

Deep Learning Versatility in New Physics Searches

Miguel Crispim Romão^a

^aLIP,

Av. Professor Gama Pinto 2, 1649-003 Lisboa, Portugal

E-mail: mcromao@lip.pt

In this communication, we will discuss the application of Deep Learning models as a discriminant step to improve sensitivity at searches for new physics. Of particular interest, we will focus on the transferability of Deep Learning models, where a neural network trained to isolate a specific signal can still provide sensitivity when deriving upper limits on a different process. This is expanded to include a discussion on the versatility of Deep Learning models to provide enough sensitivity in cases where the signal present in the sample does not follow the assumptions of an analysis.

40th International Conference on High Energy physics - ICHEP2020

July 28 - August 6, 2020

Prague, Czech Republic (virtual meeting)

1. Introduction

In HEP, one particular application of Machine Learning (ML) models is to isolate potential signals for New Physics (NP) at collider experiments from the Standard Model (SM) background. In this talk we present how DL models can provide solutions for models that can isolate unexpected Beyond the Standard Model (BSM) events, paving the way for generic NP searches.

This communication is organised as follows. In the next section we will outline the details of the data simulation, both for SM events and for a selection of BSM candidates. Next, we will present the work on the transferability of the Deep Learning (DL) models [1], where it was shown that a DL model trained on a concrete BSM candidate can still provide competitive discrimination power when presented with a new, previously unseen BSM realisation. This leads to the natural question if one can teach DL models to learn what the SM background is in order to be able to identify BSM candidates as anomalies, this is explored with Anomaly Detection (AD) models in the following section, where we show how DL AD models can isolate BSM signals [2], none of which used during training.

2. Data Simulation

The SM events cocktail was prepared as to target processes with final state composed of at least two leptons (*i.e.* electrons or muons), at least one jet identified as originated by the fragmentation of a b -quark and large scalar sum of transverse momentum (p_T) of all reconstructed particles in the event ($H_T > 500$ GeV), motivated by a generic search for large mass resonances with intermediate products involving top-quarks and W/Z vector bosons in the decay chain. Our main source of background is therefore composed of Z +jets, top pair production and dibosons.

We produced BSM candidates that are expected to fall into the kinematic and topological regions of the SM background prepared as discussed above. We focused on a Flavour-Changing Neutral Currents (FCNC) interaction associated with a top quark production and in six different realisations of vector-like T quarks. FCNC interactions in single top-quark production can appear in dimension-four or dimension-five operators of an effective SM extension. In the current paper we focus on the tZ production via FCNC coupling [3, 4] since it produces a final state belonging to the $tZ + X$ category we are targeting. For the Vector-Like Quark (VLQ) we focus on this tZ final state, considering the pair production of T quarks via a SM gluon and assuming that T belongs to a weak isospin doublet, where the $\text{BR}(T \rightarrow tZ)$ is approximately $1/2$ [5, 6]. Another class of models predicting the existence of vector-like T -quarks are the composite Higgs Models with partial compositeness [7]. In such models, new massive colour octets are naturally expected and thus the pair production of T -quarks can occur via this new heavy gluon, producing events with kinematic properties expected to be different from the $T\bar{T}$ production via a SM gluon [8]. In the presented work we consider as benchmark a heavy gluon with a mass of 3.0 TeV and the same T masses and weak isospin charge as for the standard production case. To illustrate the effect on the signal kinematics, different T masses were considered: 1.0, 1.2 and 1.4 TeV.

3. Transferability of Supervised Deep Learning Classifiers

One of the characteristics of DL models is they are composed of differential blocks, called layers, which are usually stacked together. This allows DL models to learn hierarchical representations from the data. As a consequence, DL models can learn features that are useful for similar tasks on different data. A tantalising example is the one of computer-vision, where it was shown that DL models learn first features that are transferable to images of classes not shown in training [9], allowing one to reuse parts of DL models trained on, say to differentiate between cats and dogs to another, say to differentiate cars from aeroplanes. In [1], we showed that supervised DL models trained to separate a specific BSM candidate signal from the SM background still manifested discrimination capacity when presenting another BSM signal candidate not seen during training.

