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

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

Abstract

We have searched for Long-Lived Particles (LLPs) with lifetimes greater than 0.1 nanoseconds produced at high energies in the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider, which are predicted by both the Standard Model (SM) and many Beyond the Standard Model (BSM) theories, including those addressing dark matter, the matter-antimatter asymmetry in the universe, and supersymmetry. The search uses Higgs boson decays via a massive LLP pair to a final state with two bottom quark-antiquark pairs: $H \to XX \to b\bar{b}b\bar{b}$. The analysis exploits the new depth segmentation upgrade of the CMS Hadronic Calorimeter (HCAL) to target the distinct topology of a jet-pair from LLPs that decays within the HCAL. Using Monte Carlo-generated signal samples along with W+ jet background samples, we develop and compare the performances of various neural network (NN)-based classifiers. Among the explored architectures, we produce 3D convolutional neural networks (CNNs) trained on 'images' of the jet energy distribution across the HCAL depth-layers and in eta-phi space within each layer, which yield up to 0.9 signal efficiency at just 10^{-4} background (false positive) efficiency. These findings have motivated further investigations into model interpretability, through the application of Gradient-weighted Class Activation Mapping (GradCAM) to visualise the importance of topological features and gain deeper insight into the model's learning process.

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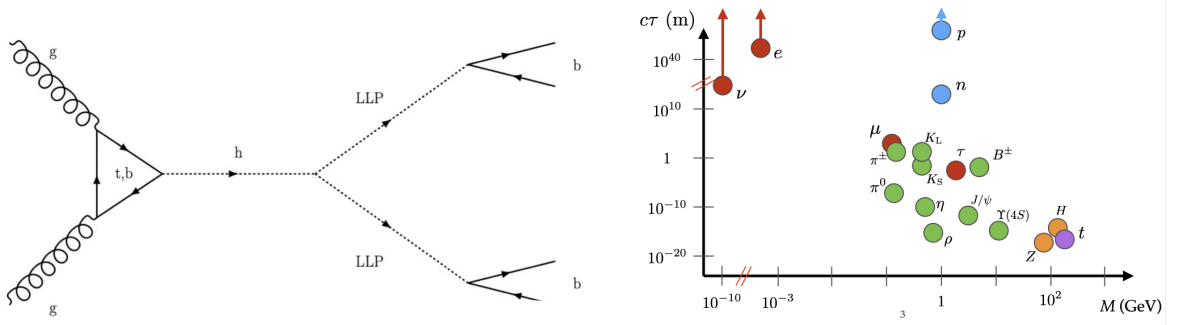
Figure 1: [Left] Feynman diagram showing the formation of a Higgs boson via gluon-gluon fusion and its subsequent hypothesized decay into a pair of LLPs which in turn decay to $b\bar{b}$ pairs. [Right] Plot of SM particles' lifetimes (τ) in units of $c\tau$ as a function of their mass in GeV. [4]

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.
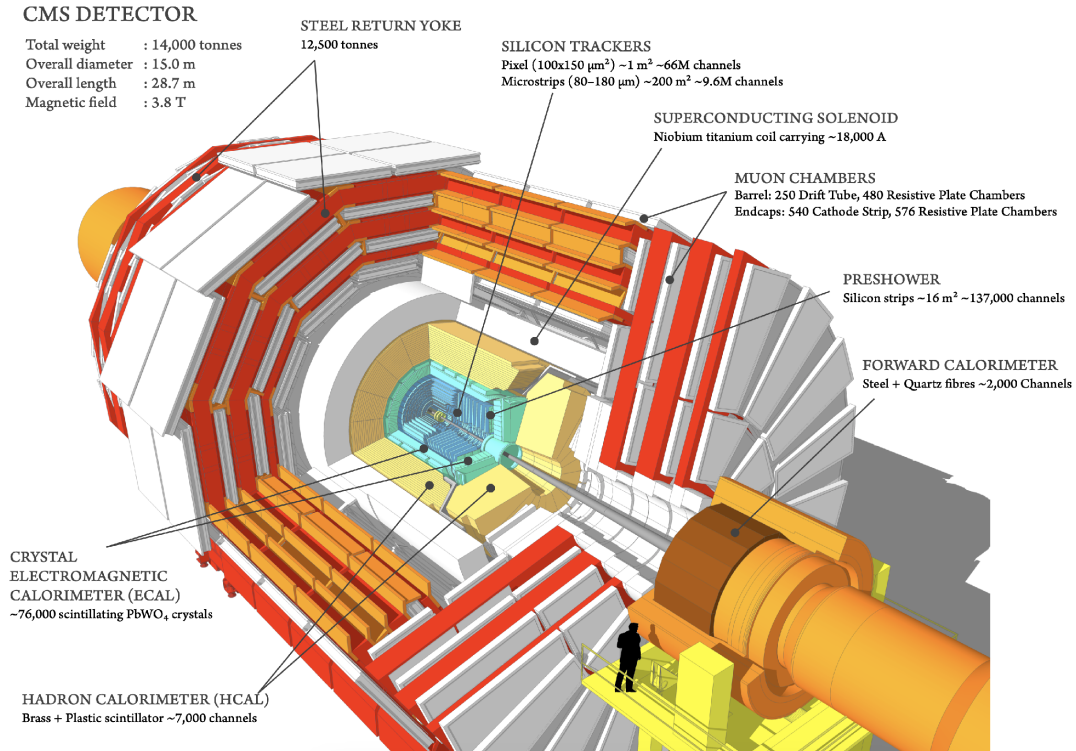

Figure 2: Diagram to show the components of the CMS detector [5].

