

[24] J. Schechter and J. W. F. Valle, *Phys. Rev. D*, vol. 22, p. 2227, 1980.

[25] M. Magg and C. Wetterich, *Phys. Lett. B*, vol. 94, p. 61, 1980.

[26] R. Foot, H. Lew, X. G. He, and G. C. Joshi, *Z. Phys. C*, vol. 44, p. 441, 1989.

[27] E. Ma, *Phys. Rev. Lett.*, vol. 81, p. 1171, 1998.

[28] A. M. Abdullahi and e. a. Barham Alzás, "The present and future status of heavy neutral leptons," *Journal of Physics G: Nuclear and Particle Physics*, vol. 50, no. 2, p. 020501, jan 2023. [Online]. Available: http://dx.doi.org/10.1088/1361-6471/ac98f9

[29] A. Blondel, E. Graverini, N. Serra, and M. Shaposhnikov, "Search for heavy right handed neutrinos at the fcc-ee," *Nuclear and Particle Physics Proceedings*, vol. 273–275, pp. 1883–1890, April–June 2016.

[30] T. Asaka and M. Shaposhnikov, "The νmsm, dark matter and baryon asymmetry of the universe," *Physics Letters B*, vol. 620, pp. 17–26, 2005.

[31] A. Merle and M. Totzauer, "kev neutrino model building," *Annals of Physics*, vol. 360, pp. 419–453, 2015.

[32] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall, "Detection of an unidentified emission line in the stacked x-ray spectrum of galaxy clusters," *The Astrophysical Journal*, vol. 789, no. 1, p. 13, jun 2014. [Online]. Available: http://dx.doi.org/10.1088/0004-637X/789/1/13

[33] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, and J. Franse, "Unidentified line in x-ray spectra of the andromeda galaxy and perseus galaxy cluster," *Physical Review Letters*, vol. 113, no. 25, dec 2014. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.113.251301

[34] E. Fernández-Martínez, M. González-López, J. Hernández-García, M. Hostert, and J. López-Pavón, "Effective portals to heavy neutral leptons," *Journal of High Energy Physics*, vol. 2023, no. 9, sep 2023. [Online]. Available: http://dx.doi.org/10.1007/JHEP09(2023)001

[35] D. Collaboration, "Search for neutral heavy leptons produced in z decays," *Z. Phys. C*, vol. 74, p. 57, 1997, [Erratum: doi:10.1007/BF03546181].

[36] C. Collaboration, "Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{(s)} = 13$ tev," *JHEP*, vol. 07, p. 081, 2022.

[37] ——, "Search for long-lived heavy neutral leptons with lepton flavour conserving or violating decays to a jet and a charged lepton," 2023, submitted to JHEP.

[38] "Search for heavy neutral leptons in final states with electrons, muons, and hadronically decaying tau leptons in proton-proton collisions at $\sqrt{s} = 13$ tev," CERN, Geneva, Tech. Rep., 2024, submitted to the Journal of High Energy

Physics. All figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/EXO-22-011 (CMS Public Pages). [Online]. Available: https://cds.cern.ch/record/2890510

[39] E. S. for Particle Physics Preparatory Group, "European strategy for particle physics - briefing book," 2019.

[40] S. Antusch, O. Fischer, and J. Rubio, "Heavy neutrino-antineutrino oscillations," *JHEP*, vol. 10, p. 094, 2016.

[41] A. Blondel, C. Verhaaren, J. Alimena, M. Bauer, P. Azzi, R. Ruiz, M. Neubert, O. Mikulenko, M. Ovchynnikov, M. Drewes, J. Klaric, C. Rizzi, A. Sfyrla, T. Sharma, S. Kulkarni, A. Thamm, R. Gonzalez Suarez, and L. Rygaard, "Searches for long-lived particles at the future fcc-ee," 2022, contribution to Snowmass 2021. [Online]. Available: https://cds.cern.ch/record/2805010

[42] M. Benedikt, F. Zimmermann, and CERN, "Fcc feasibility study status, 7th fcc physics workshop," Presented at the 7th FCC Physics Workshop, Annecy, January 2024, on behalf of FCC collaboration & FCCIS DS team. [Online]. Available: https://indico.cern.ch/event/1307378/contributions/5726327/attachments/2788949/4863152/240129-FCC-FS-Status.pdf

[43] A. Blondel and P. Janot, "Fcc-ee overview: new opportunities create new challenges," *Eur. Phys. J. Plus*, vol. 137, p. 92, 2022. [Online]. Available: https://doi.org/10.1140/epjp/s13360-021-02154-9

[44] C. Helsens and G. Ganis, "Offline computing resources for fcc-ee and related challenges," *The European Physical Journal Plus*, vol. 137, 01 2022.

[45] S. Knapen and A. Thamm, "Direct discovery of new light states at the fccee," *Eur. Phys. J. Plus*, vol. 136, p. 936, 2021. [Online]. Available: https://doi.org/10.1140/epjp/s13360-021-01874-2

[46] A. Abada, M. Abbrescia, S. AbdusSalam *et al.*, "Fcc-ee: The lepton collider," *Eur. Phys. J. Spec. Top.*, vol. 228, pp. 261–623, 2019, received 20 December 2018; Published 04 June 2019; Issue Date June 2019. [Online]. Available: https://doi.org/10.1140/epjst/e2019-900045-4

[47] N. Bacchetta, J. J. Blaising, E. Brondolin, M. Dam, D. Dannheim, K. Elsener, D. Hynds, P. Janot, A. M. Kolano, E. Leogrande, L. Linssen, A. Nürnberg, E. F. Perez, M. Petrič, P. Roloff, A. Sailer, N. Siegrist, O. Viazlo, G. G. Voutsinas, and M. A. Weber, "Cld – a detector concept for the fcc-ee," 2019.

