UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

[†]thomas.critchley@etu.unige.ch
[†]anna.sfyrla@unige.ch
[†]pantelis.kontaxakis@unige.ch

Machine Learning Techniques to Probe HNLs at the FCC-ee

CHiPP 2024 Annual Meeting

June 19th 2024

Thomas Critchley[†]

Supervsiors: *Prof. Anna Sfyrla*[†], *Dr. Pantelis Kontaxakis*[†]

FUTURE CIRCULAR COLLIDER

Heavy Neutral Leptons

- Neutrinos are massive objects with very small masses, as shown by baseline neutrino oscillation experiments
- Low-scale inverse seesaw mechnism allows us to search for heavy right handed neutrinos with Yukawa couplings $O(10^{-6})$ in a mass range between 10 – 100 GeV
- In our analysis, we search for the electron final state with two jets, in the (pseudo-) Dirac HNL model between 10 - 80 GeV with mixing angles between $10^{-4} < |U_{eN}|^2 < 10^{-10}$

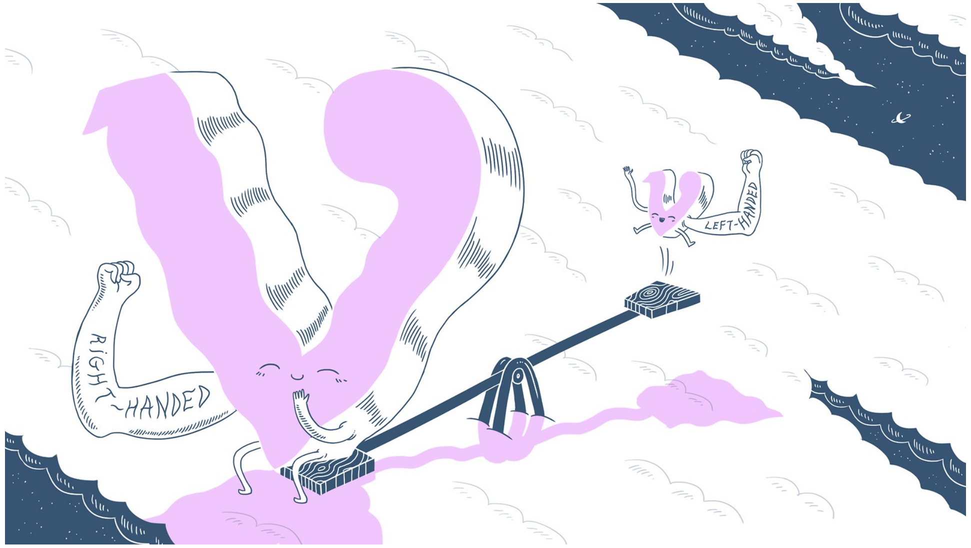


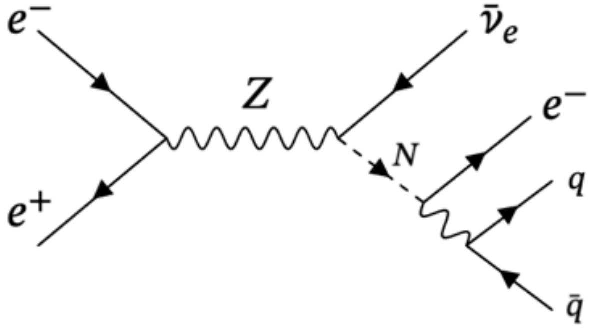
Image: Symmetry magazine

$$M = \begin{bmatrix} 0 & m_D & 0 \\ m_D & \mu_N & m_R \\ 0 & m_R & \mu_S \end{bmatrix} \longrightarrow \begin{pmatrix} m_\nu & 0 & 0 \\ 0 & m_N & 0 \\ 0 & 0 & m_S \end{pmatrix}$$

$$\mu_N, \mu_S \ll m_D \ll m_R.$$

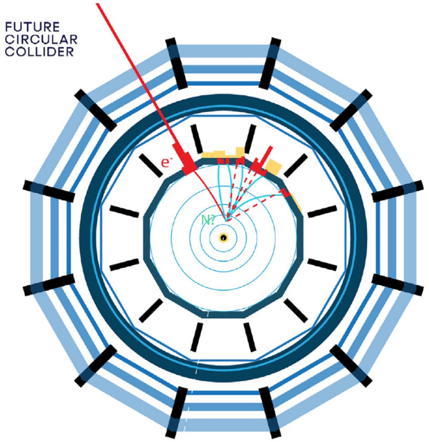
$$m_\nu \approx \frac{\mu_S}{m_D^2} \frac{m_D^2}{m_D^2 + m_R^2},$$

$$m_N, m_S \approx \sqrt{\frac{m_D^2 + m_R^2}{2}} \mp \frac{\mu_S}{2} \frac{m_R^2}{m_D^2 + m_R^2}.$$

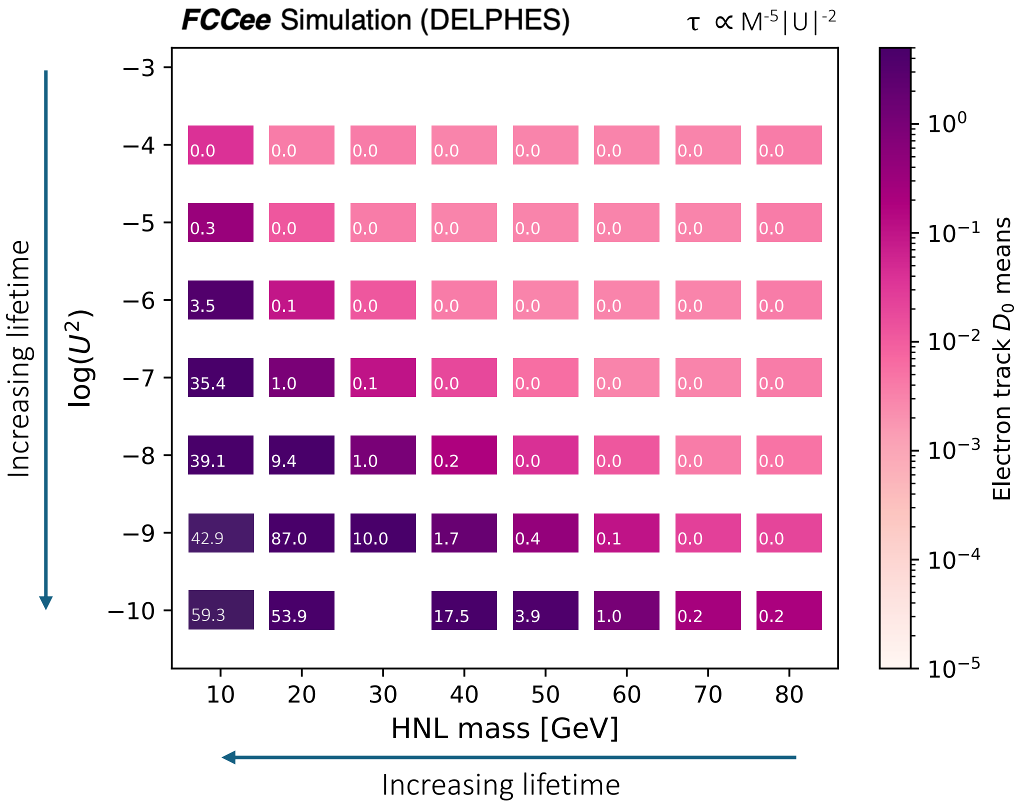


Example LNC diagram for $e^+e^- \to Z \to N\nu \to eqq$

HNL Phenomology at the FCCee

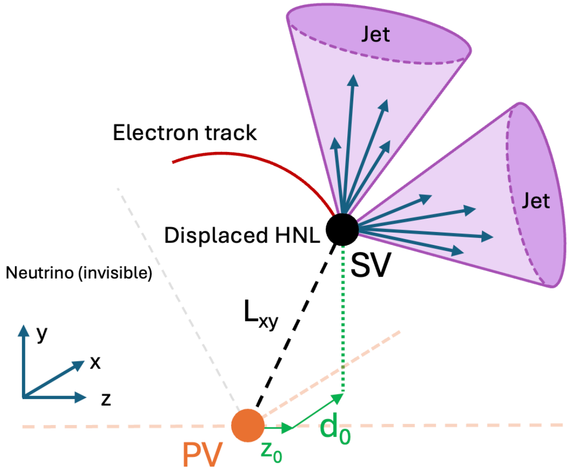


- The FCC-ee will produce some 10^{12} Z bosons in the Z-channel run (~3 years of data taking), giving a pileup free high luminosity environment to search for HNLs, and improve upon limits such as those set by LEP



