Model-Independent ML Search for New Physics

CERN UM-REU Final Report

Katarzyna Krzyżańska
Princeton University, Princeton NJ
Supervisor: Benjamin Nachman

August 20, 2021

Contents

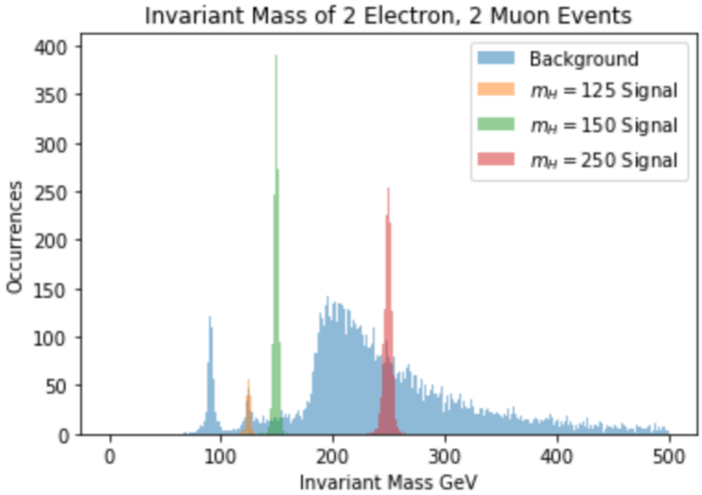

Figure 1: The distribution of the invariant mass for the background and signal simulations.

1 Introduction

Experiments in modern particle physics often search for physics beyond the Standard Model. Such searches test proposed theoretical models, but these models can be limited in scope or insensitive to deviations from what is specifically being tested. Several alternative model-independent algorithms have been suggested, where a neural network (NN) can be trained to learn directly from the data. [1, 2, 3, 4] are examples of such studies with varying model dependence, including analyses done with unlabeled data [2] or unsupervised training [4]. Another approach to anomaly detection is to focus on systems with very well known backgrounds, such as the production of four leptons via intermediate resonances. Here we may create a NN that is signal model independent while still relying on a background model. We could compare the data obtained from the ATLAS experiment to an accurate simulation of the background, letting the NN search for discrepancies between the two. These discrepancies have the potential to be signals arising from new physics beyond the standard model.

2 Data

The particular decay at the heart of my project is the decay of Higgs-like particles to two bosons, which then decay to two electrons and two muons: $H \to XY \to \mu^+\mu^-e^+e^-$ [5]. This is the dominant four lepton Higgs decay, where we do not distinguish between permutations of the leptons, e.g. $2\mu 2e$ and $2e2\mu$. I used a simulation of 100,000 background events resulting from Higgs decay, of which 31,451 were the $2\mu 2e$ decay for an efficiency of about 31.5%. For the purposes of testing and training, I also examined three simulated signals with Higgs masses of 125, 150, and 250 GeV with efficiencies of 7.7%, 9.8%, and 16.9%, respectively, as in figure 1. Ultimately the process of anomaly detection described below is itself signal model independent, but these test signals encompass a wide enough range of possibility to be considered proof of concept.

The simulated data can be interpreted to encompass many parameters. That is, for a given $2\mu 2e$ decay event, we can look at different information to provide the NN with an input of different dimensions. For example, the one-dimensional case examines only the mass of all four leptons together, $m_{4\ell}$. The three dimensional case examines $m_{4\ell}$ as well as the masses of the two lepton pairs separately as $m_{2\mu}$ and m_{2e}. The twelve dimensional case is either the momentum vector (p_x, p_y, p_z) or (p_T, η, ϕ) for each lepton, where p_T is the transverse momentum, η is the pseudorapidity, and ϕ is the polar angle coordinate. Generally, I found that the performance of the NN improved with more dimensions. However, for the sake of time and efficiency, most of the work in this report concerns the $m_H = 125$ GeV case in three dimensions.