[48] M. Antonello, "Idea: A detector concept for future leptonic colliders," *Il Nuovo*

Cimento, vol. 43C, p. 27, 2020, colloquia: IFAE 2019. On behalf of the RD-FA Collaboration, INFN, Sezione di Milano - Milano, Italy.

[49] F. Collaboration, "Noble liquid calorimetry for fcc: Allegro detector," 2023, presented at the FCC Week 2023. [Online]. Available: https://indico.in2p3.fr/event/25348/contributions/95702/attachments/62395/85273/ALLEGRO.pdf

[50] W. Elmetenawee, G. Chiarello, A. Corvaglia, F. Cuna, N. De Filippis, E. Gorini, F. Grancagnolo, M. Maggi, A. Miccoli, M. Panareo, M. Primavera, G. Tassielli, and A. Ventura, "The tracking performance for the idea drift chamber," 11 2022, p. 362.

[51] M. Gronau, C. N. Leung, and J. L. Rosner, "Extending limits on neutral heavy leptons," *Phys. Rev. D*, vol. 29, p. 2539, 1984.

[52] D. Gorbunov and M. Shaposhnikov, "How to find neutral leptons of the νmsm?" *JHEP*, vol. 0710, p. 015, 2007.

[53] A. Atre, T. Han, S. Pascoli, and B. Zhang, "The search for heavy majorana neutrinos," *JHEP*, vol. 0905, p. 030, 2009.

[54] C. G. Cely, A. Ibarra, E. Molinaro, and S. Petcov, "Higgs decays in the low scale type i see-saw model," *Phys. Lett. B*, vol. 718, pp. 957–964, 2013.

[55] P. Abreu *et al.*, "Search for neutral heavy leptons produced in z decays," *Z. Phys. C*, vol. 74, pp. 57–71, 1997, [Erratum: Z.Phys.C 75, 580 (1997)].

[56] T. Sharma, "Phenomenology of dirac and majorana type heavy neutral leptons and their searches at the fcc-ee," 2022, presented 12-07-2022. [Online]. Available: https://cds.cern.ch/record/2816135

[57] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 2014, no. 7, jul 2014. [Online]. Available: http://dx.doi.org/10.1007/JHEP07(2014)079

[58] D. Alva, T. Han, and R. Ruiz, "Heavy majorana neutrinos from wγ fusion at hadron colliders," *Journal of High Energy Physics*, vol. 2015, no. 2, Feb 2015.

[59] C. Degrande, O. Mattelaer, R. Ruiz, and J. Turner, "Fully automated precision predictions for heavy neutrino production mechanisms at hadron colliders," *Physical Review D*, vol. 94, no. 5, Sep 2016.

[60] N. D. Christensen and C. Duhr, "Feynrules – feynman rules made easy," *Computer*

Physics Communications, vol. 180, no. 9, pp. 1614–1641, Sep 2009.

[61] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, "Ufo – the universal feynrules output," *Computer Physics Communications*, vol. 183, no. 6, pp. 1201–1214, Jun 2012.

[62] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, "Feynrules 2.0 — a complete toolbox for tree-level phenomenology," *Computer Physics Communications*, vol. 185, no. 8, pp. 2250–2300, Aug 2014.

[63] T. D. Collaboration, "Delphes simulation for fcc-ee using idea detector concept," https://github.com/delphes/delphes/blob/master/cards/delphes_card_IDEA.tcl, 2024.

[64] A. Ilg, "Design and performance of the idea vertex detector at fcc-ee in full simulation," in *Proceedings of the European Physical Society Conference on High Energy Physics (EPS-HEP2023)*, ser. PoS(EPS-HEP2023), vol. 600. European Physical Society, 2023, available online at https://inspirehep.net/files/edad889c55945f6a6f54b8029ddf3a1a.

[65] CERN, "Edm4hep," https://edm4hep.web.cern.ch/, accessed: 2024-04-18.

[66] HEP-FCC, "Common analysis framework for the future circular collider," https://github.com/HEP-FCC/FCCAnalyses, 2023.

[67] E. Guiraud, A. Naumann, and D. Piparo, "Tdataframe: functional chains for root data analyses (v1.0)," 2017. [Online]. Available: https://doi.org/10.5281/zenodo.260230

[68] D. Piparo, P. Canal, E. Guiraud, X. Pla, G. Ganis, G. Amadio, A. Naumann, and E. Tejedor, "Rdataframe: Easy parallel root analysis at 100 threads," in *EPJ Web of Conferences*, vol. 214. EDP Sciences, 2019, p. 06029. [Online]. Available: https://www.epj-conferences.org/articles/epjconf/abs/2019/19/epjconf_chep2018_06029/epjconf_chep2018_06029.html

[69] R. Brun and F. Rademakers, "Root - an object oriented data analysis framework," in *Proceedings of the AIHENP'96 Workshop*, vol. 389, Lausanne, sep 1997, pp. 81–86, see also "ROOT" [software], Release 6.30/06, 09/04/2024, https://zenodo.org/search?page=1&size=20&q=conceptrecid:848818&all_versions&sort=-version.