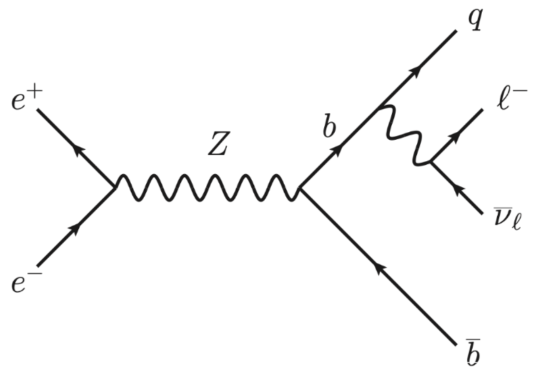
- For many of the mass points considered, a displaced topology arises from consequential lifetime ($\tau \propto M^{-5}|U|^{-2}$)

- We can tease these signals apart from prompty decaying mass points using any metric to paramaterise the lifetime, e.g. the decay length, the D_0 etc.

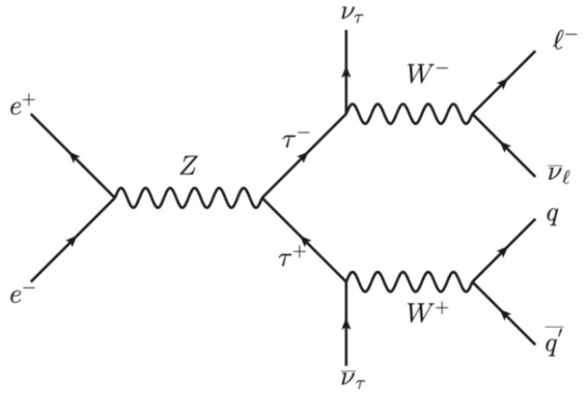


Analysis methods

- We consider three SM background processes which dominate the interaction, Z → bb, cc or Z → 4 body final state.



Z → bb production feynman diagram



Example of 4 body final state background

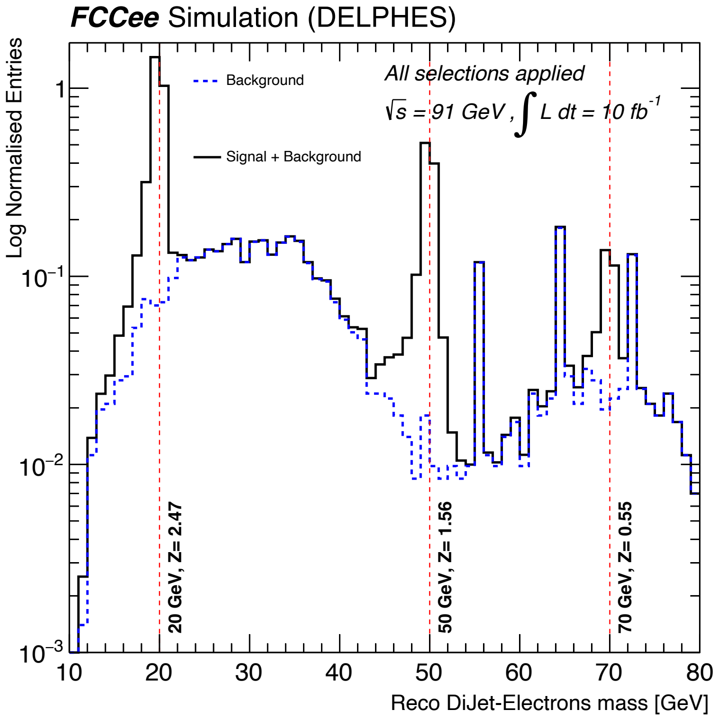
Process	$\boldsymbol{\sigma}$(pb)	**Monte-Carlo events**	**Production** $\mathcal{L}$ (fb^{-1})
Z → bb	6.65×10^3	4.39×10^8	6.60×10^1
Z → cc	5.22×10^3	4.98×10^8	1.15×10^2
Z → 4body	1.40×10^{-2}	1.00×10^5	7.14×10^3

Process	$\boldsymbol{\sigma}$ (pb)	**Monte-Carlo events**	**Events at** $\mathcal{L}$ =	
			10 fb^{-1}	150 ab^{-1}
Z → bb	6.65×10^3	4.39×10^8	6.65×10^7	9.98×10^{11}
Z → cc	5.22×10^3	4.98×10^8	5.22×10^7	7.82×10^{11}
Z → 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

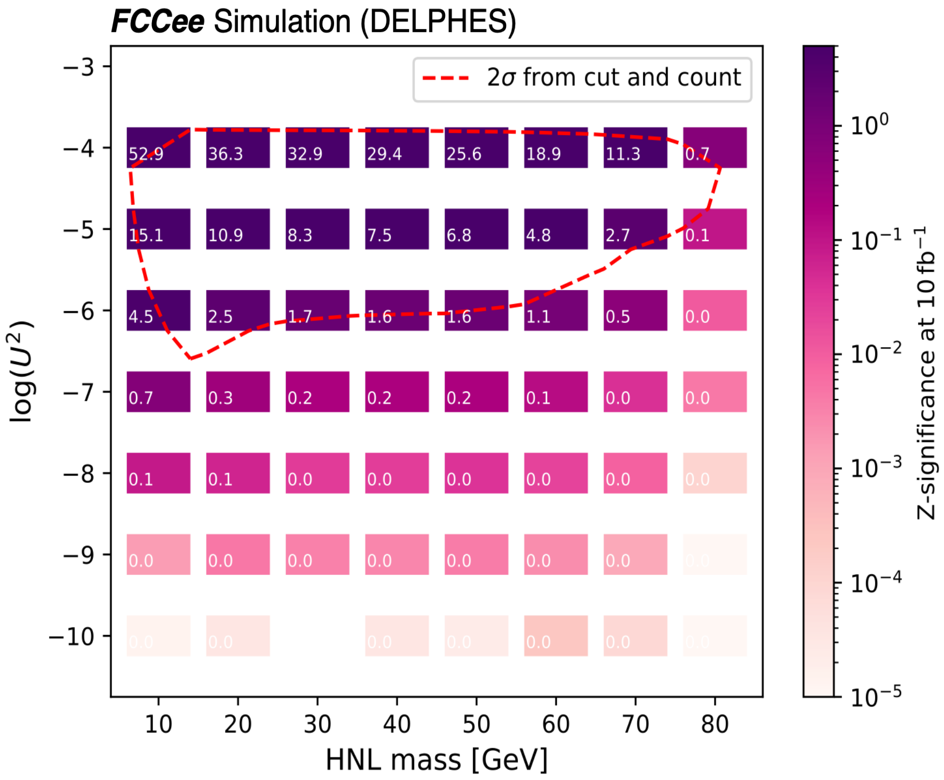
Quite limited by statitsics, only have around 10 fb^{-1} of lumi with which to model the 150 ab^{-1} FCC lumi, so we only scale to the full lumi in the final result, and elsewhere work at 10 fb^{-1}

