Search for Long-Lived Particles with HCAL Segmentation in CMS at the Large Hadron Collider

Katia Avanesov
Mentor: Harvey Newman
Co-mentor: Kiley Kennedy

1 Introduction

Long-Lived Particles (LLPs), often classified as particles with lifetimes greater than 10^{-10} seconds, feature in both the Standard Model (SM) – such as charged pions, neutrons, electrons, and protons – as well as in many Beyond the Standard Model (BSM) theories, such as those relating to dark matter, matter anti-matter asymmetry, and supersymmetry [1]. In this paper, we study the process, $H \to XX \to b\bar{b}b\bar{b}$, in which it is hypothesized that a Higgs boson first decays to an unknown long-lived particle (LLP), X, which in turns decays to the commonly observed final state consisting of pairs of bottom and anti-bottom quarks. The process is illustrated below in Figure 1. Our analysis will also provide insight into heavy twin Higgs models [2]: given that the masses of the Higgs and LLP are left as free parameters within our model, we may scan across various mass combinations, including heavier Higgs masses. The Higgs itself has been a focal point of study at the LHC ever since its discovery in 2012 [3], so the possibility of its coupling to an exotic particle, or the existence of its heavier twin, offers exciting potential for the discovery of New Physics.

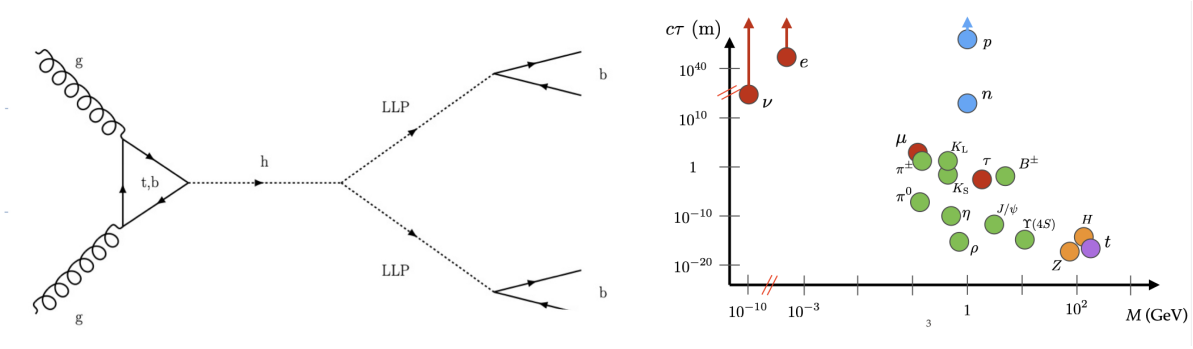

Figure 1: [Left] Feynman diagram showing the formation of a Higgs boson via gluon-gluon fusion and its subsequent hypothesized decay into a pair of LLPs which in turn decay to $b\bar{b}$ pairs. [Right] Plot of SM particles' lifetimes (τ) in units of $c\tau$ as a function of their mass in GeV. [4]

Many of the particles that CMS physicists have been interested in decay almost instantaneously after they are produced, such as the Higgs itself with a lifetime on the order of 10^{-22} seconds. A short-lived particle would therefore decay effectively at the center of the proton-proton interaction point, subsequently producing many decay products such as leptons, photons, and hadrons. The subdetectors of CMS experiment, shown in Figure 2, are optimized to detect the many signatures that the original particle leaves behind. For instance, the CMS silicon-based inner tracker records the trajectory of charged particles through the readout of electrical signals, while the electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) record the energies of electrons and photons, and hadronic particles, respectively. The information from these many sub-detectors is then used to reconstruct the masses, momenta, and charges of the decay products, from which the identity of the original particle can then be deduced.