[70] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, "New clustering algorithm for multijet cross sections in e+e- annihilation," *Physics Letters B*, vol. 269, no. 3, pp. 432–438, Oct 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/037026939190196W

[71] F. Collaboration, "Delphes fccee physics events winter 2023 production (idea detector)," 2023. [Online]. Available: http://fcc-physics-events.web.cern.ch/fcc-physics-events/FCCee/winter2023/Delphesevents_IDEA.php

[72] D. Moulin, "Probing heavy neutral leptons in the ejj channel at the fcc-ee: Kinematic signatures and discovery potential," Master's thesis, University of Geneva, jun 2023. [Online]. Available: https://dpnc.unige.ch/MASTERS/MASTER_MOULIN_Dimitris.pdf

[73] L. Lyons, "Five sigma revisited," *CERN Courier*, Jul 2023. [Online]. Available: https://cerncourier.com/a/five-sigma-revisited/

[74] "Formulae for estimating significance," CERN, Geneva, Tech. Rep., 2020, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-025. [Online]. Available: http://cds.cern.ch/record/2736148

[75] L. Rygaard, "Long-lived heavy neutral leptons at the fcc-ee," Master's thesis, Uppsala University, 2023, accessed: 2024-04-12. [Online]. Available: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1679659

[76] R. Frühwirth and A. Strandlie, *Vertex Fitting*. Cham: Springer International Publishing, 2021, pp. 143–158. [Online]. Available: https://doi.org/10.1007/978-3-030-65771-0_8

[77] CMS Collaboration, "PAT Examples: Tracking, Vertexing and b-Tagging," 2024, [Accessed: 2024-06-05]. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookPATExampleTrackBJet

[78] A. Rizzi, F. Palla, and G. Segneri, "Track impact parameter based b-tagging with cms," *Scuola Normale Superiore di Pisa and INFN Pisa*, 2024. [Online]. Available: https://inspirehep.net/files/7644d94f5a048def74feabd4166aa464

[79] A. S. Cornella, W. Doorsamy, B. Fuksc, G. Harmsen, and L. Mason, "Boosted decision trees in the era of new physics: a smuon analysis case study," *arXiv preprint arXiv:2109.11815*, 2021. [Online]. Available: https://arxiv.org/abs/2109.11815

[80] Y. Wang, Z. Pan, J. Zheng, L. Qian, and L. Mingtao, "A hybrid ensemble method for pulsar candidate classification," *Astrophysics and Space Science*, vol. 364, 08 2019.

[81] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016,

pp. 785–794. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785

[82] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[83] P. Cunningham and S. J. Delany, "Algorithmic bias and regularisation in machine learning," *CoRR*, vol. abs/2005.09052, 2020. [Online]. Available: https://arxiv.org/abs/2005.09052

[84] K. Framling, "Feature importance versus feature influence and what it signifies for explainable ai," 2023. [Online]. Available: https://arxiv.org/abs/2308.03589

[85] K. Albertsson, S. Gleyzer, A. Hoecker, L. Moneta, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, and S. Wunsch, *TMVA 4: Toolkit for Multivariate Data Analysis with ROOT - Users Guide*, 2021. [Online]. Available: https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf

[86] D. I. MacKenzie, J. E. Hines *et al.*, "Fundamental principles of statistical inference," in *Occupancy Estimation and Modeling (Second Edition)*, 2018. [Online]. Available: https://www.sciencedirect.com/topics/mathematics/logit-link-function

[87] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[88] X. Contributors, "Model training," 2023. [Online]. Available: https://xgboost.readthedocs.io/en/latest/tutorials/model.html

[89] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[90] S. Mahapatra, "Why deep learning over traditional machine learning?" *Towards Data Science*, March 2018, accessed: 2024-05-28. [Online]. Available: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063

[91] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy,

W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, sep 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[92] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[93] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[94] L. Breiman and A. Cutler, "Permutation feature importance," 2023, accessed: 2024-05-28. [Online]. Available: https://scikit-learn.org/stable/modules/permutation_importance.html

[95] J. Milgram, M. Cheriet, and R. Sabourin, "A comparison of methods for multiclass support vector machines," *Neural Networks*, vol. 16, no. 8, pp. 1605–1615, 2006.

[96] J. M. Johnson and T. M. Khoshgoftaar, "Deep learning in class imbalance problems," *Journal of Big Data*, vol. 6, no. 1, p. 27, 2019.

[97] A. Fernandez, S. Garcia, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, "Class-imbalanced data: A review of the avenues of resampling," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 163–243, 2018.

[98] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?" 2022.

[99] D. McElfresh, S. Khandagale, J. Valverde, V. P. C, B. Feuer, C. Hegde, G. Ramakrishnan, M. Goldblum, and C. White, "When do neural nets outperform boosted trees on tabular data?" 2023.

[100] I. Krasnov, "Hnl see-saw: lower mixing limit and pseudodegenerate state," 7 2023.

[101] G. Polesello and N. Valle, "Sensitivity of the fcc-ee to decay of an hnl into a muon and two jets," INFN, Pavia, Physics Performance Meeting, May 2023, 22 May 2023 - Physics Performance Meeting.

[102] L. Bellagamba, "Hnls in the two muon final state at the fcc," Presentation at the FCC Workshop, 2023, accessed: 2024-05-28. [Online]. Available: https://indico.cern.ch/event/1282355/contributions/5387478/attachments/

2640912/4570195/hnl_2mu_fcc.pdf

A Additional Theory

A.1 Gamma Matrices of the Dirac Equation

To unify special relativity with quantum mechanics, the Schrödinger equation must be reformulated such that rather than the classical energy momentum relation, $E = p^2/2m$, it can be described using the relativistic Einstein relation, $E^2 = p^2 + m^2$. The result of this is the Klein-Gordon equation, which is second order in time and space and unfortunately predicts rather unphysical situations, such as negative probability densities. Hence, Dirac searched for a form of equation to satisfy Einstein's equation that was first order in space and time. It must then have the basic form of Equation 36, where $\boldsymbol{\alpha}$ is a 3-vector coefficient and β is a constant.