By contrast, LLPs such as those hypothesized in the $H \rightarrow XX \rightarrow b\bar{b}b\bar{b}$ 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

2

signature in and of itself. By narrowing our search down to LLPs that decay in the HCAL, we end up with a very distinct target jet topology, compared to much of what our main sources of background would look like, 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.
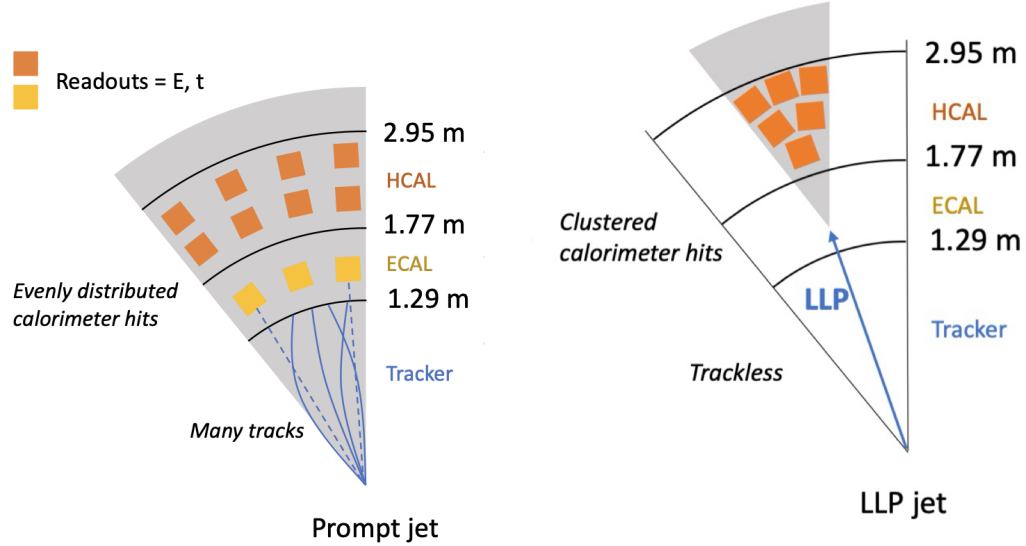

Figure 3: Illustration of a cross-section of the CMS detector to show the difference in signatures left by a prompt jet (left) and a delayed LLP jet (right). [4].

A novel Level 1 Trigger has been designed specifically for LLP selection. Due to the sheer number of collisions, occurring at a rate of billions of proton-proton collisions per second, it is not possible to store all the events in memory. Therefore, a series of triggers are developed to determine the selection of events that are kept, and which are discarded. The Level 1 trigger is the first of such triggers, which processes events at a rate of 40MHz and reduces the output of information to 100kHz.

The LLP trigger is composed of several metrics. Firstly, it triggers on jets which deposit a higher energy fraction in the outermost layers of the HCAL. To measure this, one takes advantage of the recent detector upgrade which introduced depth segmentation in the HCAL. Namely, the HCAL has now been divided into four distinct layers, represented by the different colors on the HCAL barrel in Figure 4 below, where the trigger selects jets with high energy fractions in the third and fourth outermost layers.

Secondly, the trigger also measures the time delay between the expected arrival of a jet and its actual arrival. For jets produced by LLPs, we would expect to detect a time delay for two reasons. Firstly, there will be a time delay simply because the LLP will not decay immediately and so the jet will not be formed instantaneously either. Secondly, by the time the LLP will have decayed, it would have traveled a given distance from the center of the proton-proton interaction point. This means that if we were to reconstruct the trajectories of the particles in the jet, then their trajectories would not originate from the center of the beam but would rather be formed at a displaced vertex. As a result, we see that this would create a path difference between a jet which simply traveled straight from the center to the point where energy is deposited in the HCAL versus if it were first originated from a displaced vertex. This path difference would thus contribute its own time delay. As a result, the LLP trigger is also designed to select events with at least one jet which arrived late. The trigger is designed to have three categories: prompt jets are labeled as events with a time delay of less than 6 ns, slightly delayed jets have a delay between 6-7 ns, while delayed jets are those with a delay of greater than 7 ns. The time of arrival of these jets is calculated

relative to the HCAL clock which records relative to a particular bunch crossing. A custom data-taking run was performed to calibrate and align these time measurements.

Finally, the trigger has a minimum threshold jet energy of 40GeV and an implicit pseudorapidity, η cut of 1.4, in the sense that the trigger is designed to only detect events occurring in the barrel of the detector as opposed to the end-cap, see Figure 4 below.