Cut and Count summary

- Cut and count study was replicated to match the cuts made in D. Moulin thesis (2023), as a benchmark for optimisation



Invariant mass dist after all cuts



2σ delimitation after all cuts

Cuts chosen:

Variable	Selection
Missing energy	> 12 GeV
Leading electron energy	> 35 GeV
3D di-jet Angle	< 2.4 rad
Di-jet – Electron ΔR	< 3

Normalising factor:

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

Significance:

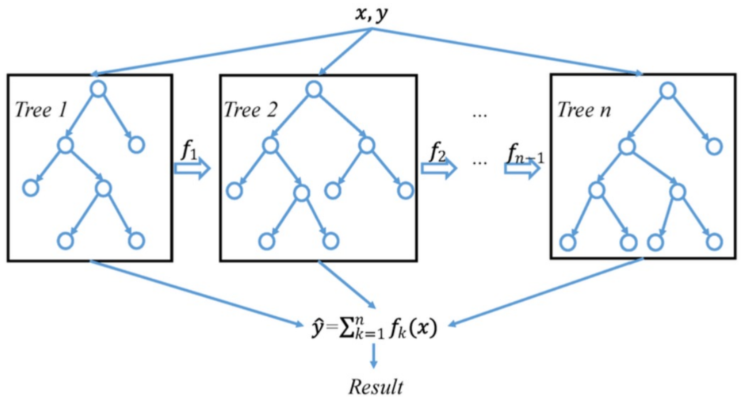
$$Z = \sqrt{2\left(n \cdot \ln\left[\frac{n\,(b+\sigma^2)}{b^2 + n\sigma^2}\right] - \frac{b^2}{\sigma^2}\ln\left[1 + \frac{\sigma^2(n-b)}{b\,(b+\sigma^2)}\right]\right)}$$

Optimisation strategy

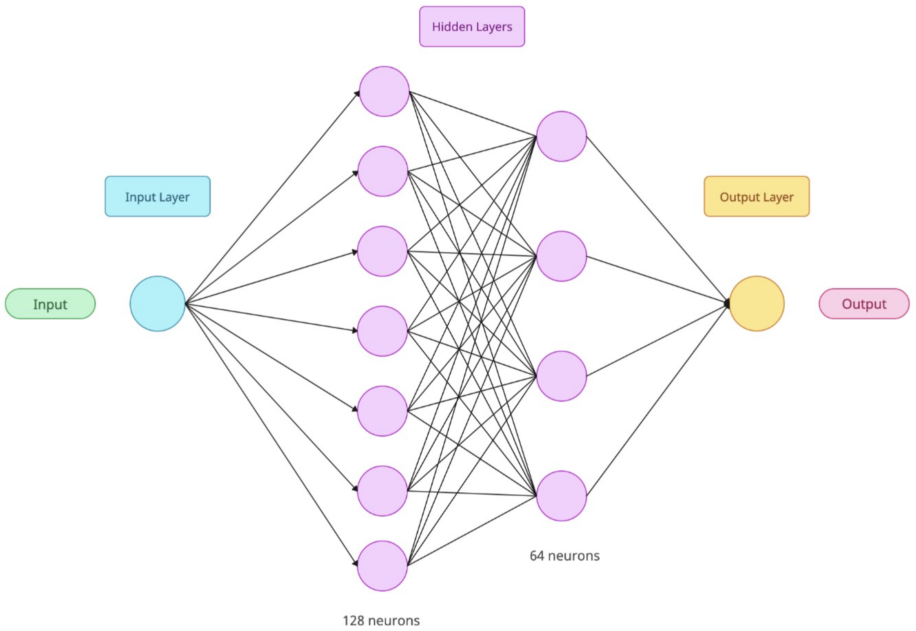
- Boosted Decision Trees (BDTs) and Deep Neural Networks (DNNs) are the natural extension of the cut and count study, we can make a single optimised cut on the BDT/DNN output rather than having to make sequential cuts on specific variables, giving more flexibility and utilising any correlation between discriminating variables

- A ML model can be trained for each individual mass point, meaning we need not focus on some benchmark mass points to find global cuts – this limits our capacity to fully exploit features like prompt and LLPs which a ML can naturally find!

- For the BDTs, XGBoost is used in conjunction with TMVA



- For the DNN models, Keras in Tensorflow is used

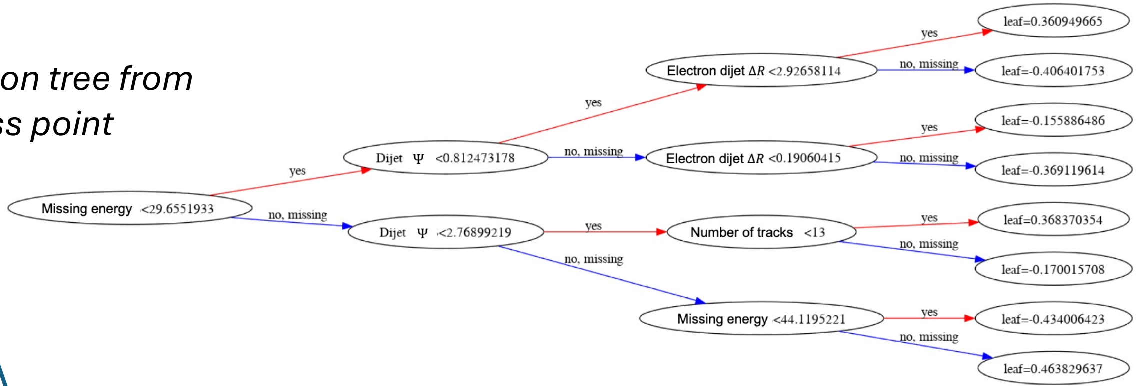


- In both cases, we use the following features to train:

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	$n_{\text{tracks}}, n_{\text{primary tracks}}, \chi^2_{\text{vertex}}$

BDT workflow

“simple” single decision tree from 10 GeV $|U|^2 = 10^{-4}$ mass point



Data Preparation:
- Stage 1 flat ntuples loaded
- Filter applied: E > 15 GeV
- Training and testing split
- Data sets saved using TMVA

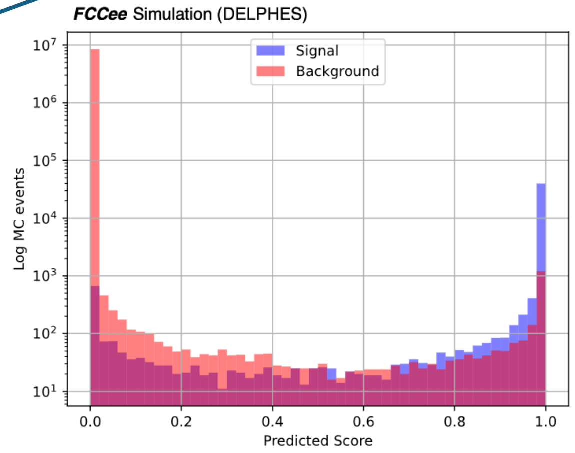
Model training:
- GridSearchCV
- Decision trees made
- Model saved in ROOT file (TMVA)

Model predictions:
- Models applied to test
- Normalisation to 10 fb^{-1}
- BDT cut chosen based on optimal significance (as with cut and count)