$$\hat{E}\psi = (\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \beta m)\psi, \tag{36}$$

The coefficients $\boldsymbol{\alpha}$ and β are constrained tightly by the fact that a solution to the Dirac equation must also satisfy the Klein Gordon Equation. The choices for $\boldsymbol{\alpha}$ and β are somewhat arbitrary; in the Dirac–Pauli representation we write them in terms of the Identity matrix and the Pauli spin-matrices:

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad \text{and} \quad \alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix},$$

with

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Such that the Dirac equation is manifestly covariant (i.e. Lorentz invariant), we must write it in terms of Lorentz 4-vectors so that it has the appropriate properties under selected spacetime transformations. To this end, the gamma matrices are introduced, with $\gamma^\mu = (\gamma^0, \gamma^1, \gamma^2, \gamma^3)$ allowing the Dirac equation to be covariantly written in the form of $(i\gamma^\mu \partial_\mu - m)\psi = 0$. In this way, the gamma matrices can be defined as in Equation 37, where the index k = 1,2,3.

$$\gamma^0 = \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad \text{and} \quad \gamma^k = \beta\alpha_k = \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix}, \tag{37}$$

A.2 Neutrino Oscillation Survival Probability

Starting with the PMNS matrix and the nine imposed unitarity constraints (since $U^{\dagger}U = I$), we can consider the linear superposition of the mass eigenstates which form the weak eigenstate, as seen in Figure 69 below.

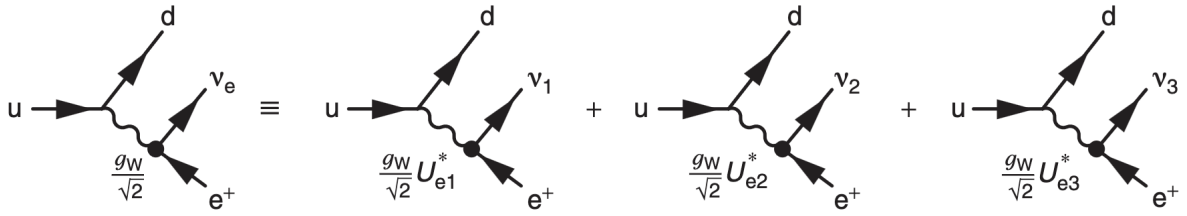


Figure 69: The Feynman diagram for β^+ decay broken down into the contributions from the different mass eigenstates (Image: M. Thompson [19]).

The time evolution of the wave function, ψ is given by the time evolution of the mass eigenstates dictated by the complex phase $-i\phi$, where $\phi = p_i \cdot x_i = (E_i t - \boldsymbol{p}_i \cdot \boldsymbol{x})$.

$$|\psi(\mathbf{x},t)\rangle = U^*_{\mathrm{e}1}\,|v_1\rangle\,e^{-i\phi_1} + U^*_{\mathrm{e}2}\,|v_2\rangle\,e^{-i\phi_2} + U^*_{\mathrm{e}3}\,|v_3\rangle\,e^{-i\phi_3} \tag{38}$$

Equation 38 can be reformulated by writing the composition of each mass eigenstate in terms of the weak eigenstates and as a result we find the result given in Equation 39.

$$\begin{aligned}|\psi(\mathbf{x},t)\rangle =& U^*_{\mathrm{e}1}\left(U_{\mathrm{e}1}\,|v_{\mathrm{e}}\rangle + U_{\mu 1}\,|v_\mu\rangle + U_{\tau 1}\,|v_\tau\rangle\right)e^{-i\phi_1}\\ &+ U^*_{\mathrm{e}2}\left(U_{\mathrm{e}2}\,|v_{\mathrm{e}}\rangle + U_{\mu 2}\,|v_\mu\rangle + U_{\tau 2}\,|v_\tau\rangle\right)e^{-i\phi_2}\\ &+ U^*_{\mathrm{e}3}\left(U_{\mathrm{e}3}\,|v_{\mathrm{e}}\rangle + U_{\mu 3}\,|v_\mu\rangle + U_{\tau 3}\,|v_\tau\rangle\right)e^{-i\phi_3}\end{aligned} \tag{39}$$

We can further break down Equation 39 by gathering up the terms for each weak eigenstate and find Equation 40.

$$\begin{aligned}|\psi(\mathbf{x},t)\rangle =& \left(U^*_{\mathrm{e}1}U_{\mathrm{e}1}e^{-i\phi_1} + U^*_{\mathrm{e}2}U_{\mathrm{e}2}e^{-i\phi_2} + U^*_{\mathrm{e}3}U_{\mathrm{e}3}e^{-i\phi_3}\right)|v_{\mathrm{e}}\rangle\\ &\left(U^*_{\mathrm{e}1}U_{\mu 1}e^{-i\phi_1} + U^*_{\mathrm{e}2}U_{\mu 2}e^{-i\phi_2} + U^*_{\mathrm{e}3}U_{\mu 3}e^{-i\phi_3}\right)|v_\mu\rangle\\ &\left(U^*_{\mathrm{e}1}U_{\tau 1}e^{-i\phi_1} + U^*_{\mathrm{e}2}U_{\tau 2}e^{-i\phi_2} + U^*_{\mathrm{e}3}U_{\tau 3}e^{-i\phi_3}\right)|v_\tau\rangle\end{aligned} \tag{40}$$

The unitarity constraints take the form e.g. $U_{e1}U^*_{e1} + U_{e2}U^*_{e2} + U_{e3}U^*_{e3} = 1, U_{e1}U^*_{\mu 1} + U_{e2}U^*_{\mu 2} + U_{e3}U^*_{\mu 3} = 0$. The survival probability for an electron neutrino can be written as $P(\nu_e \to \nu_e) = |\langle\nu_e|\psi(\mathbf{x},t)\rangle|^2$. By writing Equation 40 in the form $|\psi(\mathbf{x},t)\rangle = c_e|\nu_e\rangle + c_\mu|\nu_\mu\rangle + c_\tau|\nu_\tau\rangle$, we can say that$P(\nu_e \to \nu_e) = |U^*_{e1}U_{e1}e^{-i\phi_1} + U^*_{e2}U_{e2}e^{-i\phi_2} + U^*_{e3}U_{e3}e^{-i\phi_3}|^2$.