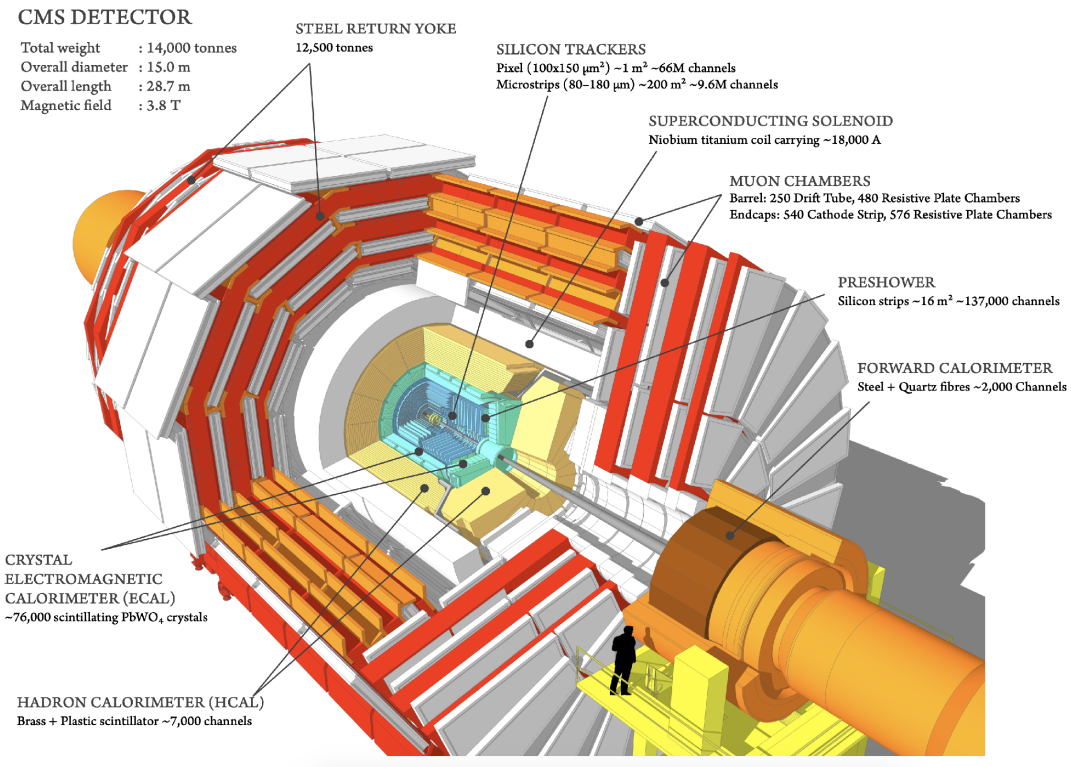

Figure 2: Diagram to show the components of the CMS detector [5].

By contrast, LLPs such as those hypothesized in the $H \to XX \to b\bar{b}b\bar{b}$ interaction could travel up to several meters before it decays, and thereby leave relatively fewer signatures within the detector, especially within its innermost parts such as the inner tracker or the ECAL. However, we can think of this absence of signal as a signature in and of itself. Indeed, this is the aim of the joint effort of the Caltech and Princeton group in their analysis of searching for Long-Lived Particles: to exploit their distinct topology, or lack-thereof.

One principle of our search strategy lies in selecting events with LLPs decaying within the HCAL because their signatures look very different compared to our main sources of background, such as QCD, $Z+$ and $W+$ jets, which form 'prompt' jets, as opposed to delayed ones. These differences are illustrated below in Figure 3. Firstly, we observe that a higher fraction of delayed jets' energy would be deposited in the outermost layers of the HCAL, since we require that the LLP decays inside of it. We expect a more even distribution of jet energy fraction from the prompt jets, if not more being deposited within the inner layers, given that the HCAL is made of very dense material and so jets produced outside of the HCAL are less likely to travel as far inside of it and deposit its energy far into its depths. Similarly, due to the dense HCAL material, we also see that the jets produced from with the HCAL are less 'spread out' compared to jets produced immediately within the LHC beam, as a result of their displaced vertices.