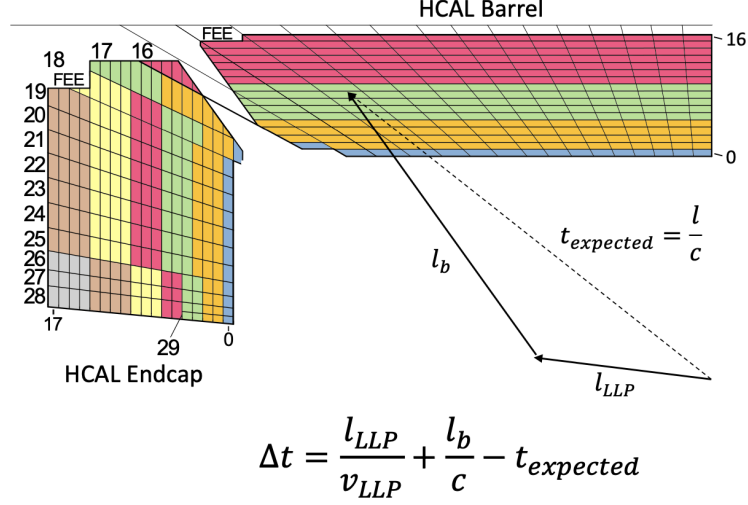$$\Delta t = \frac{l_{LLP}}{v_{LLP}} + \frac{l_b}{c} - t_{expected}$$

Figure 4: Diagram to show the resulting path difference due to the displaced vertex. Further, observe that the four colors denoting the calorimeter at the barrel represent the four layers of segmentation.

With the LLP Trigger in place, our goal now lies in optimizing the offline analysis to further maximise our signal selection, especially at very high values of background rejection. This is essential because of 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 5, 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.
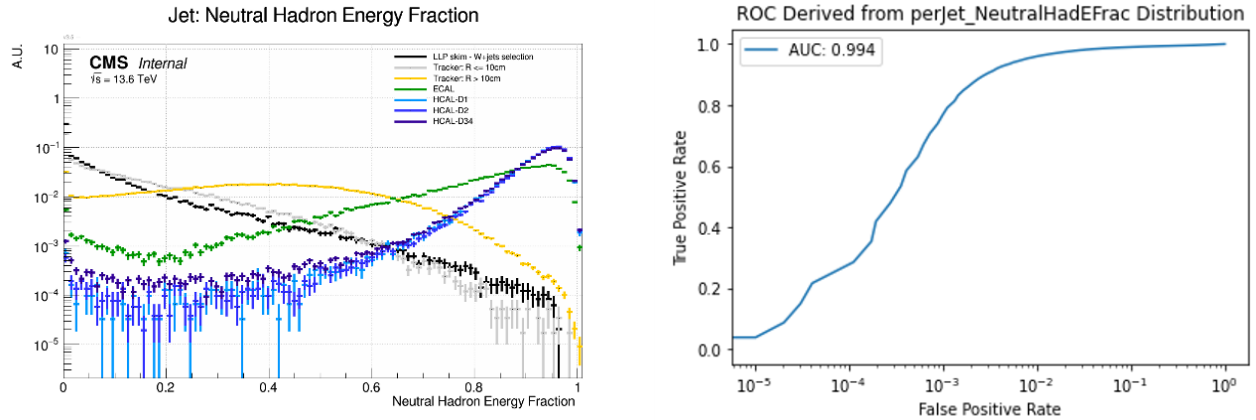
Figure 5: [Left] Distribution of the jet neutral hadron fraction for LLPs decaying within the HCAL (HCALD1, HCALD2, HCALD34 in dark blue, light blue, purples) and background, $W+$ jets (LLP skim, $W+$ jets selection in black) [4] [Right] ROC constructed directly from this distribution.

2 Methodology

We have progressed from the use of (Gradient) Boosted Decision Trees in ROOT's built-in multi-variate analysis framework (TMVA), to the construction of a Dense Neural Network (NN) trained on high level features of the jets, taking advantage of the highly discriminating custom variables. We have also gone on to train 3D Convolutional Neural Networks (3DCNNs) on the distribution of energy deposits within the HCAL across all four of its layers, and within 10x10 pixels of its η-ϕ space, where ϕ is the azimuthal angle.

All the training is performed on Monte-Carlo generated signal samples of LLPs that have decay within any of the four layers of the HCAL. The signal samples are split into different parameters of mass combinations of Higgs mass (MH) and LLP mass (MS), each containing on the order of 40k - 70k events. Table 1 presents the combinations of masses, as well as characteristic decay length ($c\tau$) of the LLP, which is also left as a free parameter in the model. All identical mass combinations are grouped when training classifiers since decay length has not yet been incorporated into these studies. For our background samples, we use data from a skim of Run 3 selecting $W+$ jets as it is particularly depleted in signal.

Table 1: Table of Higgs Mass, LLP Mass, and characteristic decay length ($c\tau$) for different mass combinations.

Higgs Mass [GeV]	LLP Mass [GeV]	LLP cτ [m]
125	15	1
125	15	3
125	15	10
125	50	3
125	50	10
250	120	10
350	80	0.5
350	160	10

To further improve upon the background rejection, we propose to use our developed classifiers as part of an 'asymmetric double tag' strategy. This involves developing a second, uncorrelated classifier, known as the 'inclusive tagger' that will be trained on jets decaying anywhere within the detector. In this way, we develop a per-event selection by requiring that the leading jet decays within the HCAL and passes the 'depth tagger', while the sub-leading jet passes the inclusive tagger. If the two classifiers are indeed uncorrelated, though this remains to be verified, we would expect to obtain a background efficiency equal to approximately the sum of the orders of background efficiency attained by the two classifiers.