The complex number identity $|z_1+z_2+z_3|^2 \equiv |z_1|^2+|z_2|^2+|z_3|^2+2\Re\{z_1z_2^*+z_1z_3^*+z_2z_3^*\}$

can then be used to simplify our expression in conjunction with the unitarity condition to show: $P(\nu_e \to \nu_e) = 1 + 2|U_{e1}|^2|U_{e2}|^2\Re\{e^{i(\phi_2-\phi_1)} - 1\} + 2|U_{e1}|^2|U_{e3}|^2\Re\{e^{i(\phi_3-\phi_1)} - 1\} + 2|U_{e2}|^2|U_{e3}|^2\Re\{e^{i(\phi_3-\phi_2)} - 1\}$.

By noting that $\Re\{e^{i(\phi_j-\phi_i)} - 1\} = \cos(\phi_j - \phi_i) - 1 = -2\sin^2\left(\frac{\phi_j-\phi_i}{2}\right) = -2\sin^2\Delta_j$, where $\Delta_{ji} = \frac{\phi_j-\phi_i}{2} = \frac{(m_j^2-m_i^2)L}{4E_\nu}$ we find that the survival probability is indeed given by Equation 10 as stated in Section 2.2.1.

B Process Cards

B.1 MG5 Process Card for Dirac HNLs

```
set default unset_couplings 99
set group_subprocesses Auto
set ignore_six_quark_processes False
set loop_optimized_output True
set loop_color_flows False
set gauge unitary
set complex_mass_scheme False
set max_npoint_for_channel 0
import model sm
define p = g u c d s b t u~ c~ d~ s~ b~ t~
define j = g u c d s b t u~ c~ d~ s~ b~ t~
define l+ = e+ mu+
define l- = e- mu-
define vl = ve vm vt
define vl~ = ve~ vm~ vt~
import model SM_HeavyN_Dirac_CKM_Masses_LO
define e = e+ e-
define nue = ve ve~
generate e+ e- > n1~ ve , (n1~ > e+ jj)
add process e+ e- > n1 ve~ , (n1 > e- jj)
output HNL_Dirac_ejj_10GeV_1e-4Ve
launch HNL_Dirac_ejj_10GeV_1e-4Ve
done
# set to electron beams (0 for ele, 1 for proton)
set lpp1 0
set lpp2 0
set ebeam1 45.594
set ebeam2 45.594
set no_parton_cut
# Here set mass of the electron HNL
set mn1 10
# set mass of muon HNL, made heavy here
set mn2 10000
# set mass of tau HNL, made heavy here
set mn3 10000
# set electron mixing angle
set ven1 1e-4
set WN1 auto
set time_of_flight 0
set nevents 100000
done
```

Listing 1: MG5 Process Card for Dirac HNLs

B.2 MG5 Process Card for 4-body background

```
set default unset_couplings 99
set group_subprocesses Auto
set ignore_six_quark_processes False
set loop_optimized_output True
set loop_color_flows False
set gauge unitary
set complex_mass_scheme False
set max_npoint_for_channel 0
import model sm
define p = g u c d s u~ c~ d~ s~ b b~
define j = g u c d s u~ c~ d~ s~ b b~
define l+ = e+ mu+
define l- = e- mu-
define vl = ve vm vt
define vl~ = ve~ vm~ vt~
define vlpm = vl vl~
define e = e+ e-
generate e+ e- > e vlpm jj
output enuqq
launch enuqq
done
set nevents 100000
set ptj 0
set ptl 0
set etal 5
set mmjj 5
set drjj 0
set drjl 0
set maxjetflavor 5
set lpp1 0
set lpp2 0
set ebeam1 45.8
set ebeam2 45.8
set time_of_flight 1
done
```

Listing 2: MG5 Process Card for 4-body background

C Additional Event Statistics

C.1 Signal Sample Cross Sections

Table 14: Summary table showing the cross section, raw-events, and production luminosity for each of the signal points considered in this analysis.