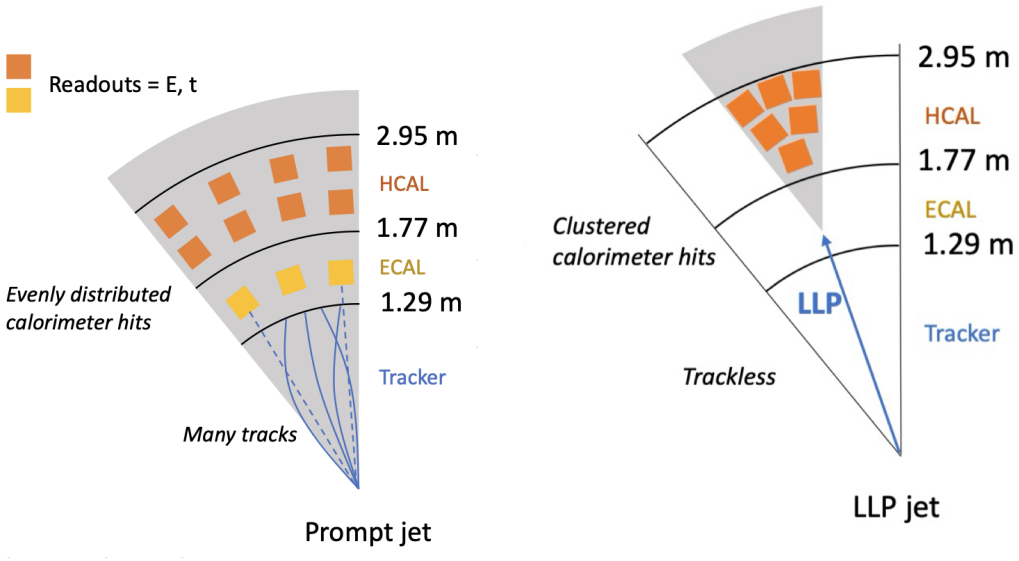

Figure 3: Illustration of a cross-section of the CMS detector to show the difference in signatures left by a prompt jet (left) and a delayed LLP jet (right). [4].

2 Previous Work

Last year, in collaboration with the Caltech and Princeton Group, led by Kiley Kennedy, I developed a range of Machine Learning-based classifiers for LLP jet-tagging, representing a significant innovation for the team, which had previously relied only classical classification methods such as boosted decision trees. The implementation of more advanced classification techniques is necessary for our analysis due to the minute cross-section of the Higgs decay at a mass of 125 GeV and at a total center of mass energy, $\sqrt{s} = 13.6$TeV, namely $59.2^{+5\%}_{-7\%}$ pb [6]. With an integrated luminosity of 160 fb^{-1} reported so far for Run 3 in September 2024, [7], we expect to find on the order of just 10 million Higgs decay events across the entirety of the Run 3 dataset. Thus, within our dataset of 100 million events that passed our L1 trigger, we would expect to be left with even fewer signal events. Moreover, the number of Higgs decaying to LLPs is further limited by the 60% branching ratio of the bottom quark decay channel, [6]. Therefore, while it is relatively straightforward to understand the main differences between LLP and prompt jets, the challenge lies in producing a classifier that is sufficiently powerful enough to select the signal with high efficiency while maintaining the lowest possible efficiency for the background, on the order of 10^{-6}.

To illustrate this challenge, consider the Receiver-Operating Characteristic (ROC) curve below, showing the trade-off between signal selection (True Positive Rate) and background selection (False Positive Rate) for various thresholds of a classifier's score. In Figure 4, the score is simply that of the value of the Jet Neutral Hadron fraction, which is one of our most discriminating variables between LLPs and $W+$ jets, one source of background. We see that while this ROC curve already manages to achieve an impressive 0.994 area under the curve (AUC), it exhibits a very poor signal selection at high background rejection (≈ 0.25 at 10^{-4}) . We therefore see that our project warrants the use of more powerful classification methods to enhance the signal and background discrimination.