The NN is designed to compare experimental data with simulation. Here, the "simulation" includes only background events, and the "data" consists of background events injected with some amount of signal events. Throughout I will refer to this signal amount either as the exact number of events, the corresponding cross-section, or the percentage of the signal relative to the entire $2e2\mu$ data set; e.g., a 5% 125 GeV signal means 5% of 7,700 events, or 385 events.

In order for the NN to accurately represent the experimental situation, it must accurately account for the integrated luminosity of the LHC. The current run of the LHC, Run 2, has an integrated luminosity of 150 fb^{-1}, with the scheduled Run 3 update set to increase it to 300 fb^{-1} [6]. I reduced the simulated background events to appropriately match the Run 2 luminosity using

$$\sigma_{pp\to h} Br() \int L dt = (\text{total events}) f \tag{1}$$

for a cross-section $\sigma_{pp\to h}$ and branching ratio $Br()$ obtained from [7, 8], and $\int L dt = 150\text{fb}^{-1}$. Ultimately I found that the fraction f required that we use 23.19% of the 31,451 background events.

3 The NN

3.1 Loss Distributions

I created a NN using Keras under Tensorflow in Python. A key part of the NN is the loss function I used, which was obtained from [1]:

$$L[f] = \sum_{(x,y)} [(1-y)\frac{N(R)}{\mathcal{N}_{\mathcal{R}}}(e^{f(x)} - 1) - y f(x)] \tag{2}$$

Here y is the input data, $f(x)$ is the output predicted by the NN, and $\frac{N(R)}{\mathcal{N}_{\mathcal{R}}}$ is the fraction of the data used in training, which here is 0.5. In course of training, the NN minimizes the loss with every iteration until it plateaus to some approximate final value. This particular loss function happens to have useful statistical properties where the loss reflects the test statistic used to compute p-values.

Partly this is useful in terms of calibrating the NN, but it also allows us to more closely examine the distribution of the final loss for different NNs found for the same background

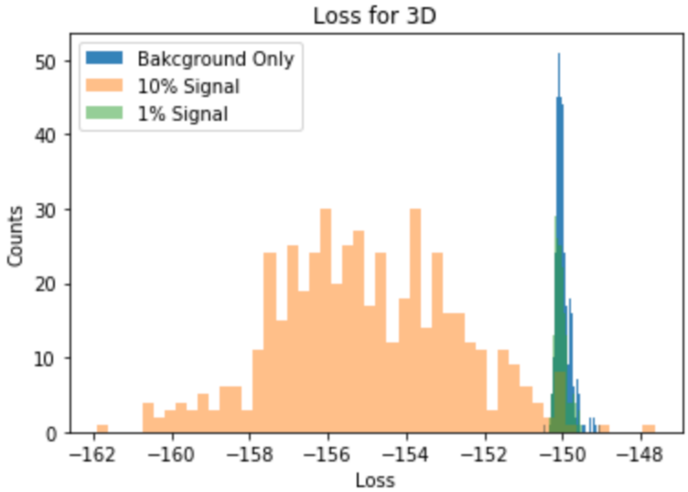

Figure 2: The loss distributions for 10%, 1%, and 0% signal. The analysis in this case has been performed for the data set before accounting for experimental integrated luminosity, and so this does not reflect the true expected loss distributions.

signal mix. That is, say we look at a "simulation" data set with 10% signal. For that data set, we train a large number of NNs, each of which produce a slightly different final loss. The loss distribution for 10% signal should be fairly different from the loss distribution for 1% signal, or that of just the background, as seen in figure 2. In fact, we can find the median p-value of the loss distribution for 10% signal relative to the background distribution. Because of the loss function's convenient statistical properties, this median p-value reflects the confidence with which it can detect an anomalous signal using the NN as well as the confidence of distinguishing the 10% signal "simulation" loss from the background loss.

We can compute the median p-value for a range of varying signal amounts, and ultimately produce a curve comparing median p-value to signal cross section. We know that the median p-value will decrease as signal increases- the more signal there is, the easier it is to detect- so we are interested in the point at which the p-value first crosses 0.05.