3   Results

3.1   Training on Jet Variables

Figures 6 and 7 below display the performance of two models trained on a set of high level jet metrics (variables), which include tracking information, jet energy, mass, and momentum, information about the fraction of the jet's energy deposited in different layers of the HCAL, as well as custom variables such as $S_{\phi\phi} \equiv \frac{\sum \Delta\phi_i^2 \cdot E_i}{E_{\text{tot}}}$. This custom variable can be interpreted as a measurement of the 'spread' of a given jet, and serves as another powerful discriminating variable, as we would expect from recalling the illustration shown above in Figure 3. Overall, 41 metrics were used as inputs.

The first model trained is a Gradient-Boosted Decision Tree (BDTG) implemented within ROOT's Multi-Variate Analysis framework (TMVA) and used as a binary classifier for LLP signals and $W+$ jets background. The second model trained on these jet variables was a Dense Neural Network (NN) in Tensorflow's Keras, a more optimized machine learning library in Python. This model was more versatile as it could be trained as both a multi-classifier, distinguishing between LLPs decaying in depths 1-2, and depths 3-4 of the HCAL layer, as well as a binary classifier. From the plots below, it is seen that switching to the Tensorflow framework, has already allowed for some improvements. Namely, let us first focus on the ROC curve corresponding to the MH350-MS160 dataset (light grey, outermost curve in Figure 5), which is interpreted as the optimal performance of the model, given that it was trained specifically on the signatures arising from this mass combination; the model possesses a signal efficiency of 50% at a background efficiency of 10^{-4}. By contrast, looking at the rightmost plot in Figure 7, we see that the Dense NN possesses a higher 62% signal efficiency at the same value of background rejection, and even exhibits almost 20% sensitivity at just 10^{-5} background efficiency. Despite these improvements, these classifiers still did not provide an adequate ratio of signal to background efficiency for obtaining meaningful statistics in our full analysis. We must therefore continue our search for a more powerful classifier.
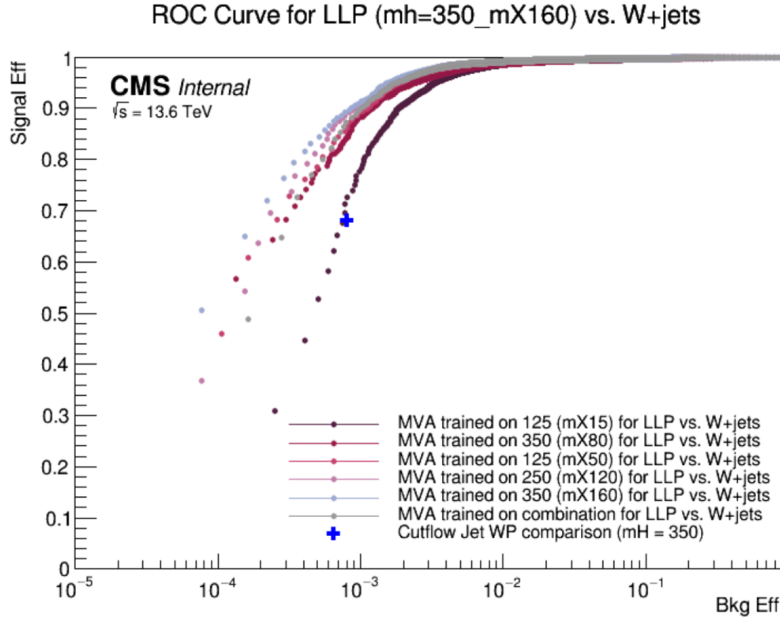

Figure 6: ROC curves displaying the performance of TMVA's Gradient-Boosted Decision Tree binary classifier, trained on the MH350-MS160 sample and evaluated on all other mass samples.
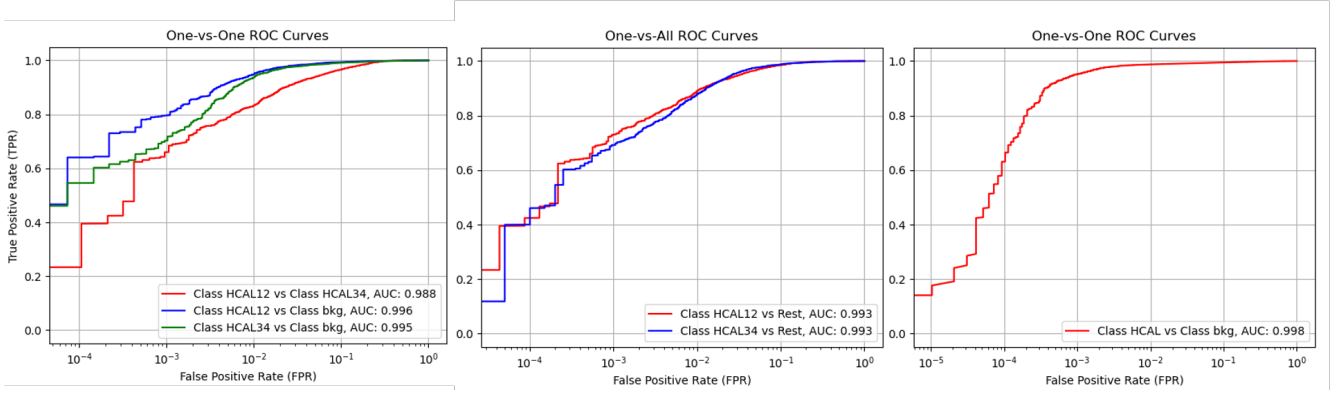
Figure 7: Series of ROC curves displaying the performance of Dense Neural Network implemented in Tensorflow's Keras and trained on the MH350-MS160 dataset. [Left] ROC curves shown for how well the model distinguishes between the two classes. [Center] ROC curves showing how well the model distinguishes 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.