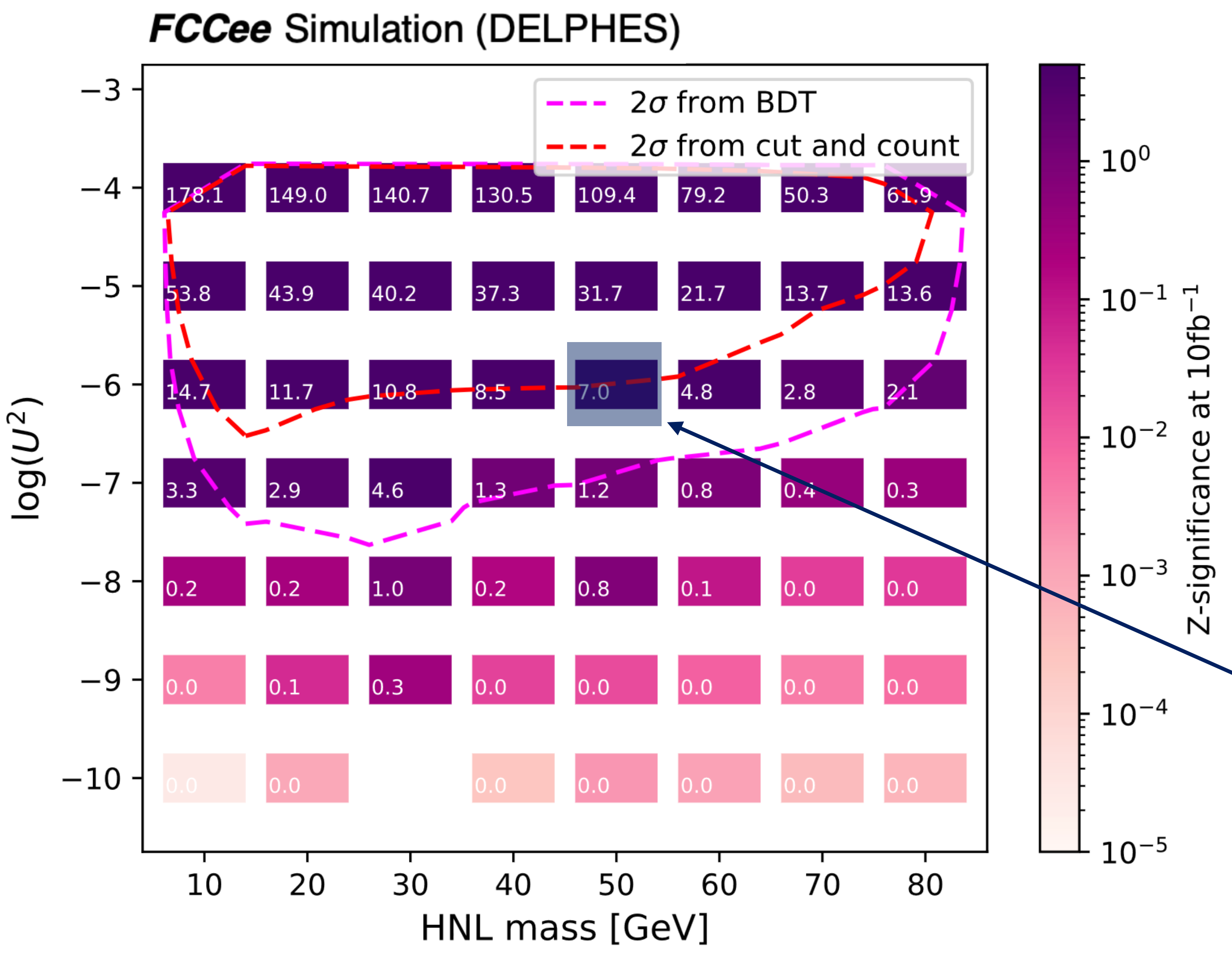
Process	Training Events	Testing Events
Total Background	5,655,708	11,311,415
20 GeV, $\|U^2\| = 10^{-6}$	26,254	26,254
50 GeV, $\|U^2\| = 10^{-6}$	29,991	29,991
70 GeV, $\|U^2\| = 10^{-6}$	32,194	32,193

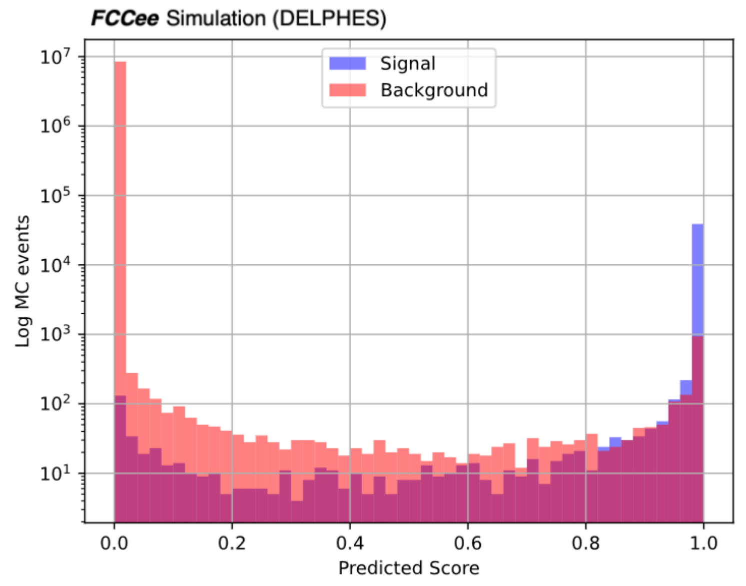
Training / testing split statistics

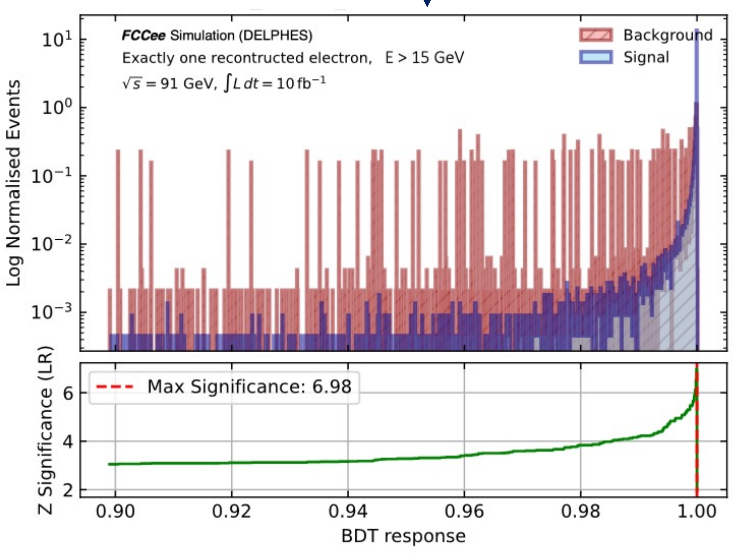
Example raw BDT classification scores for 70 GeV $|U|^2 = 10^{-6}$ mass point



BDT Result







For the 50 GeV mass point at $|U|^2 = 10^{-6}$ at 10 fb^{-1} - BDT cut of 0.999 gives 13.5 signal events and 1.12 background events

DNN workflow

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

Hyperparameter random search for DNN model, including fixed metrics such as the Adam Optimizer

Data Preparation:
- Stage 1 flat ntuples loaded
- Filter applied: E > 20 GeV
- Training and testing split
- Data sets saved in numpy
- Feature flattening

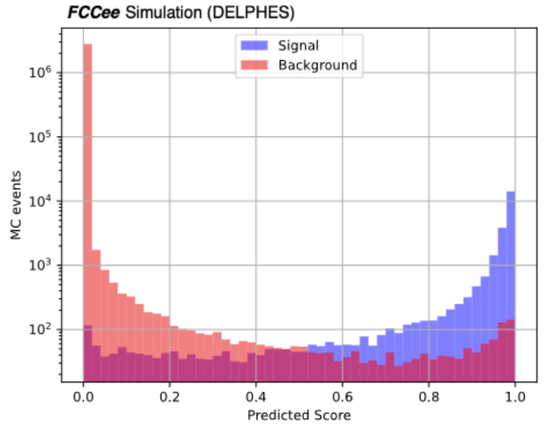
Model training:
- Random search
- Model trained for 100 epochs
- Target: minimise validation loss