3.2 ROC Curves

Another way of measuring the efficacy of a NN is by examining its ROC curve; an example of a series of ROC curves is provided in figure 3. These curves are mainly illustrative, comparing the true and false positive rates for the NN. The false positive rate is the rate at which the NN incorrectly classifies pure background events as events belonging to the background/signal mix, and the true positive rate is similarly the correct classification. Ideally we want the former to be close to 0 and the latter close to 1.

As figure 3 demonstrates for a more general case, the more signal there is, the better the ROC curve. Failing that, we can also improve the curve by ensembling. After using a given data set on a large number of existing NNs (for example, using a full signal set on many NNs trained on 40 signal events, as in figure 4), we can take the median of the results to produce a much better curve.

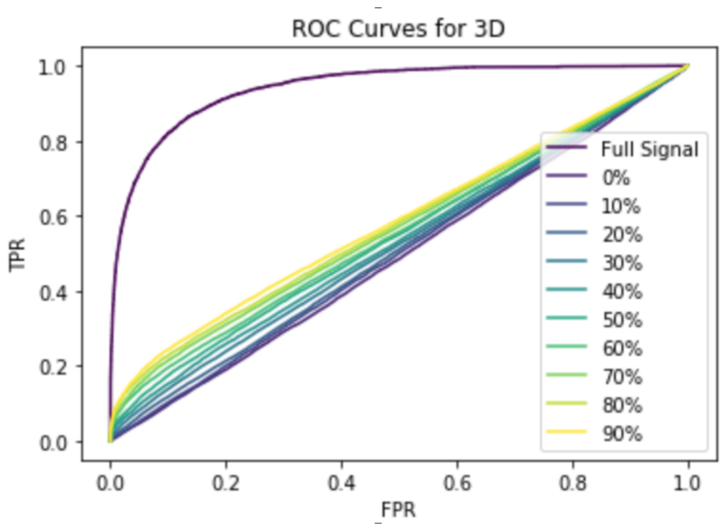

Figure 3: Some early ROC curves for varying signal percentages. As in figure 2, the data set here has not been adjusted for the experimental integrated luminosity.

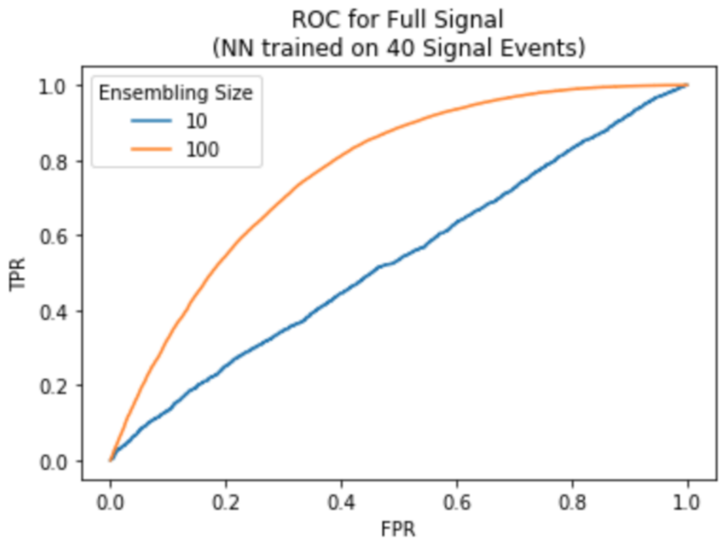

Figure 4: An example of how ensembling may improve the ROC curves. After finding the median of 100 ROC curve scores for a "simulation" set with 40 signal events, we note that the ROC curve tends to have a higher TPR rate than FPR rate, approaching the ideal case. In this instance, the data set has been adjusted for the experimental luminosity and is working under a setup nearly like the one described in section 3.3. However, the number of epochs is 40 rather than 200, which is likely less than ideal.