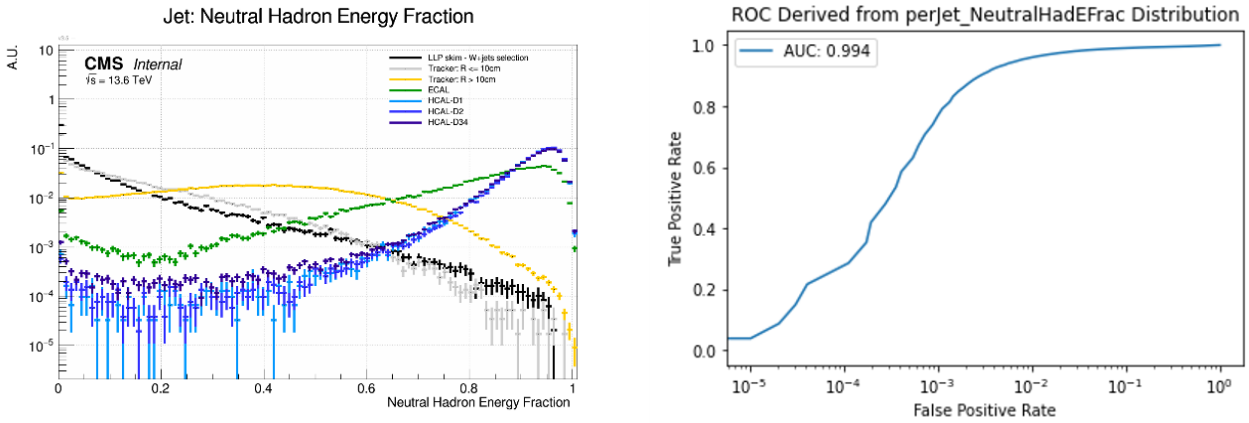

Figure 4: [Left] Distribution of the jet neutral hadron fraction for LLPs decaying within the HCAL (HCALD1, HCALD2, HCALD34 in dark blue, light blue, purples) and background, $W+$ jets (LLP skim, $W+$ jets selection in black) [4] [Right] ROC constructed directly from this distribution.

Over the course of last summer, I developed two neural network classifiers using data from the LHC. First, I trained a Dense Neural Network (DNN) on various jet metrics—such as jet energy and momentum—and achieved a signal efficiency of 60% at a background efficiency of just 0.01%. I also trained a Convolutional Neural Network (CNN) on three-dimensional "images" of jet energy deposits in the HCAL. In tests on a smaller dataset, this CNN reached up to 80% signal efficiency at the same 0.01% background efficiency. When evaluated on the full dataset, the CNN maintained over 40% signal efficiency at the same background rejection rate. These classifiers mark significant improvements on the results yielded by the boosted decision trees. Below, we present a series of ROC curves to summarize the progress made last year, and demonstrating the comparison between the models.

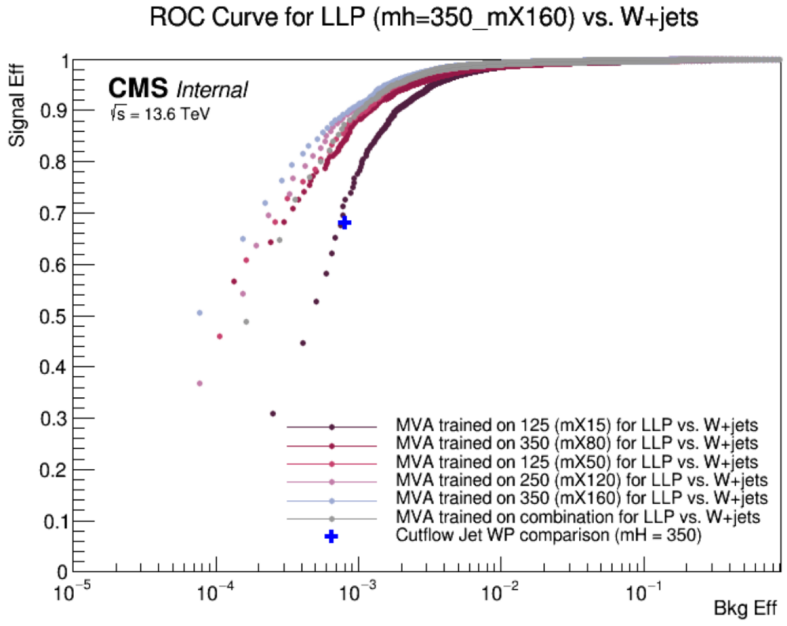

Figure 5: ROC curves displaying the performance of TMVA's Gradient-Boosted Decision Tree binary classifier, trained on the MH350-MS160 sample and evaluated on all other mass samples.