In fig. 1, we see how a model trained on a specific VLQ realisation still provides discrimination power when presented with any other sample. Of special notice we point to the fact that the FCNC signal, being the most similar to the SM background, is still being isolated, whereas the transferability between different VLQ candidates appears to have a little effect in the model performance. In addition, we observe how every model trained on every signal still holds discrimination power when presented an FCNC sample. To further illustrate this we computed the upper limits on the

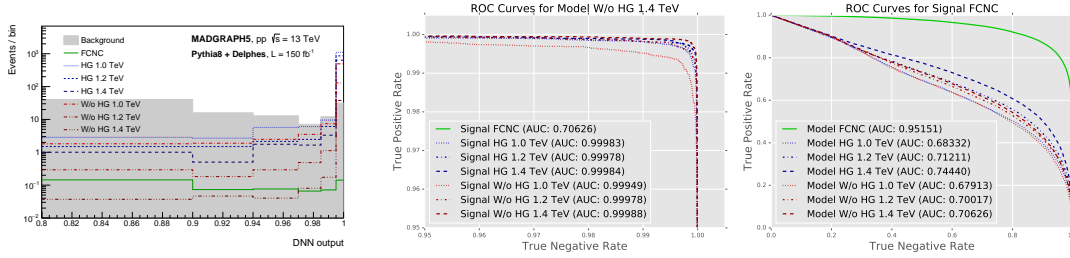


Figure 1: Discrimination of models trained on specific BSM candidates on different BSM candidates. Left: The output of a neural classifier trained on the VLQ (with SM production) with mass 1.4 TeV. Centre: The ROC curves of the same model when passing different BSM candidates. Right: The ROC of the FCNC signal across all models trained on all the BSM candidates.

expected signal strength using the CL_s methods using the outputs of each neural classifier. The results can be seen in fig. 2, where we observed that the models trained on VLQ signals provide the same limits across all other VLQ candidates. This shows that DL models are fairly robust against the changes of specific theory parameters, such as the mass of the VLQ produced, but also on the topology generating the final state, as can be seen by the absence of limit deterioration across the VLQ candidates produced from the SM to the heavy-gluon case. These results highlight that even when trained on a supervised manner, DL models can detect unexpected BSM candidates that were not using training, suggestions that DL models have the ability to learn how to isolate the SM region of the phase space, helping to isolate unexpected NP.

4. Anomaly Detection as a Tool for Searches at Colliders

Although the previous work showed that supervised neural classifiers offer the possibility to detect unexpected BSM signals on the data, it would be more satisfying if we had a model that could

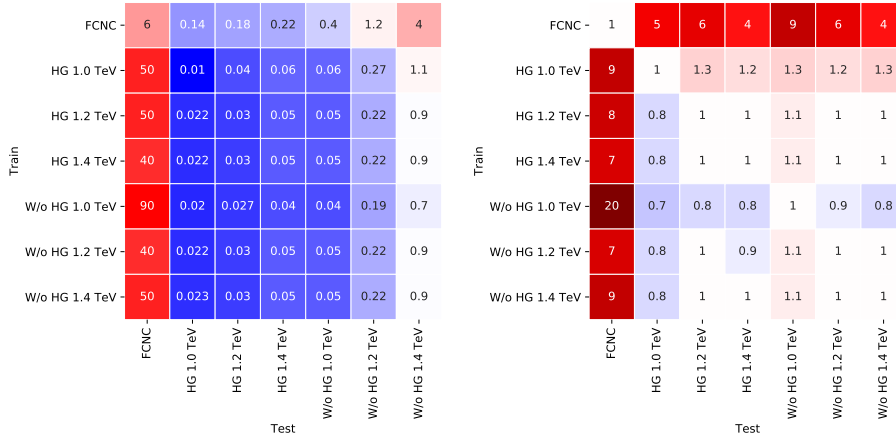


Figure 2: Upper limits on the expected signal strength. Left: Nominal value. Right: Normalised to the diagonal, which represents the case of using the supervised model on the sample containing the same type of signal that it was trained on.