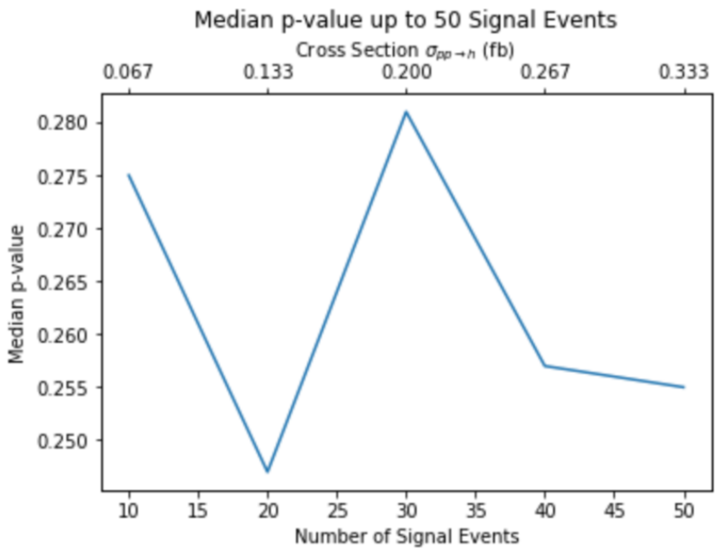

Figure 5: An intermediate result in the process of trying to determine the median p-value curve for varying amounts of signal.

3.3 Refining the NN

The NN itself also has a number of parameters I can set to alter its behavior. The qualities I looked at included the number of epochs, the batch size, the number of learnable parameters, the learning rate, and the optimizer used. Although I did perform organized tests to better determine the optimal settings, I nonetheless found iterative testing to be valuable, and so the final configuration I settled on changed even after formal tests. At the time of this report, the NN uses 200 epochs, a batch size of 200, 50 learnable parameters, a learning rate of 10^{-5}, and the Adam optimizer. I found the learning rate to have the most drastic effect on NN performance, but perhaps more importantly I found that if the number of epochs is too low, the NN simply won't finish learning and the losses obtained will be largely unhelpful.

4 Results

Much of this project was iterative, where I would create a NN under a certain data set that produced decent, but imperfect results, and then I would set out to improve both the NN and how I used the data. As such, the most recent median p-value plot I produced in figure 5 is still just another step in that process, and it is far from the ideal curve.

For the most part this is the result of an error that occurred sometime between when I refined my NNs, as in section 3.3, and when I attempted to produce this final plot. Though testing suggested that 40 epochs would be sufficient, upon closer inspection this was not the case. It would seem that at this stage the NN had not, in fact, finished learning, so the loss distributions would entirely overlap regardless of the amount of signal used and the calculated p-value was largely unhelpful. However, further tests suggests that 200 epochs would instead be sufficient. The increased length of time these NNs would require to run, coupled with the close of my internship, prevented me from giving this the attention it deserve, but the framework is in place to better examine this.

5 Conclusion

My work with machine learning in signal model independent anomaly detection serves as a proof of concept that such NN can indeed be used to pick out certain signals in the Higgs decay to four leptons. I have created a framework with which the median p-value curve can be created that could determine the minimum signal necessary for detection. Though there are further improvements that still can and ought to be made on the system, these adjustments are no so much technically difficult as time intensive, so I believe that with only a few more steps I would produce more definitive results.

6 Acknowledgements

I would like to thank Benjamin Nachman for his guidance in support over the course of this project. Thank you also to Myron Campbell, Junjie Zhu, and Mary Kate Bossard for their work with the UM-REU program. This material is based upon work supported by the National Science Foundation under Grant No. PHYS-1949923.

References

[1] R. T. D'Agnolo and A. Wulzer, "Learning new physics from a machine," *Physical Review D*, vol. 99, no. 1, pp. 1–34, 2019.

[2] A. Andreassen, B. Nachman, and D. Shih, "Simulation assisted likelihood-free anomaly detection," *Physical Review D*, vol. 101, no. 9, p. 95004, 2020.

[3] J. Collins, K. Howe, and B. Nachman, "Anomaly Detection for Resonant New Physics with Machine Learning," *Physical Review Letters*, vol. 121, no. 24, pp. 1–7, 2018.