Process	$\boldsymbol{\sigma}$(pb)	**Monte-Carlo Events**	**Production** $\mathcal{L}$ (fb^{-1})
10 GeV, $\lvert U^2\rvert = 10^{-4}$	4.06×10^{-1}	1.00×10^{5}	2.46×10^{5}
10 GeV, $\lvert U^2\rvert = 10^{-5}$	4.06×10^{-2}	1.00×10^{5}	2.46×10^{6}
10 GeV, $\lvert U^2\rvert = 10^{-6}$	4.06×10^{-3}	1.00×10^{5}	2.46×10^{7}
10 GeV, $\lvert U^2\rvert = 10^{-7}$	4.06×10^{-4}	1.00×10^{5}	2.46×10^{8}
10 GeV, $\lvert U^2\rvert = 10^{-8}$	4.06×10^{-5}	1.00×10^{5}	2.46×10^{9}
10 GeV, $\lvert U^2\rvert = 10^{-9}$	4.06×10^{-6}	1.00×10^{5}	2.46×10^{10}
10 GeV, $\lvert U^2\rvert = 10^{-10}$	4.06×10^{-7}	1.00×10^{5}	2.46×10^{11}
20 GeV, $\lvert U^2\rvert = 10^{-4}$	3.77×10^{-1}	1.00×10^{5}	2.65×10^{5}
20 GeV, $\lvert U^2\rvert = 10^{-5}$	3.77×10^{-2}	1.00×10^{5}	2.65×10^{6}
20 GeV, $\lvert U^2\rvert = 10^{-6}$	3.77×10^{-3}	1.00×10^{5}	2.65×10^{7}
20 GeV, $\lvert U^2\rvert = 10^{-7}$	3.77×10^{-4}	1.00×10^{5}	2.65×10^{8}
20 GeV, $\lvert U^2\rvert = 10^{-8}$	3.77×10^{-5}	1.00×10^{5}	2.65×10^{9}
20 GeV, $\lvert U^2\rvert = 10^{-9}$	3.77×10^{-6}	1.00×10^{5}	2.65×10^{10}
20 GeV, $\lvert U^2\rvert = 10^{-10}$	3.77×10^{-7}	1.00×10^{5}	2.65×10^{11}
30 GeV, $\lvert U^2\rvert = 10^{-4}$	3.36×10^{-1}	1.00×10^{5}	2.97×10^{5}
30 GeV, $\lvert U^2\rvert = 10^{-5}$	3.36×10^{-2}	1.00×10^{5}	2.97×10^{6}
30 GeV, $\lvert U^2\rvert = 10^{-6}$	3.36×10^{-3}	1.00×10^{5}	2.97×10^{7}
30 GeV, $\lvert U^2\rvert = 10^{-7}$	3.36×10^{-4}	1.00×10^{5}	2.97×10^{8}
30 GeV, $\lvert U^2\rvert = 10^{-8}$	3.36×10^{-5}	1.00×10^{5}	2.97×10^{9}
30 GeV, $\lvert U^2\rvert = 10^{-9}$	3.36×10^{-6}	1.00×10^{5}	2.97×10^{10}
30 GeV, $\lvert U^2\rvert = 10^{-10}$	3.36×10^{-7}	1.00×10^{5}	2.97×10^{11}
40 GeV, $\lvert U^2\rvert = 10^{-4}$	2.87×10^{-1}	1.00×10^{5}	3.49×10^{5}
40 GeV, $\lvert U^2\rvert = 10^{-5}$	2.87×10^{-2}	1.00×10^{5}	3.49×10^{6}
40 GeV, $\lvert U^2\rvert = 10^{-6}$	2.87×10^{-3}	1.00×10^{5}	3.49×10^{7}
40 GeV, $\lvert U^2\rvert = 10^{-7}$	2.87×10^{-4}	1.00×10^{5}	3.49×10^{8}
40 GeV, $\lvert U^2\rvert = 10^{-8}$	2.87×10^{-5}	1.00×10^{5}	3.49×10^{9}
40 GeV, $\lvert U^2\rvert = 10^{-9}$	2.87×10^{-6}	1.00×10^{5}	3.49×10^{10}
40 GeV, $\lvert U^2\rvert = 10^{-10}$	2.87×10^{-7}	1.00×10^{5}	3.49×10^{11}
50 GeV, $\lvert U^2\rvert = 10^{-4}$	2.27×10^{-1}	1.00×10^{5}	4.41×10^{5}
50 GeV, $\lvert U^2\rvert = 10^{-5}$	2.27×10^{-2}	1.00×10^{5}	4.41×10^{6}

Continued on next page

Table 14 continued from previous page

Process	$\boldsymbol{\sigma}$(pb)	**Monte-Carlo Events**	**Production** $\mathcal{L}$ (fb^{-1})
50 GeV, $\|U^2\| = 10^{-6}$	2.27×10^{-3}	1.00×10^{5}	4.41×10^{7}
50 GeV, $\|U^2\| = 10^{-7}$	2.27×10^{-4}	1.00×10^{5}	4.41×10^{8}
50 GeV, $\|U^2\| = 10^{-8}$	2.27×10^{-5}	1.00×10^{5}	4.41×10^{9}
50 GeV, $\|U^2\| = 10^{-9}$	2.27×10^{-6}	1.00×10^{5}	4.41×10^{10}
50 GeV, $\|U^2\| = 10^{-10}$	2.27×10^{-7}	1.00×10^{5}	4.41×10^{11}
60 GeV, $\|U^2\| = 10^{-4}$	1.59×10^{-1}	1.00×10^{5}	6.29×10^{5}
60 GeV, $\|U^2\| = 10^{-5}$	1.59×10^{-2}	1.00×10^{5}	6.29×10^{6}
60 GeV, $\|U^2\| = 10^{-6}$	1.59×10^{-3}	1.00×10^{5}	6.29×10^{7}
60 GeV, $\|U^2\| = 10^{-7}$	1.59×10^{-4}	1.00×10^{5}	6.29×10^{8}
60 GeV, $\|U^2\| = 10^{-8}$	1.59×10^{-5}	1.00×10^{5}	6.29×10^{9}
60 GeV, $\|U^2\| = 10^{-9}$	1.59×10^{-6}	1.00×10^{5}	6.29×10^{10}
60 GeV, $\|U^2\| = 10^{-10}$	1.59×10^{-7}	1.00×10^{5}	6.29×10^{11}
70 GeV, $\|U^2\| = 10^{-4}$	9.06×10^{-2}	1.00×10^{5}	1.10×10^{6}
70 GeV, $\|U^2\| = 10^{-5}$	9.06×10^{-3}	1.00×10^{5}	1.10×10^{7}
70 GeV, $\|U^2\| = 10^{-6}$	9.06×10^{-4}	1.00×10^{5}	1.10×10^{8}
70 GeV, $\|U^2\| = 10^{-7}$	9.06×10^{-5}	1.00×10^{5}	1.10×10^{9}
70 GeV, $\|U^2\| = 10^{-8}$	9.06×10^{-6}	1.00×10^{5}	1.10×10^{10}
70 GeV, $\|U^2\| = 10^{-9}$	9.06×10^{-7}	1.00×10^{5}	1.10×10^{11}
70 GeV, $\|U^2\| = 10^{-10}$	9.06×10^{-8}	1.00×10^{5}	1.10×10^{12}
80 GeV, $\|U^2\| = 10^{-4}$	1.85×10^{-1}	1.00×10^{5}	5.41×10^{5}
80 GeV, $\|U^2\| = 10^{-5}$	1.85×10^{-2}	1.00×10^{5}	5.41×10^{6}
80 GeV, $\|U^2\| = 10^{-6}$	1.85×10^{-3}	1.00×10^{5}	5.41×10^{7}
80 GeV, $\|U^2\| = 10^{-7}$	1.85×10^{-4}	1.00×10^{5}	5.41×10^{8}
80 GeV, $\|U^2\| = 10^{-8}$	1.85×10^{-5}	1.00×10^{5}	5.41×10^{9}
80 GeV, $\|U^2\| = 10^{-9}$	1.85×10^{-6}	1.00×10^{5}	5.41×10^{10}
80 GeV, $\|U^2\| = 10^{-10}$	1.85×10^{-7}	1.00×10^{5}	5.41×10^{11}