learn the SM event distribution as to infer if a new event is likely to belong to the SM or otherwise. This is addressed by AD, where one train a model on normal samples, which we take to be the SM cocktail, after which the model will produce anomaly scores on events. This anomaly score can be seen as new variable on which we can compute upper limits on the expected signal strength. This was studied in [2], where we tested four different AD methods and benchmarked them against the same six BSM candidates presented already. We studied two shallow methods: Isolation Forest and Histogram Based Outlier System); and two deep methods: Autoencoder and Deep Support Vector Data Description (Deep-SVDD). Each method was trained only on SM events, and their outputs were used to calculate expected upper limits on signal strength using the CL_s method. The performance was assessed by comparing to the upper limits obtained when using a supervised neural discriminator. In fig. 3 we show the outputs of all four AD methods for the background sample and all BSM candidates. We see that all VLQ samples have their anomaly scores distribution shifted to the right, providing discrimination power. As for the FCNC, we observe that while the same happens to the Deep-SVDD, the distribution is very similar to the SM background, as expected. Since the distributions in fig. 3 are different between themselves and across different samples, it is important to understand how they are correlated. In fig. 4 we see that both shallow methods are highly correlated, while the deep ones are not. In particular, we notice that Deep-SVDD and the AE are considerably uncorrelated. In addition, we observe that the AD methods are picking up different events as the anomalous ones. While the shallow methods and the AE are focusing on the tails of the distribution, the Deep-SVDD is capturing other events, showing how these methods are not learning the same notion of outlyingness. As before, we produced the upper limits for signal strength and compared them to the ones one would get if using a supervised neural classifier. This comparison is shown in fig. 5. We observe that for the FCNC obtain least performance degradation with the Deep-SVDD, while for the VLQ the AE is able to produce limits which are competitive with those produced in a supervised manner. In addition, we observe how the Deep-SVDD retains similar performance across different BSM candidates, noticeably across FCNC and VLQ quarks.

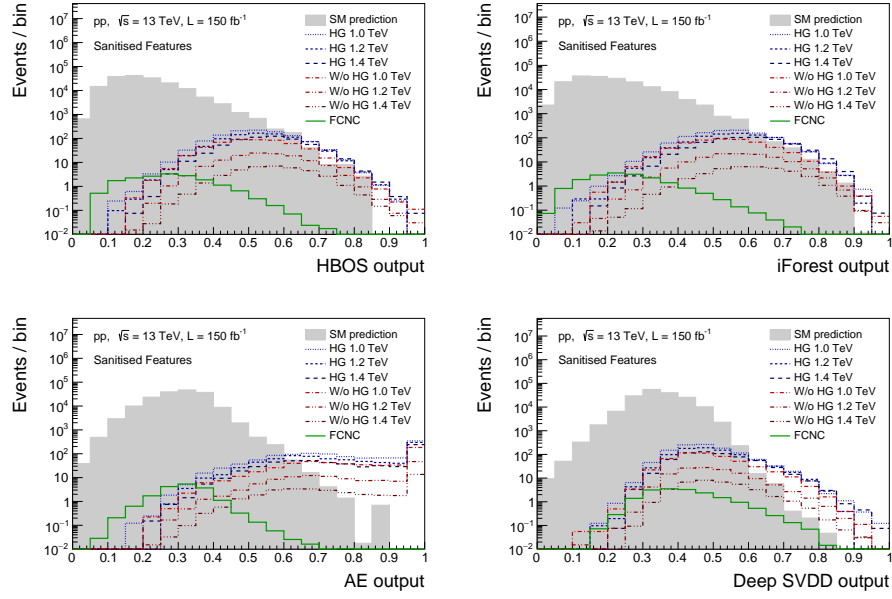


Figure 3: Outputs for the different AD methods for all samples.