[4] B. Nachman and D. Shih, "Anomaly detection with density estimation," *Physical Review D*, vol. 101, no. 7, 2020.

[5] The ATLAS Collaboration, "Search for Higgs boson decays to beyond-the-Standard-Model light bosons in four-lepton events with the ATLAS detector at s=13 TeV," *Journal of High Energy Physics*, vol. 2018, no. 6, pp. 1–50, 2018.

[6] E. C. López, "The ATLAS upgrade programme for LHC Run-3 and the HL- LHC," 2015.

[7] The LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, R. Tanaka, J. R. Andersen, P. Artoisenet, E. A. Bagnaschi, A. Banfi, T. Becher, F. U. Bernlochner, S. Bolognesi, P. Bolzoni, R. Boughezal, D. Buarque, J. Campbell, F. Caola, M. Carena, F. Cascioli, N. Chanon, T. Cheng, S. Y. Choi, A. David, P. de Aquino, G. Degrassi, D. Del Re, A. Denner, H. van Deurzen, S. Diglio, B. Di Micco, R. Di Nardo, S. Dittmaier, M. Duhrssen, R. K. Ellis, G. Ferrera, N. Fidanza, M. Flechl, D. de Florian, S. Forte, R. Frederix, S. Frixione, S. Gangal, Y. Gao, M. V. Garzelli, D. Gillberg, P. Govoni, M. Grazzini, N. Greiner, J. Griffiths, A. . V. Gritsan, C. Grojean, D. C. Hall, C. Hays, R. Harlander, R. Hernandez-Pinto, S. Hoche, J. Huston, T. Jubb, M. Kadastik, S. Kallweit, A. Kardos, L. Kashif, N. Kauer, H. Kim, R. Klees, M. Kramer, F. Krauss, A. Laureys, S. Laurila, S. Lehti, Q. Li, S. Liebler, X. Liu, H. E. Logan, G. Luisoni, M. Malberti, F. Maltoni, K. Mawatari, F. Maierhofer, H. Mantler, S. Martin, P. Mastrolia, O. Mattelaer, J. Mazzitelli, B. Mellado, K. Melnikov, P. Meridiani, D. J. Miller, E. Mirabella, S. O. Moch, P. Monni, N. Moretti, A. Muck, M. Muhlleitner, P. Musella, P. Nason, C. Neu, M. Neubert, C. Oleari, J. Olsen, G. Ossola, T. Peraro, K. Peters, F. Petriello, G. Piacquadio, C. T. Potter, S. Pozzorini, K. Prokofiev, I. Puljak, M. Rauch, D. Rebuzzi, L. Reina, R. Rietkerk, A. Rizzi, Y. Rotstein-Habarnau, G. P. Salam, G. Sborlini, F. Schissler, M. Schonherr, M. Schulze, M. Schumacher, F. Siegert, P. Slavich, J. M. Smillie, O. Stal, J. F. von Soden-Fraunhofen, M. Spira, I. W. Stewart, F. J. Tackmann, P. T. E. Taylor, D. Tommasini, J. Thompson, R. S. Thorne, P. Torrielli, F. Tramontano, N. V. Tran, Z. Trocsanyi, M. Ubiali, P. Vanlaer, M. V. Acosta, T. Vickey, A. Vicini, W. J. Waalewijn, D. Wackeroth, C. Wagner, J. R. Walsh, J. Wang, G. Weiglein, A. Whitbeck, C. Williams, J. Yu, G. Zanderighi, M. Zanetti, M. Zaro, P. M. Zerwas, C. Zhang, T. J. . E. Zirke, and S. Zuberi, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," no. July, 2013.

[8] The ATLAS Collaboration, “Measurement of inclusive and differential cross sections in the H $\rightarrow$ ZZ * $\rightarrow$ 4 decay channel in pp collisions at s=13 TeV with the ATLAS detector,” *Journal of High Energy Physics*, vol. 2017, no. 10, 2017.