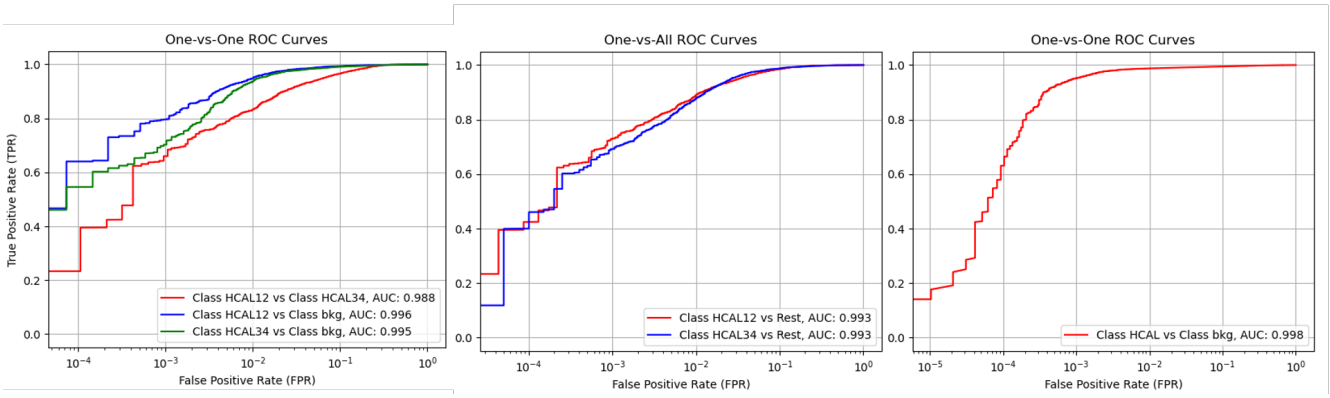

Figure 6: Series of ROC curves displaying the performance of Dense Neural Network implemented in Tensorflow's Keras and trained on the MH350-MS160 dataset. [Left] ROC curves shown for how well the model distinguishes between the two classes. [Center] ROC curves showing how well the model distinguishes one signal class from the rest. [Right] ROC curve showing the performance of the model as a binary classifier, where the two signal classes are combined into one.

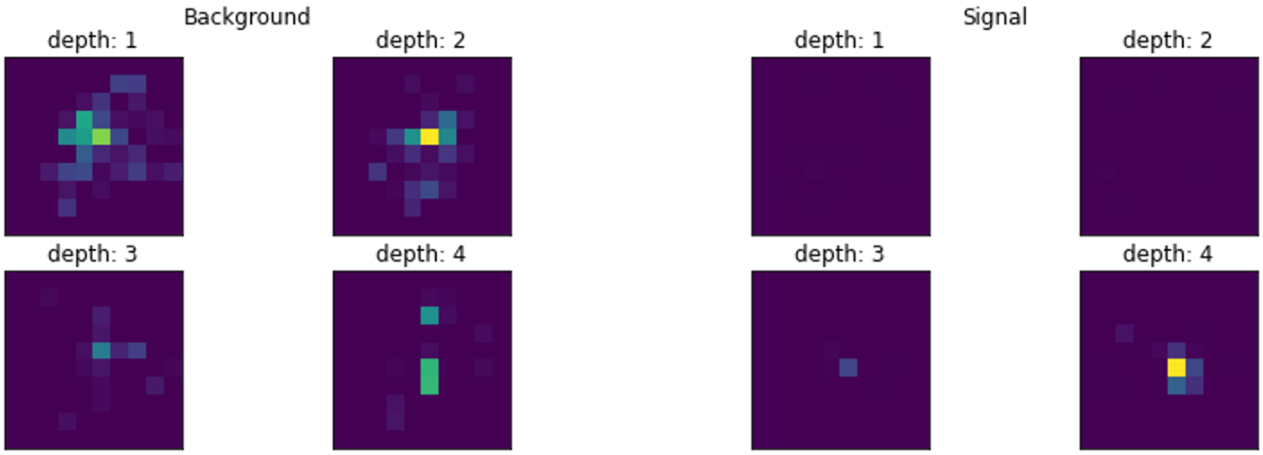

Figure 7: 3D jet 'image' samples of a background (left) and signal (right) jet. Each image is normalized to preserve the relative value of energy in each depth. These images align closely with what we would expect from our previous illustration of the differences of the LLP and prompt jets: the background has a more widespread distribution of energy across the depths of the HCAL, whereas the signal has a very localized energy deposit at the outermost layer of the HCAL.