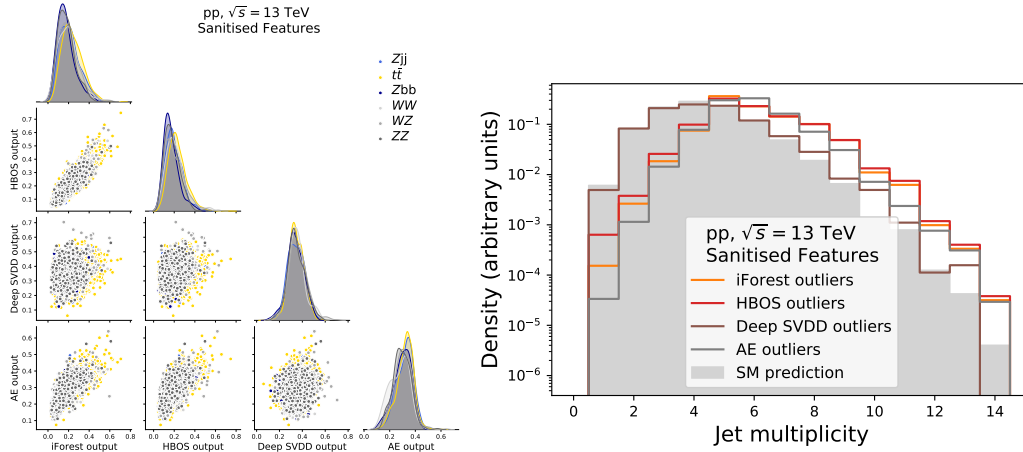


Figure 4: Left: Correlations of the outputs of the different methods. Right: The distribution of Jet multiplicity for the 10% outliers for all methods.

These results show the potential for DL AD models to provide generic discrimination power to any BSM candidate. In addition, we have also showed that different AD models learn differently what an outlying event is, supporting the idea that different AD methods are complementary and sensitive to different BSM candidates.

5. Conclusions

In this talk we presented recent work on the application of DL models to isolate BSM candidates from the SM background in the context of searches for NP. We showed that DL supervised models

Model	FCNC	HG 1.0 TeV	HG 1.2 TeV	HG 1.4 TeV	W/o HG 1.0 TeV	W/o HG 1.2 TeV	W/o HG 1.4 TeV
Supervised DNN	1	1	1	1	1	1	1
AE	30	2.9	2	1.6	1.2	0.98	1.8
Deep SVDD	10	70	30	10	30	10	10
HBOS	20	60	40	20	30	10	20
IForest	20	1e+02	60	40	60	30	30

Figure 5

provide a versatile solution that persists discrimination power when used in a BSM candidate not used during training, producing reasonable upper limits on signal strength. In addition, we showed how modern AD methods based on DL architectures provide powerful generic BSM signal candidates discriminators, being able to isolate BSM signals that were never used during training. We also showed that these methods learn differently what an outlier is, highlighting their complementarity and sensitivity to different BSM candidates.

References

- [1] M. Romão Crispim, N.F. Castro, R. Pedro, and T. Vale. Transferability of Deep Learning Models in Searches for New Physics at Colliders. *Phys. Rev. D*, 101(3):035042, 2020.
- [2] M. Crispim Romao, N.F. Castro, and R. Pedro. Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders. 6 2020.
- [3] C. Degrande et al. Automatic computations at next-to-leading order in QCD for top-quark flavor-changing neutral processes. *Phys. Rev.*, D91:034024, 2015.
- [4] G. Durieux, F. Maltoni, and C. Zhang. Global approach to top-quark flavor-changing interactions. *Phys. Rev.*, D91(7):074017, 2015.
- [5] M. Buchkremer et al. Model Independent Framework for Searches of Top Partners. *Nucl. Phys.*, B876:376–417, 2013.
- [6] J. A. Aguilar-Saavedra. Identifying top partners at LHC. *JHEP*, 11:030, 2009.
- [7] D. B. Kaplan. Flavor at SSC energies: A New mechanism for dynamically generated fermion masses. *Nucl. Phys.*, B365:259–278, 1991.
- [8] J. P. Araque, N. F. Castro, and J. Santiago. Interpretation of Vector-like Quark Searches: Heavy Gluons in Composite Higgs Models. *JHEP*, 11:120, 2015.
- [9] J. Yosinski et al. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.