3.2 Training on Jet Images

We have already seen that the signature of an LLP is intimately tied to the topology of the jet produced when it decays within the HCAL. As such, it would more natural to represent jets in a way that preserves this spatial information about their structure rather than simply storing this information in one-dimensional arrays of jet metrics, as in the case of the data inputted to the Dense NN and the BDTG. Therefore, our next logical step was to turn to the use of Convolutional Neural Networks (CNNs) and train on a dataset of "jet images". These images display the three-dimensional distribution of energy in the HCAL left behind by a single jet. The positions of the energy deposits were plotted in $\eta - \phi$ space which aligns with the coordinate system of the detector. In Figure 8, η and ϕ are plotted on the horizontal and vertical axes, respectively. One 3D image from a background jet, and one from a signal jet are depicted below. From the examples below, we may already see the stark difference between a background and signal jet, in terms of the energy distribution as well as its spread, just as we would expect from our earlier discussion of the jets' topologies.
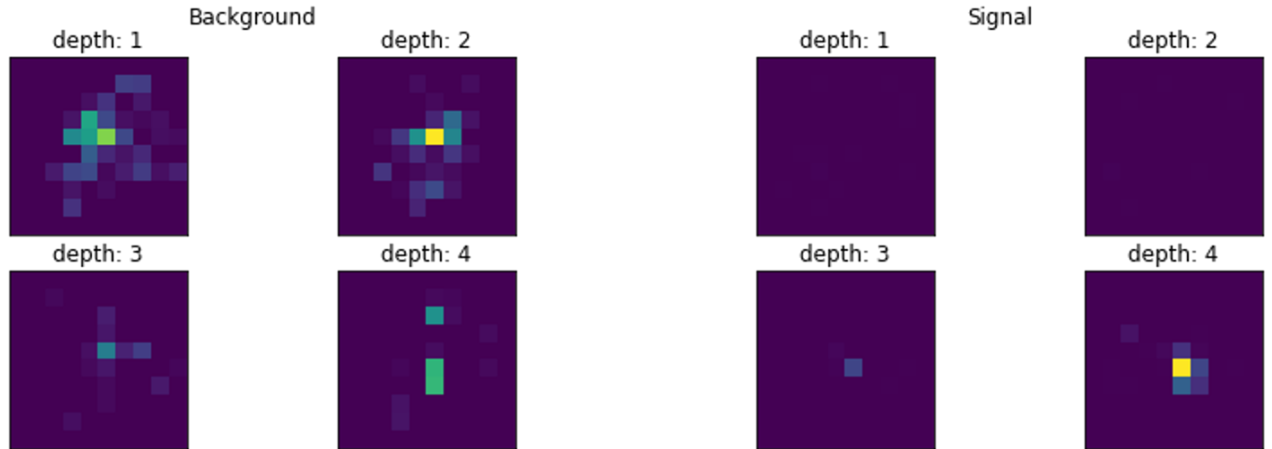

Figure 8: 3D jet images from background (left) and signal (right). Each image is normalized to preserve the relative value of energy in each depth.

In maintaining consistency across models, we show a series of ROC curves for the performance of the 3DCNN on all mass datasets, but trained solely with the MH350-MS160 dataset on the left-hand side of Figure 9. On the right, we include a ROC of a combined signal (all masses) and evaluated on an augmented dataset, where the images are mirrored and flipped so as to quadruple the size of the dataset, and yielding a total of just under 1 million samples. We see that the signal efficiency for the combined training is now up to 80% at 10^{-4}, and 40 % signal at 10^{-5}.