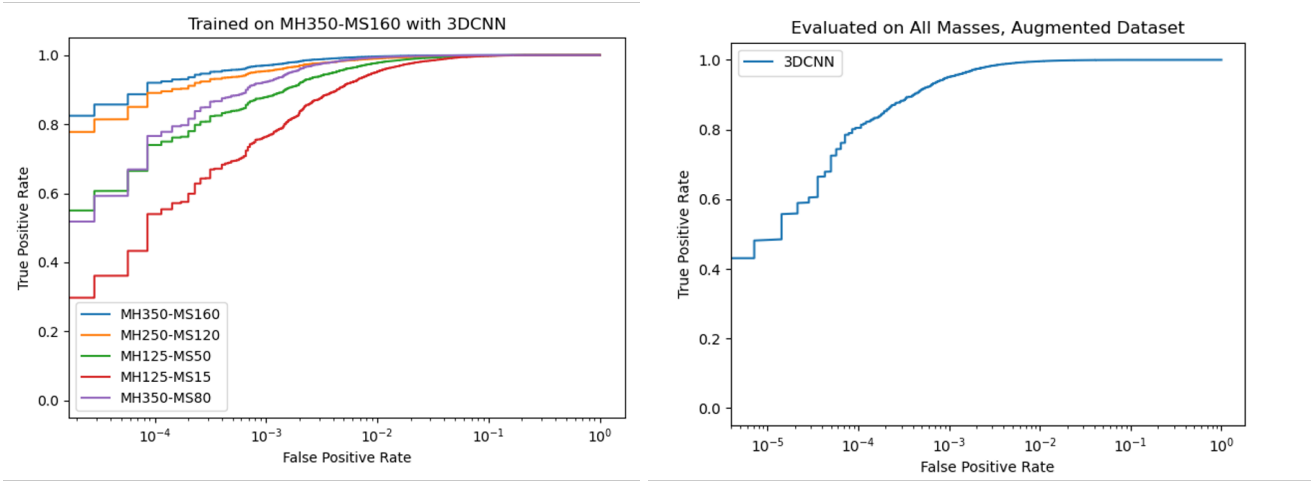

Figure 8: ROC curves showing the performance of the 3DCNN model. [Left] Mass comparison performance. [Right] Combined mass evaluation with augmentation of dataset.

With the development of these high-performing classifiers, I became increasingly interested in understanding which features the networks were using. This was essential to verify their reliability, ensuring that they captured meaningful characteristics of the jets rather than insignificant artifacts that might fail when applied to real CMS detector data (as opposed to the Monte Carlo simulations they were trained on).

To analyze the DNN and the Decision Tree, I extracted the variable importance of their inputs and observed a strong alignment in the features each model found most critical, providing further validation. For the CNN, I employed Gradient-Class Activation Mapping (GradCAM), which highlighted the regions of the jet energy deposit images that the model deemed most significant. These visualizations revealed that the CNN was focusing on energy bursts and paying attention to the distinction between clustered and widespread distributions—precisely what we would expect, given that this difference is key in distinguishing LLP jets from background jets.

3 My Project

I plan to build upon my previous work by integrating the 3DCNN into the analysis framework, which is based on an "asymmetric" or "double-tag" strategy. In this approach, we use two separate classifiers to analyze the two LLP jets produced within a single Higgs decay event (as illustrated in the Feynman diagram in Figure 1). Firstly, we choose to only select events which have at least one jet decaying within the HCAL. That jet is then evaluated by our 3DCNN ("depth tagger"), outputting a score; using some fixed score threshold, we can classify the jet as signal or background.

The other jet, which is free to decay anywhere within the detector, is evaluated by our DNN that is trained on features such as tracking information from the inner detector and the ECAL. Each event therefore gets two scores, one from the 3DCNN and one from the DNN. We can visualize these two outputs on a 2D plot, as shown in the Figure 9 below.

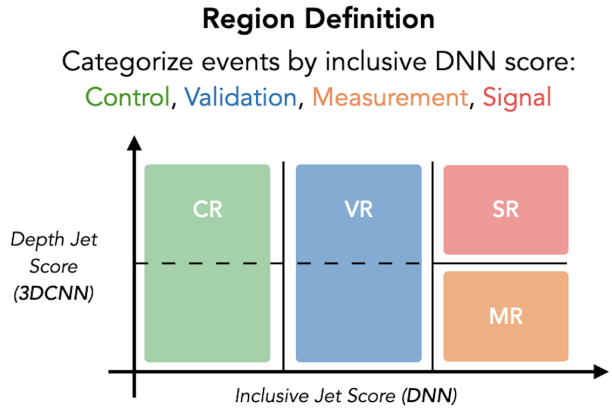

Figure 9: A visualization of the plot of the inclusive jet score output versus the 3DCNN depth tagger output. The dashed horizontal line represents a chosen fixed score threshold above which we treat events as signal, and below which, as background. CR: Control Regions. VR: Validation Region. MR: Measurement Region. SR: Signal Region.