Model predictions:
- Models applied to test
- Normalisation to 10 fb^{-1}
- DNN cut chosen based on optimal significance (as with cut and count)

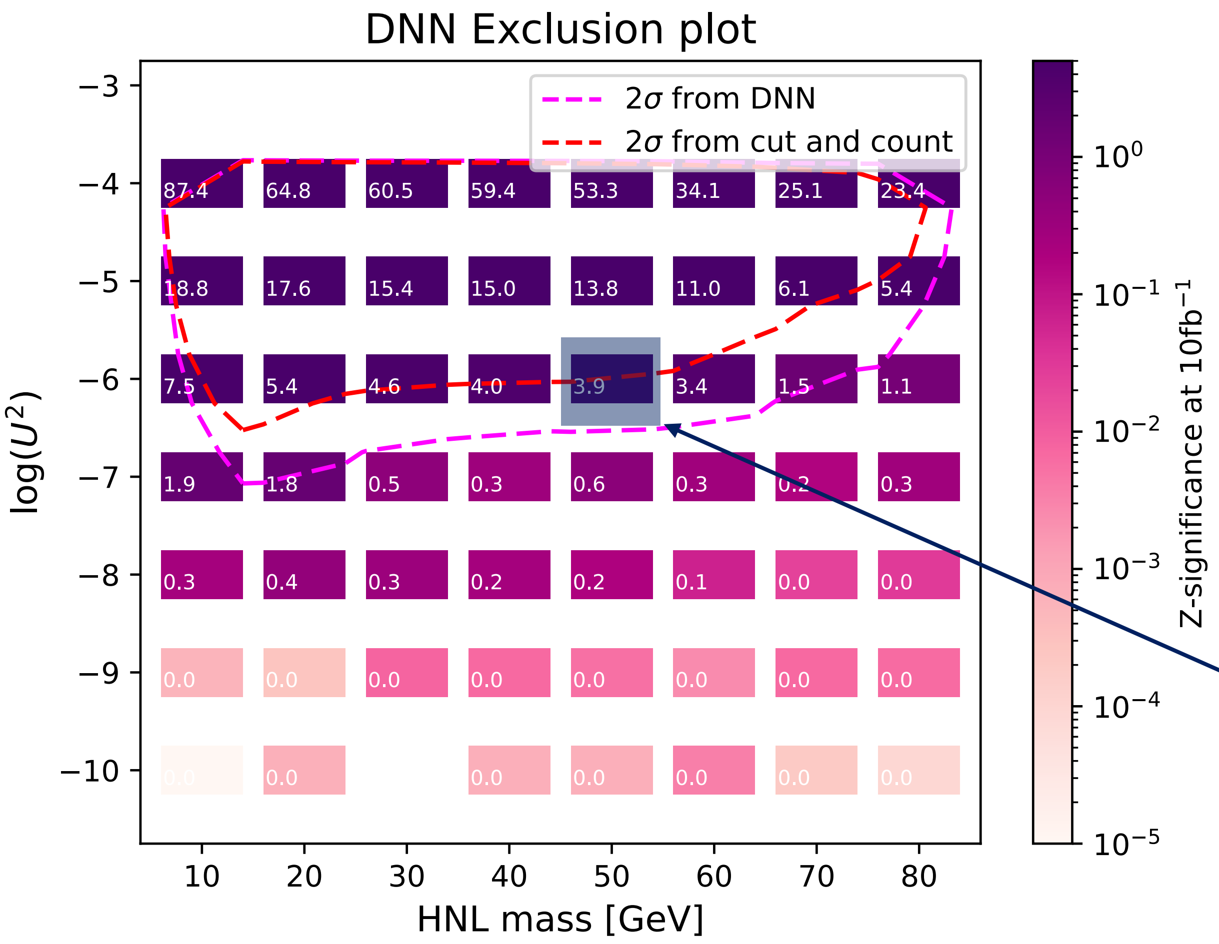
Process	Training Events	Testing Events
Total Background	2,792,099	2,792,099
20 GeV, $\lvert U^2 \rvert = 10^{-6}$	19,601	19,600
50 GeV, $\lvert U^2 \rvert = 10^{-6}$	21,471	21,471
70 GeV, $\lvert U^2 \rvert = 10^{-6}$	23,951	23,951

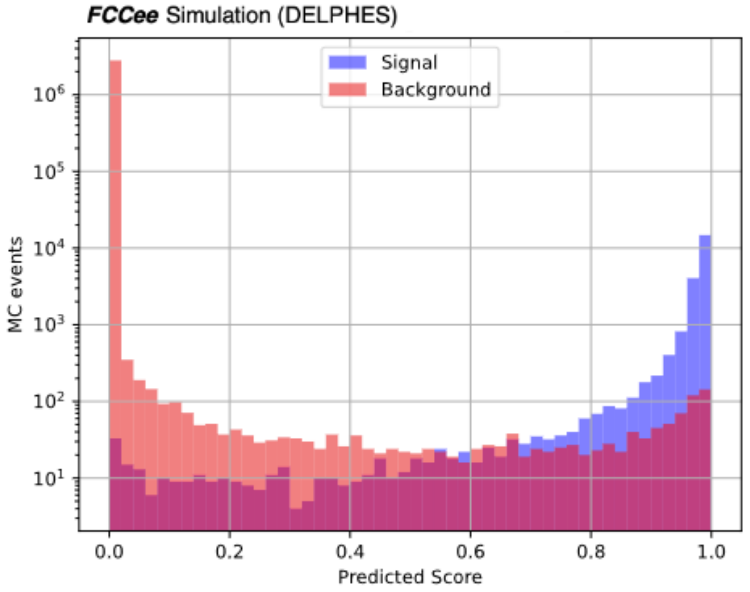
Training / testing split statistics

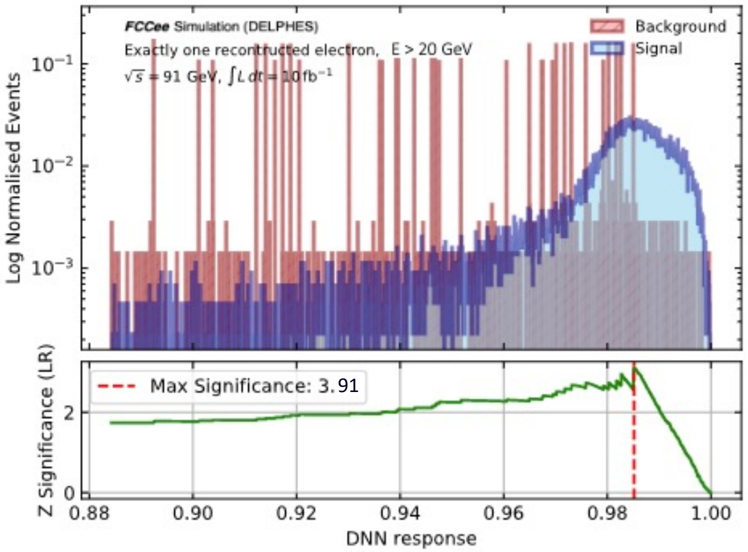
Example raw DNN classification scores for 70 GeV $|U|^2 = 10^{-6}$ mass point