7

However, looking at the individual plots, we see that this value goes up to 90% for the MH350-MS160 combination which presents a greatly significant improvement. Furthermore, from this individual curve, we see that although we do not have sufficient samples within individual files to reach 10^{-5} background rejection, we see that close to this value we already find 80% for MH350-MS160. The left-hand plot indicates that the generated signals from the different mass combinations are sufficiently distinct, as there are at least two orders of magnitude difference between the best and worst performing signal in this model. This also indicates that the use of models trained on the individual signals may further enhance the separation between the signal and background.
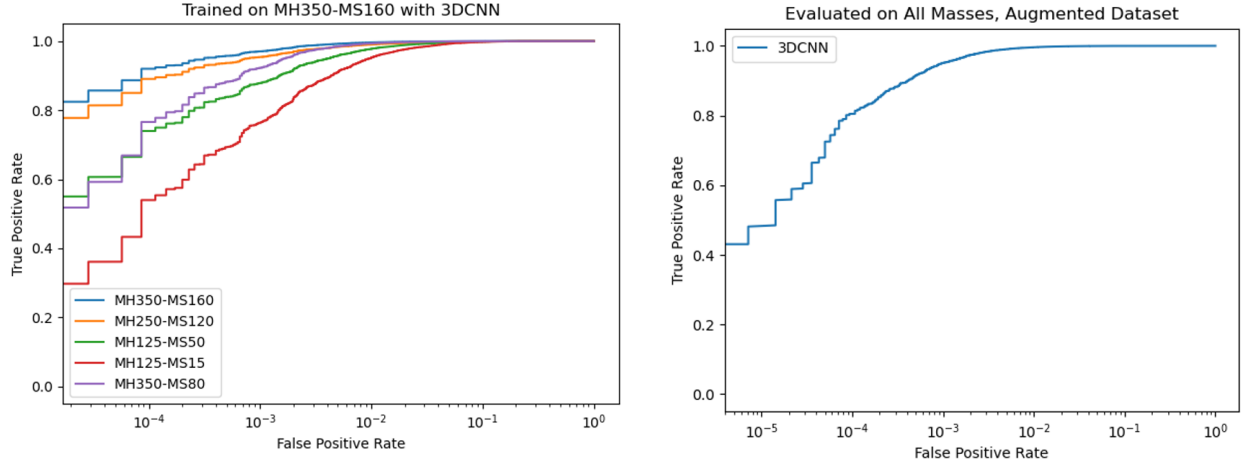

Figure 9: ROC curves showing the performance of the 3DCNN model. [Left] Mass comparison performance. [Right] Combined mass evaluation with augmentation of dataset.

3.3   Explainability

With the creation of these high-performing classifiers, we have become interested in understanding what features our networks are learning from. This is important as we want to verify the reliability of the model: to ensure that they learn significant features of the jets as opposed to some anomalous, insignificant characteristic that could lead them to fail when classifying data obtained from events in the CMS detector, as opposed to Monte Carlo generations. And so, at its core, our goal is to provide some explanation for the workings of our models.

A ranking of feature importance is directly obtained from TMVA's BDT implementation, which is computed by counting "how often the variables are used to split decision tree nodes, and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node" [8]. Table 2 below displays the top 10 most important variables determined by this BDT.

Table 2: Variable Importance Ranking for BDT

Rank	Variable	Variable Importance
1	perJet_NeutralHadEFrac	1.171e-01
2	perJet_LeadingRechitE/perJet_E	6.161e-02
3	perJet_EnergyFrac_Depth1	6.049e-02
4	perJet_PhoEFrac	5.970e-02
5	perJet_Track0dR	5.873e-02
6	perJet_S_phiphi	5.871e-02
7	perJet_MuonEFrac	5.855e-02
8	perJet_ChargedHadEFrac	5.719e-02
9	perJet_EnergyFrac_Depth4	4.701e-02
10	perJet_EnergyFrac_Depth2	4.668e-02

The highest-ranked variable is the jet neutral hadron fraction, which is consistent with our earlier discussion showing that just using jet neutral hadron fraction already produces a very powerful discrimination. Additionally,

three of the four depth energy fractions are included in the BDT's top 10 most important variables which aligns with our illustration of the displaced jets' distinct topology.

A DNN does not directly provide information about variable importance, and instead requires other methods to probe this 'black box'. One such method is through the computation of mutual information. Conceptually, mutual information can be thought of as a way to quantify how much information can be derived about one variable from the observation of another, [9]. Therefore, a higher mutual information score will demonstrate that more information is shared between two variables. In the context of mutual information between a classifier's score and an input feature, the mutual information would tell us the extent to which the score has been informed by that feature, and hence reflects the most important feature.

Table 3: Variable Importance Ranking for DNN based on Mutual Information

Rank	Variable Importance	Mutual Information
1	perJet_NeutralHadEFrac	6.015e-01
2	perJet_Track0Pt	4.759e-01
3	perJet_EnergyFrac_Depth1	4.555e-01
4	perJet_ChargedHadEFrac	4.539e-01
5	perJet_Track1Pt	4.353e-01
6	perJet_Track2Pt	3.819e-01
7	perJet_S_phiphi	3.663e-01
8	perJet_Track0dR	3.581e-01
9	perJet_EnergyFrac_Depth2	3.028e-01
10	perJet_S_etaphi	3.024e-01

We observe an overlap in the selection of important variables between the DNN and the BDT, indicating that the DNN is learning tractable features and functioning in a similar manner to the BDT. This overlap suggests that, although the DNN improves classification power, its reasoning remains consistent and interpretable. Therefore, this provides confidence that the model is robust and capable of performing reliably in the full analysis.