C.2 Filter Efficiencies

Table 15: BDT filter efficiencies for background and signal processes with the filter condition requiring the lead electron energy > 15 GeV. Additionally, the number of final state reconstructed electrons is required to be exactly one.

Process	No Filter	Lead Electron Energy > 15 GeV	Efficiency (%)
Background			
$Z \to bb$	438,538,637	13,074,288	2.98
$Z \to cc$	498,091,935	3,875,708	0.78
$Z \to 4body$	100,000	17,127	17.13
Signals			
10 GeV, $\|U^2\| = 10^{-4}$	100,000	50,270	50.27
10 GeV, $\|U^2\| = 10^{-5}$	100,000	50,198	50.20
10 GeV, $\|U^2\| = 10^{-6}$	100,000	50,266	50.27
10 GeV, $\|U^2\| = 10^{-7}$	38,667	19,319	49.96
10 GeV, $\|U^2\| = 10^{-8}$	100,000	23,017	23.02
10 GeV, $\|U^2\| = 10^{-9}$	100,000	3,050	3.05
10 GeV, $\|U^2\| = 10^{-10}$	100,000	308	0.31
20 GeV, $\|U^2\| = 10^{-4}$	100,000	52,269	52.27
20 GeV, $\|U^2\| = 10^{-5}$	100,000	52,586	52.59
20 GeV, $\|U^2\| = 10^{-6}$	100,000	52,508	52.51
20 GeV, $\|U^2\| = 10^{-7}$	100,000	52,391	52.39
20 GeV, $\|U^2\| = 10^{-8}$	76,400	39,971	52.32
20 GeV, $\|U^2\| = 10^{-9}$	9,707	4,947	50.96
20 GeV, $\|U^2\| = 10^{-10}$	45,948	8,888	19.34
30 GeV, $\|U^2\| = 10^{-4}$	100,000	54,236	54.24
30 GeV, $\|U^2\| = 10^{-5}$	100,000	54,278	54.28
30 GeV, $\|U^2\| = 10^{-6}$	100,000	54,230	54.23
30 GeV, $\|U^2\| = 10^{-7}$	100,000	54,316	54.32
30 GeV, $\|U^2\| = 10^{-8}$	100,000	54,334	54.33
30 GeV, $\|U^2\| = 10^{-9}$	100,000	54,192	54.19
40 GeV, $\|U^2\| = 10^{-4}$	100,000	56,398	56.40
40 GeV, $\|U^2\| = 10^{-5}$	100,000	56,397	56.40
40 GeV, $\|U^2\| = 10^{-6}$	100,000	56,391	56.39
40 GeV, $\|U^2\| = 10^{-7}$	100,000	56,361	56.36
40 GeV, $\|U^2\| = 10^{-8}$	100,000	56,391	56.39