DNN Result

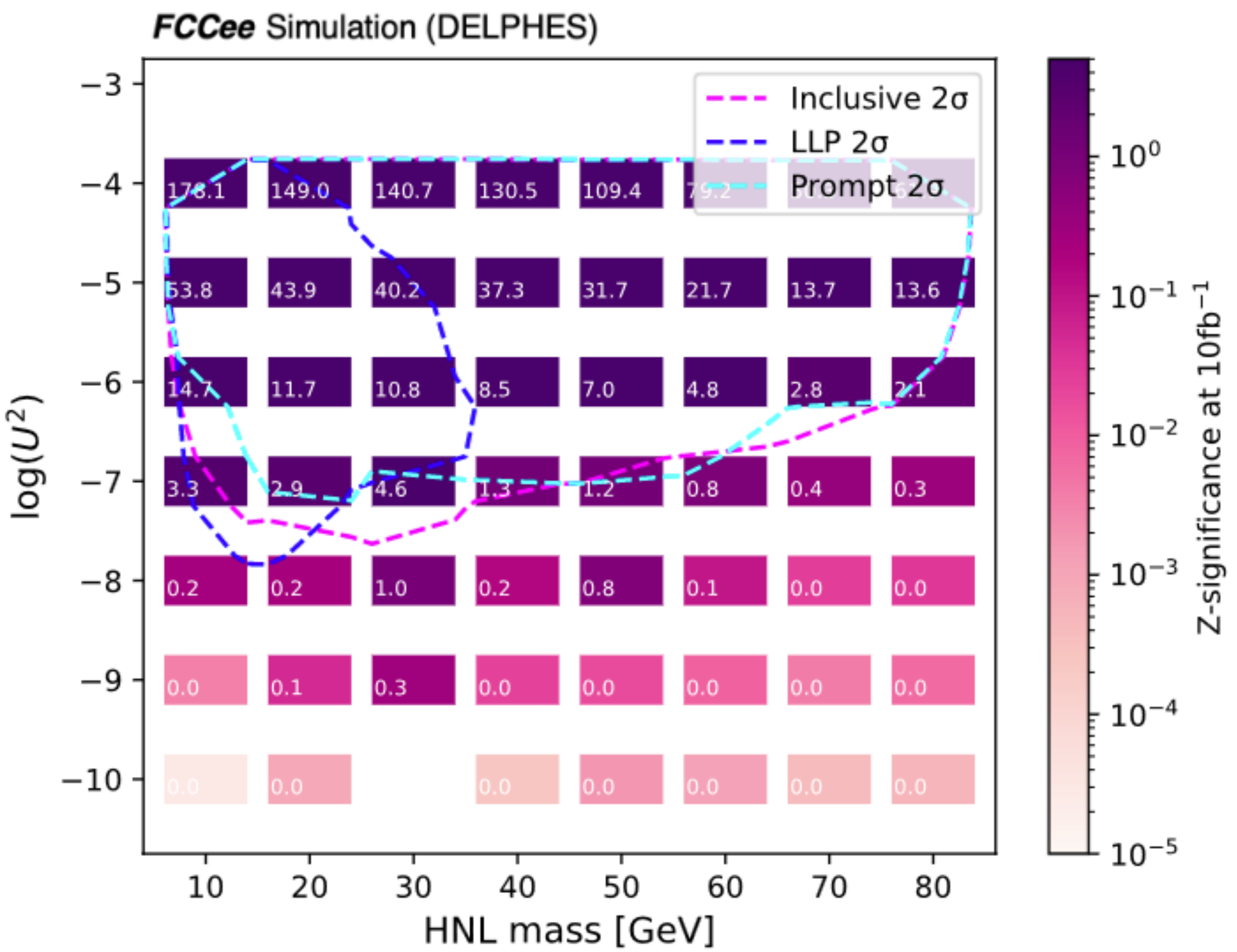




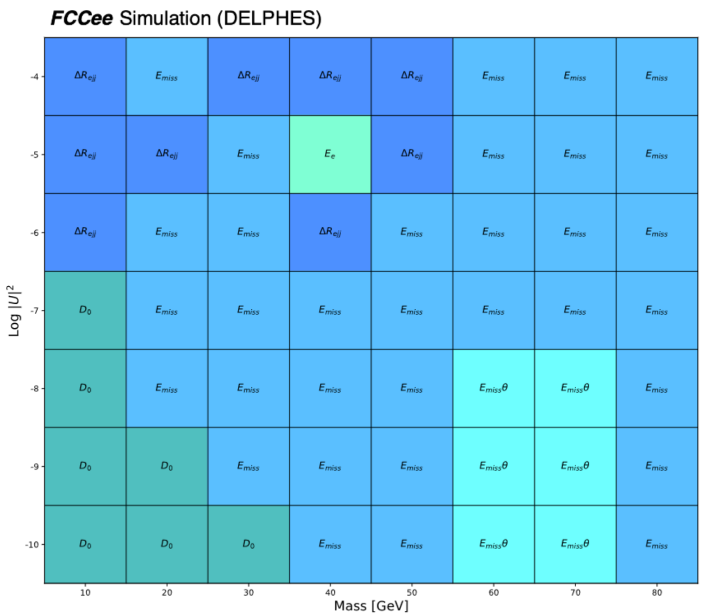


For the 50 GeV mass point at $|U|^2 = 10^{-6}$ at 10 fb^{-1} - DNN cut of 0.986 gives 13.5 signal events and 1.12 background events