One successful method of understanding the workings of a Convolutional Neural Network is Gradient-weighted Class Activation Mapping, (GradCAM), [10]. GradCAM works by passing the image through the final convolution layer of the network. This last layer is known for extracting high-level features and patterns within the data [11]. The activation map is then weighted by the gradient of the final output score of the predicted class with respect to each pixel. In this way, each pixel of the activation map is weighted by how important it is in contributing to the choice of the final label. It is therefore important to note that this particular method looks at which features in the image correspond to what is most class-like in that image. Our implementation of GradCAM was first verified on the Fashion MNIST dataset [12]. This is because this dataset contains images of everyday clothing and therefore it is easy to validate the output of GradCAM against intuition. From the image below, we see that the GradCAM has correctly identified the shoe using its outline, which aligns with the way we ourselves might identify a shoe from an image.
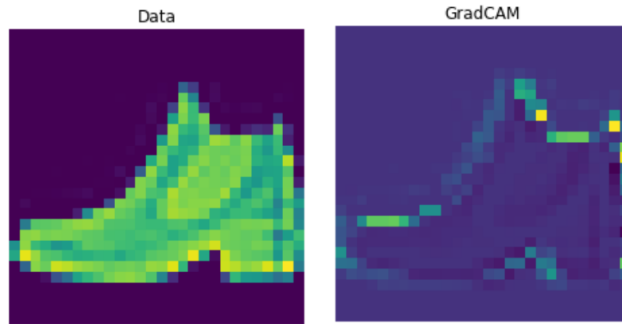

Figure 10: [Left] Image of a data sample taken from the Fashion MNIST dataset, depicting a shoe. [Right] The output of the GradCAM.

Applying GradCAM to our 3D jet images, we observe that the model assigns the greatest importance to cells with the highest energy deposits, as well as some secondary hot towers. This is illustrated in Figures 11 and 12.

However, the model's activation pattern differs between signal and background images. Firstly, the shape of the activation pattern is distinct between the two cases. For the background jet, the activations closely resemble the 3x3 kernel shape used in the last convolutional layer, resulting in a GradCAM that shows localized activations for primary and secondary hot towers. Even when several hot towers are close together, the activations overlap while maintaining the kernel-like shape, suggesting that the model perceives these towers as separate energy deposits. By contrast, the activations in the signal image more closely mimic the shape of the hot tower pattern, indicating that the model interprets this topology as a single energy cluster rather than individual hits. This interpretation is further supported by the observation that, in the background image, the model assigns greater importance to the surroundings of the highest energy cell, as seen by the darker center of the activation. However, in the signal image, the activations are more centered and aligned with the highest energy cells, reflecting a more direct focus on the signal's core features. Overall, this GradCAM analysis suggests that the model is effectively learning relevant features: recognizing the clustered, singular nature of the signal versus the more spread out, disorganized background. This indicates that the model is reliable and has strong potential for full-scale analysis.
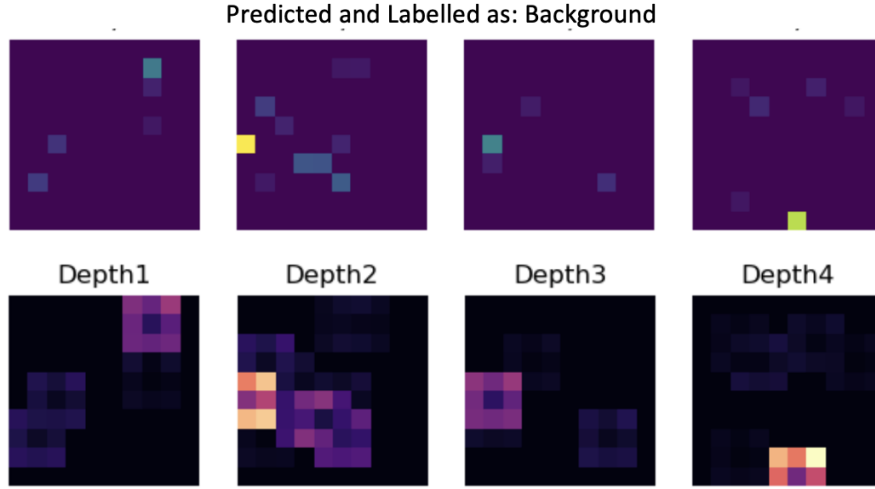