Continued on next page

Table 15 continued from previous page

Process	No Filter	Lead Electron Energy > 15 GeV	Efficiency (%)
40 GeV, $\|U^2\| = 10^{-9}$	100,000	56,366	56.37
40 GeV, $\|U^2\| = 10^{-10}$	7,196	4,031	56.02
50 GeV, $\|U^2\| = 10^{-4}$	100,000	59,958	59.96
50 GeV, $\|U^2\| = 10^{-5}$	100,000	60,106	60.11
50 GeV, $\|U^2\| = 10^{-6}$	100,000	59,982	59.98
50 GeV, $\|U^2\| = 10^{-7}$	100,000	59,991	59.99
50 GeV, $\|U^2\| = 10^{-8}$	100,000	60,126	60.13
50 GeV, $\|U^2\| = 10^{-9}$	100,000	59,949	59.95
50 GeV, $\|U^2\| = 10^{-10}$	100,000	59,941	59.94
60 GeV, $\|U^2\| = 10^{-4}$	100,000	64,604	64.60
60 GeV, $\|U^2\| = 10^{-5}$	100,000	64,604	64.60
60 GeV, $\|U^2\| = 10^{-6}$	100,000	64,598	64.60
60 GeV, $\|U^2\| = 10^{-7}$	100,000	64,587	64.59
60 GeV, $\|U^2\| = 10^{-8}$	100,000	64,534	64.53
60 GeV, $\|U^2\| = 10^{-9}$	100,000	64,604	64.60
60 GeV, $\|U^2\| = 10^{-10}$	100,000	64,600	64.60
70 GeV, $\|U^2\| = 10^{-4}$	100,000	64,387	64.39
70 GeV, $\|U^2\| = 10^{-5}$	100,000	64,388	64.39
70 GeV, $\|U^2\| = 10^{-6}$	100,000	64,387	64.39
70 GeV, $\|U^2\| = 10^{-7}$	100,000	64,387	64.39
70 GeV, $\|U^2\| = 10^{-8}$	100,000	64,400	64.40
70 GeV, $\|U^2\| = 10^{-9}$	100,000	64,393	64.39
70 GeV, $\|U^2\| = 10^{-10}$	100,000	64,399	64.40
80 GeV, $\|U^2\| = 10^{-4}$	100,000	48,708	48.71
80 GeV, $\|U^2\| = 10^{-5}$	100,000	48,709	48.71
80 GeV, $\|U^2\| = 10^{-6}$	100,000	48,708	48.71
80 GeV, $\|U^2\| = 10^{-7}$	100,000	48,706	48.71
80 GeV, $\|U^2\| = 10^{-8}$	100,000	48,715	48.72
80 GeV, $\|U^2\| = 10^{-9}$	100,000	48,725	48.73
80 GeV, $\|U^2\| = 10^{-10}$	100,000	48,701	48.70

Table 16: BDT filter efficiencies for background and signal processes as part of the data pre-processing, where the filter conditions require the lead electron energy > 20 GeV, additionally, the number of final state reconstructed electrons is required to be exactly one.

Process	Initial Events	Final Events	Efficiency (%)
Background			
$Z \to bb$	438,538,637	4,692,700	1.07
$Z \to cc$	498,091,935	890,278	0.179
$Z \to 4body$	100,000	13,375	13.375
Signals			
10 GeV, $\|U^2\| = 10^{-4}$	100,000	37,037	37.037
10 GeV, $\|U^2\| = 10^{-5}$	100,000	37,000	37.000
10 GeV, $\|U^2\| = 10^{-6}$	100,000	16,944	16.944
10 GeV, $\|U^2\| = 10^{-7}$	100,000	2,250	2.250
10 GeV, $\|U^2\| = 10^{-10}$	100,000	228	0.228
20 GeV, $\|U^2\| = 10^{-4}$	100,000	39,079	39.079
20 GeV, $\|U^2\| = 10^{-5}$	100,000	39,234	39.234
20 GeV, $\|U^2\| = 10^{-6}$	76,400	29,755	38.946
20 GeV, $\|U^2\| = 10^{-7}$	9,707	3,677	37.880
20 GeV, $\|U^2\| = 10^{-10}$	45,948	6,580	14.321
30 GeV, $\|U^2\| = 10^{-4}$	100,000	40,102	40.102
30 GeV, $\|U^2\| = 10^{-5}$	100,000	40,144	40.144
30 GeV, $\|U^2\| = 10^{-6}$	100,000	40,237	40.237
30 GeV, $\|U^2\| = 10^{-7}$	100,000	40,201	40.201
30 GeV, $\|U^2\| = 10^{-10}$	100,000	N/A	N/A
40 GeV, $\|U^2\| = 10^{-4}$	100,000	41,295	41.295
40 GeV, $\|U^2\| = 10^{-5}$	100,000	41,293	41.293
40 GeV, $\|U^2\| = 10^{-6}$	100,000	41,299	41.299
40 GeV, $\|U^2\| = 10^{-7}$	100,000	41,291	41.291
40 GeV, $\|U^2\| = 10^{-10}$	7,196	2,925	40.648
50 GeV, $\|U^2\| = 10^{-4}$	100,000	42,874	42.874
50 GeV, $\|U^2\| = 10^{-5}$	100,000	43,063	43.063
50 GeV, $\|U^2\| = 10^{-6}$	100,000	43,076	43.076
50 GeV, $\|U^2\| = 10^{-7}$	100,000	42,941	42.941
50 GeV, $\|U^2\| = 10^{-10}$	100,000	42,896	42.896

Continued on next page

Table 16 continued from previous page

Process	Initial Events	Final Events	Efficiency (%)
60 GeV, $\|U^2\| = 10^{-4}$	100,000	46,242	46.242
60 GeV, $\|U^2\| = 10^{-5}$	100,000	46,242	46.242
60 GeV, $\|U^2\| = 10^{-6}$	100,000	46,219	46.219
60 GeV, $\|U^2\| = 10^{-7}$	100,000	46,271	46.271
60 GeV, $\|U^2\| = 10^{-10}$	100,000	46,266	46.266
70 GeV, $\|U^2\| = 10^{-4}$	100,000	47,901	47.901
70 GeV, $\|U^2\| = 10^{-5}$	100,000	47,902	47.902
70 GeV, $\|U^2\| = 10^{-6}$	100,000	47,912	47.912
70 GeV, $\|U^2\| = 10^{-7}$	100,000	47,910	47.910
70 GeV, $\|U^2\| = 10^{-10}$	100,000	47,934	47.934
80 GeV, $\|U^2\| = 10^{-4}$	100,000	35,831	35.831
80 GeV, $\|U^2\| = 10^{-5}$	100,000	35,831	35.831
80 GeV, $\|U^2\| = 10^{-6}$	100,000	35,825	35.825
80 GeV, $\|U^2\| = 10^{-7}$	100,000	35,837	35.837
80 GeV, $\|U^2\| = 10^{-10}$	100,000	35,792	35.792