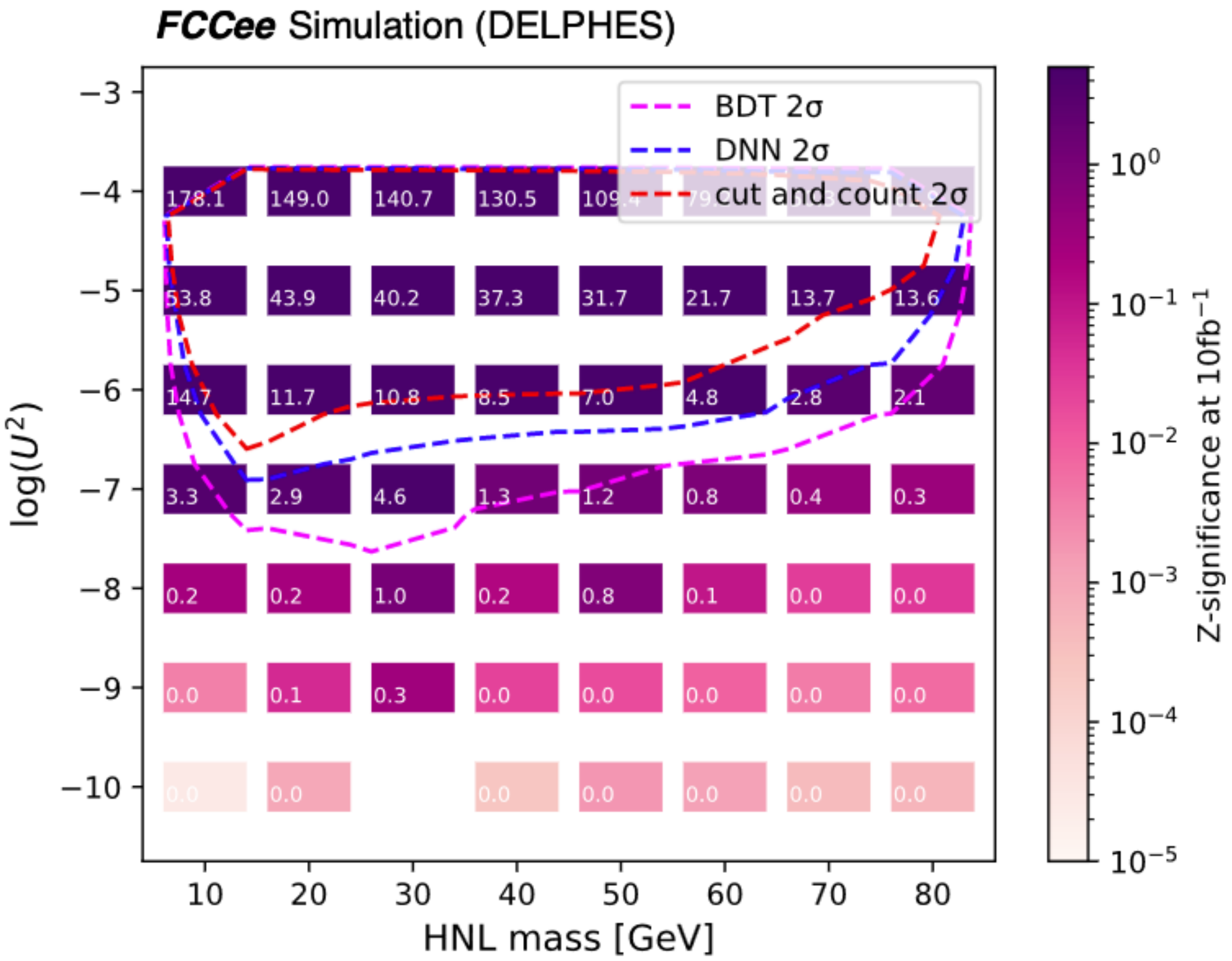
LLP study with BDTs



- Attempted to separate the signal using filter involving the impact parameter significance
- Prompt decays are targetting using $\sigma_{d_0} < 5$ (cyan) and LLPs are targeted for $\sigma_{d_0} > 5$ (blue)
- We find very little improvement (if any) likely because the BDT already uses d_0 as a the most important variable for the LLPs

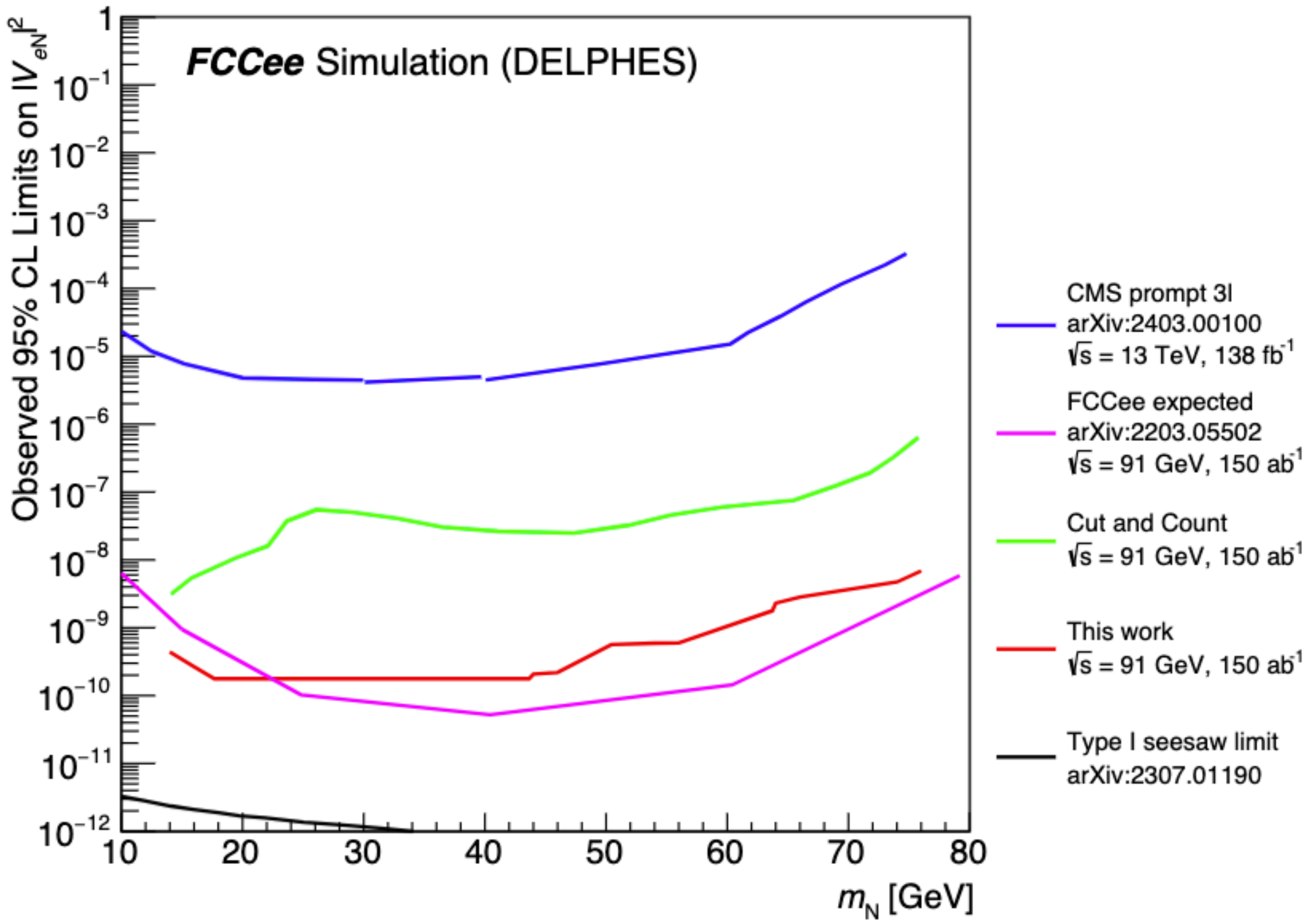


Comparing the strategies



- BDT models so far elicit almost 2 orders of magnitude more delimitation in the couplings compared to the cut and count at the biggest difference, DNN gives ~ 1 order of magnitude.
- Study not yet robust enough to truly claim that the BDT is “better” but instead we can say that it does require much less optimisation to yield great results
- More work on hyperparameter optimisation, feature engineering etc being done on for the DNN until the submission of my thesis – so still some time to improve this result!

Conclusion: our study in context



- Scaling to 150 ab^{-1}, with no estimation on statistical or systematic uncertainties; hence, we can only interpret the plot on the left in terms of how it compares to the cut and count, and we see that it indeed delimits a much broader region of the phase space.
- We begin to crest upon the projected FCC-ee limit, despite working with only ~50% of the branching ratio – though, as said – this should be interpreted only as a guide for improvement strategies since we do not have the associated uncertainties
- Nevertheless, ML seems to be capable of hugely improving our limits, possibly across all final states!
- **Increasing MC statistics** in signal region **crucial** for robust studies

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Thank you for listening! I would be happy to take any questions