Figure 11: [Top] 3D image of a background jet, with each image corresponding to the depth marked below. [Bottom] The output of GradCAM for this image. The image depths correspond with that of the original image shown above so as to make a comparison of each depth between original image and GradCAM.
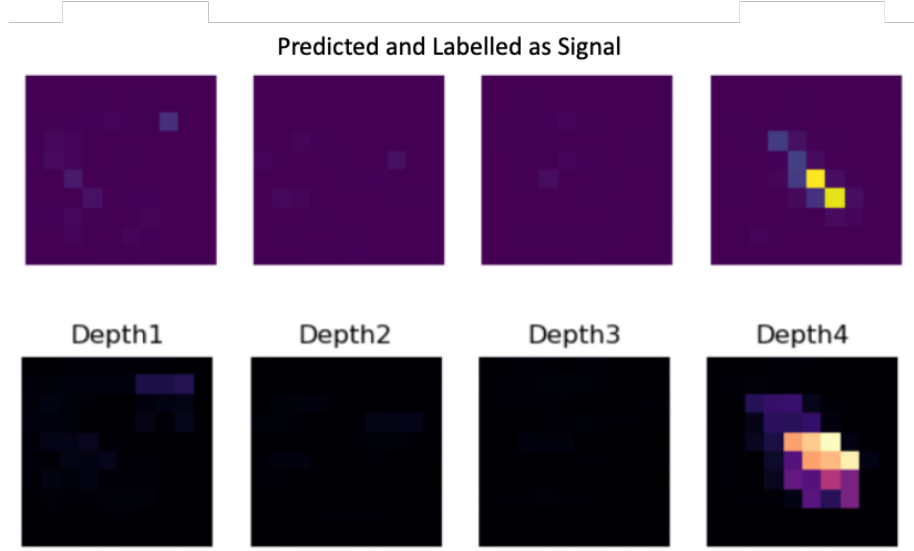
Figure 12: [Top] 3D image of a signal jet, with each image corresponding to the depth marked below. [Bottom] The output of GradCAM for this image. The image depths correspond with that of the original image shown above so as to make a comparison of each depth between original image and GradCAM.

4 Conclusion

The work presented here offers a promising exploration of machine learning-based methods for tagging delayed jets and identifying long-lived particles, with the most notable achievement being a 90% signal efficiency at just 0.01% background efficiency. This, coupled with deeper insights into the model's interpretability and feature extraction, marks significant progress.

However, before a full analysis can be performed, the most pressing limitation is the lack of sufficient data and MC samples to probe the models' performance at even lower levels of background rejection with sufficient statistical significance. Currently, our models are evaluated on roughly between 10^4 and 10^5 events, meaning that the lower bound of background efficiency that we can probe is between 10^{-5} and 10^{-4}. By aggregating more event samples, we may can investigate further into regions of down to 10^{-6} background efficiency.

Gathering more data will not only help to explore the model's potential at even lower background efficiencies but also help the model reach its maximal potential of signal efficiency at the current levels of background rejection. This is exemplified in the plot below, which shows the rising trend in signal efficiency at 0.01% background efficiency as we feed more training data into the 3D CNN. The curve suggests that we are approaching a limiting value, indicating that while there is room for improvement, it may only bring a marginal benefit. Instead, we note that it may be time to consider upgrading the 3DCNN model to a more complex architecture or even to look ahead to more powerful structures and algorithms altogether, such as the representation of jet images as graph neural networks instead.

In conclusion, this study provides a promising first insight into the potential of using ML-based LLP tagging, laying the groundwork for further exploration. With many exciting avenues for future research, we anticipate that continued developments will yield even greater advancements and deeper understanding in the area of exotic searches and long-lived particles.
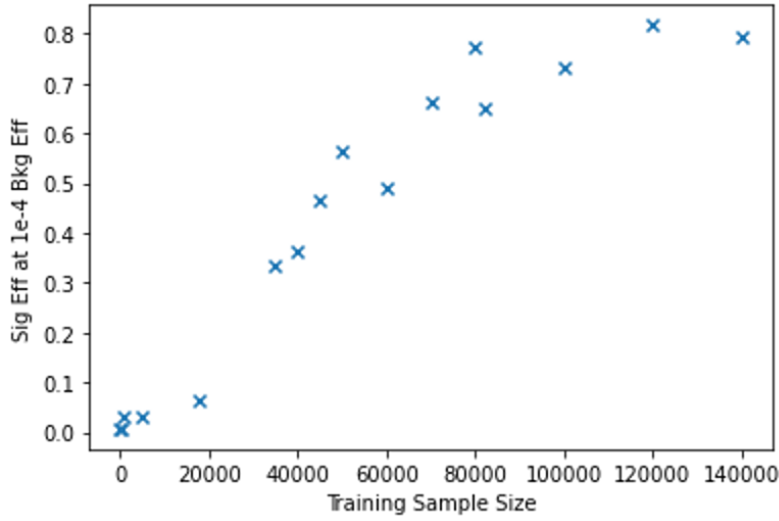
Figure 13: Plot of the value of signal efficiency achieved by the 3DCNN at a background efficiency of 10^{-4} against the number of training samples provided to the model. This classifier is trained and tested on a dataset comprised of MH350-MS160 and MH250-MS120 data, given that they showed consistently similar performances across all model versions (for example, see Figure 9 above). Fluctuations in the signal efficiency are present due to slight differences in convergence of the model's training. To achieve higher values of training samples, the training dataset was augmented, but the number of testing samples is constant.

5 Acknowledgements

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

[8] K. Albertsson, S. Gleyzer, A. Hoecker, L. Moneta, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, and S. Wunsch, "TMVA 4: Chapter 8.13.4," *arXiv preprint physics/0703039*, 2007.

[9] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, p. e87357, 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0087357

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: visual explanations from deep networks via gradient-based localization," *International journal of computer vision*, vol. 128, pp. 336–359, 2020.

[11] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that?" *arXiv preprint arXiv:1611.07450*, 2016.

[12] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," 2017. [Online]. Available: https://arxiv.org/abs/1708.07747