In the simplest case, we assume that the 3DCNN and DNN scores are uncorrelated— this is reasonable if the jets truly behave as independent objects. This then implies that within any specific bin of DNN scores, the fraction of events classified as signal or background by the 3DCNN remains constant. We can validate this assumption by examining the control (CR) and validation (VR) regions (see the figure above): we check for the proportion of signal and background (base don the 3DCNN score threshold). If the proportions within the CR and VR bins are consistent, we can extrapolate this consistency into the signal region (SR), allowing us to predict how many events should appear there, based on the number of events within the measurement regions (MR). Comparing the actual number of events in the SR to this prediction then informs us whether there is a significant discrepancy that might indicate new physics.

However, several complexities need to be addressed:

Score Thresholds: A score threshold needs to be chosen such that jets with scores above this threshold are treated as signal. This is why the studies from last year are essential because they inform us about the proportion of background rejection and signal selection at a given threshold.

Varying Mis-Tag Rates: In reality, we observe that the mis-tag rate varies with the physical parameters of the jets, such as its momentum, ϕ, etc. Therefore, it is necessary to study how the mis-tag rate varies as a function of each parameter.

Thresholds and Binning: We also need to choose appropriate thresholds for the DNN and decide how to define our control (CR), validation (VR), and measurement (MR) regions.

Uncertainties due to Neural Networks: Finally, we must account for and quantify the uncertainties introduced by the neural network models themselves due to their own mis-tag rates (i.e., assigning a high score to a background jet, given that background rejection will not be 100% at any reasonable threshold we choose).

These are the key challenges I aim to tackle this summer, with the ultimate goal of refining our double-tag strategy and preparing the analysis for a final measurement.

4 Project Timeline

Pre-SURF Work as a part of Dr Kennedy's group during the Winter and Spring term. I am currently optimizing the 3DCNN network performance, and testing its stability during training.

Weeks 1-3 Set up workflow for the asymmetric double-tag strategy and obtain some baseline results with some nominally chosen parameters.

Weeks 4-6 Perform in-depth studies on the variation of the mis-tag rate with the physical parameters of the jets.

Weeks 6-8 Optimize the search strategy by experimenting with setting a variety of thresholds and bins to explore the full parameter space of this model.

Weeks 9-10 Formalize final results and work on final paper.

References

[1] C. et al., "Long-lived particles at the energy frontier: the mathusla physics case," *Reports on Progress in Physics*, vol. 82, no. 11, p. 116201, Oct. 2019. [Online]. Available: http://dx.doi.org/10.1088/1361-6633/ab28d6

[2] Z. Chacko, H.-S. Goh, and R. Harnik, "Natural electroweak breaking from a mirror symmetry," *Phys. Rev. Lett.*, vol. 96, p. 231802, Jun 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.96.231802

[3] S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[4] G. Kopp, "Searching for llps in run 3 with a dedicated hcal timing trigger," Presentation at FNAL Pileup, February 2024. [Online]. Available: https://indico.cern.ch/event/1384246/#5-searching-for-llps-in-run-3

[5] "Cms detector introduction - cms open data workshop 2023," CMS Open Data Workshop, 2023. [Online]. Available: https://cms-opendata-workshop.github.io/workshop2023-lesson-cms-detector/01-introduction/index.html

[6] S. Navas *et al.*, "Review of particle physics," *Phys. Rev. D*, vol. 110, no. 3, p. 030001, 2024.

[7] CERN. (2023) Accelerator report: Lhc run 3 achieves record-breaking integrated luminosity. Accessed: 2024-09-21. [Online]. Available: https://www.home.cern/news/news/accelerators/accelerator-report-lhc-run-3-achieves-record-breaking-integrated-luminosity