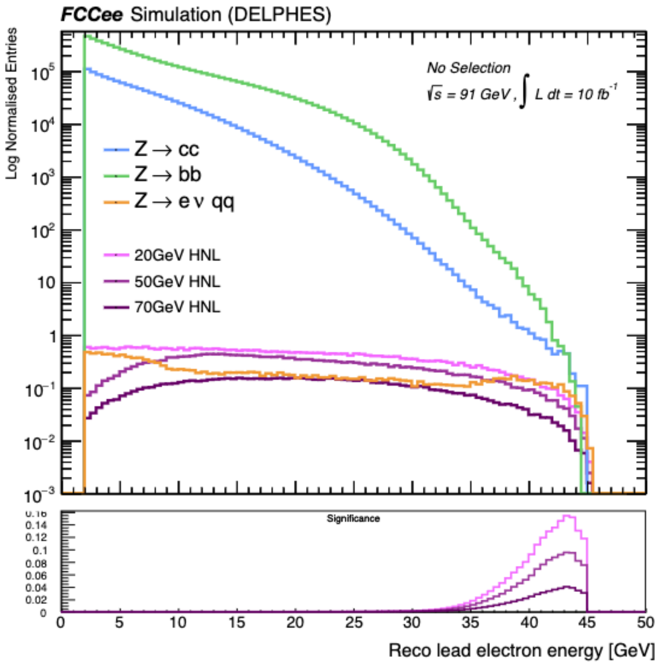
FUTURE CIRCULAR COLLIDER

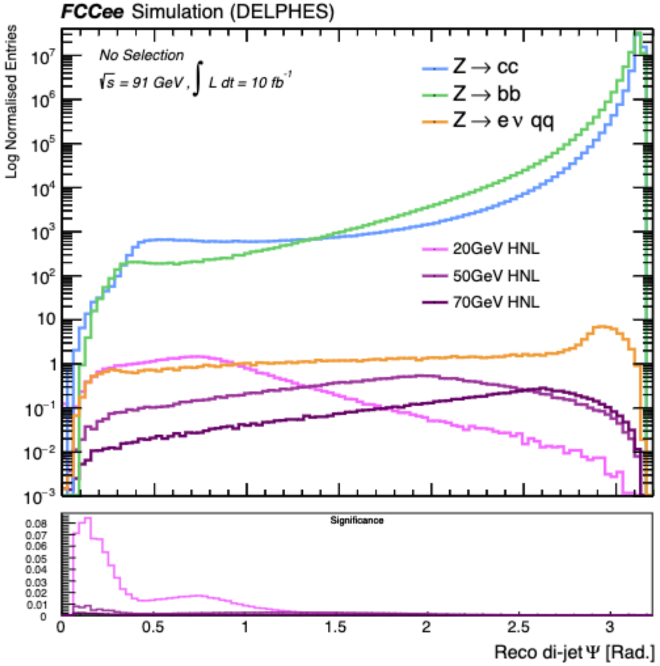
UNIVERSITÉ DE GENÈVE

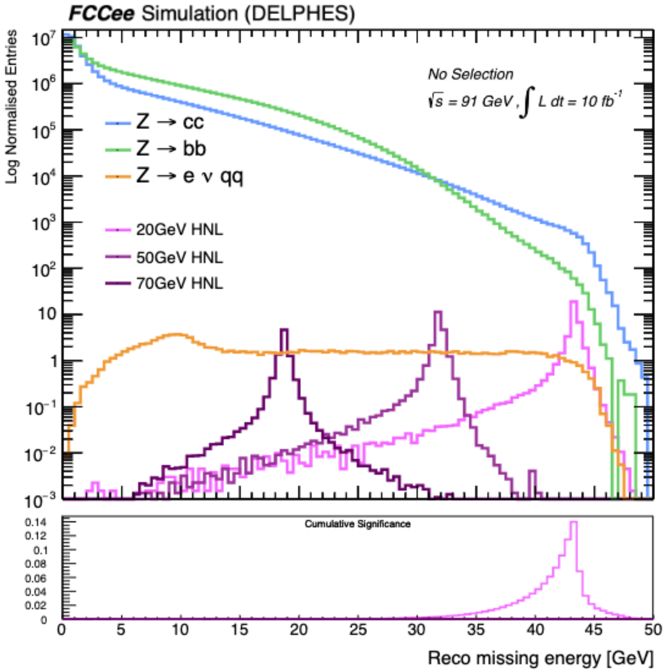
FACULTÉ DES SCIENCES

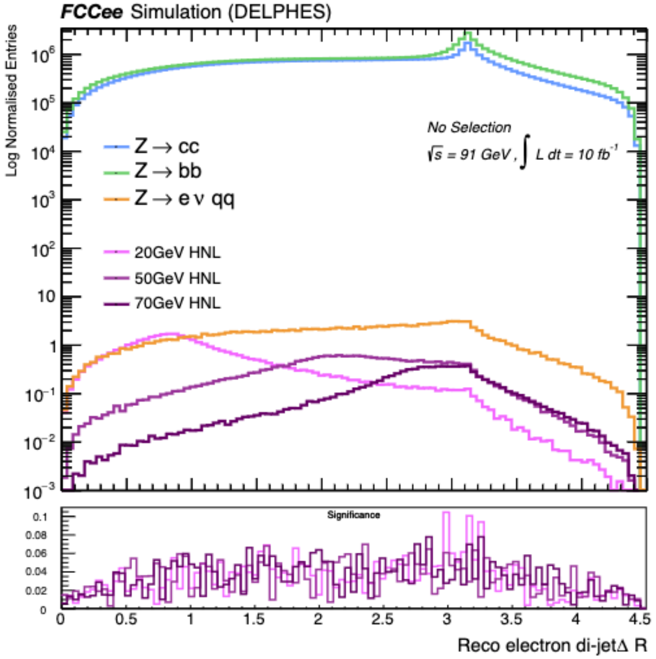
-- Additional slides --

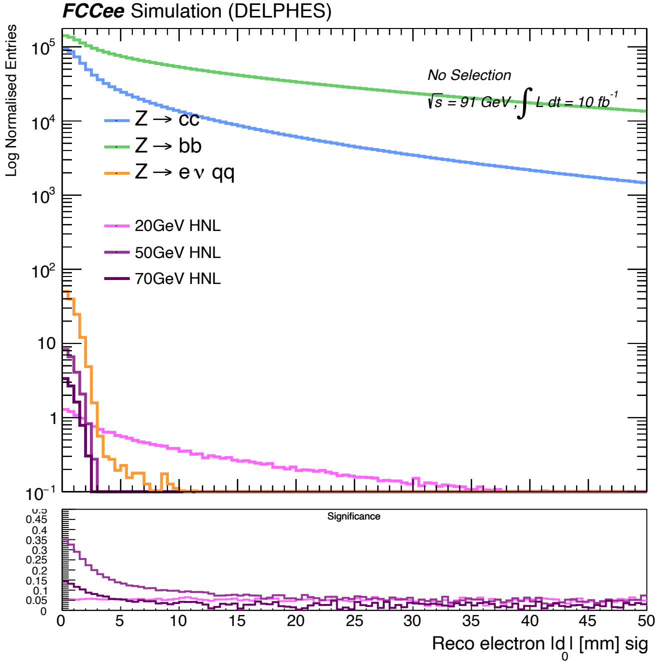
FUTURE CIRCULAR COLLIDER


FCCee Simulation (DELPHES)
No Selection
$\sqrt{s}$ = 91 GeV , $\int$ L dt = 10 fb^{-1}
Z → cc
Z → bb
Z → e ν qq
20GeV HNL
50GeV HNL
70GeV HNL
Log Normalised Entries
Significance
Reco lead electron energy [GeV]


FCCee Simulation (DELPHES)
No Selection
$\sqrt{s}$ = 91 GeV , $\int$ L dt = 10 fb^{-1}
Z → cc
Z → bb
Z → e ν qq
20GeV HNL
50GeV HNL
70GeV HNL
Log Normalised Entries
Significance
Reco di-jet Ψ [Rad.]


FCCee Simulation (DELPHES)
No Selection
$\sqrt{s}$ = 91 GeV , $\int$ L dt = 10 fb^{-1}
Z → cc
Z → bb
Z → e ν qq
20GeV HNL
50GeV HNL
70GeV HNL
Log Normalised Entries
Cumulative Significance
Reco missing energy [GeV]


FCCee Simulation (DELPHES)
No Selection
$\sqrt{s}$ = 91 GeV , $\int$ L dt = 10 fb^{-1}
Z → cc
Z → bb
Z → e ν qq
20GeV HNL
50GeV HNL
70GeV HNL
Log Normalised Entries
Significance
Reco electron di-jetΔ R


FCCee Simulation (DELPHES)
No Selection
$\sqrt{s}$ = 91 GeV , $\int$ L dt = 10 fb^{-1}
Z → cc
Z → bb
Z → e ν qq
20GeV HNL
50GeV HNL
70GeV HNL
Log Normalised Entries
Significance
Reco electron |d$_0$| [mm] sig

DNN vs BDT feature importance