# Transferability of deep learning models in searches for new physics at colliders

M. Crispim Romão[1,*] N. F. Castro[1,2,†] R. Pedro[1,‡] and T. Vale[1,2,§]

[1]*LIP, Avenida Professor Gama Pinto 2, 1649-003 Lisboa, Portugal*
[2]*Departamento de Física, Escola de Ciências, Universidade do Minho, 4710-057 Braga, Portugal*

In this work we assess the transferability of deep learning models to detect beyond the standard model signals. For this we trained deep neural networks on three different signal models: $tZ$ production via a flavor changing neutral current, pair production of vectorlike $T$-quarks via standard model gluon fusion and via a heavy gluon decay in a grid of three mass points: 1, 1.2 and 1.4 TeV. These networks were trained with $t\bar{t}$, $Z + \text{jets}$ and dibosons as the main backgrounds. Limits were derived for each signal benchmark using the inference of networks trained on each signal independently, so that we can quantify the degradation of their discriminative power across different signal processes. We determine that the limits are compatible within uncertainties for all networks trained on signals with vectorlike $T$-quarks, whether they are produced via heavy gluon decay or standard model gluon fusion. The network trained on flavor changing neutral current signal, while struggling the most on the other signals, still produces reasonable limits. These results indicate that deep learning models are capable of providing sensitivity in the search for new physics even if it manifests itself in models not assumed during training.

## I. INTRODUCTION

Although machine learning has a long history in high energy physics (HEP), we have recently witnessed a surge in interest in new methods and algorithms emerging from deep learning [1]. Deep learning models differ from those of traditional machine learning as they are composed of a stack of layers with nonlinear functions that have the capacity to learn hierarchical features from the inputs [2]. Indeed, it has been shown that deep learning models for a computer version trained on a certain task can be adapted to a different, albeit similar, task [3] as the layers closer to the inputs learn low-level features that progressively become higher level as they are transformed by the subsequent layers. In computer vision this manifests as the first layers learn about localized pixel variations, the following learn about textures and patterns, and finally the last encode high-level features such as *dog* or *cat*. This has led to reusing deep learning models trained on a specific task, say to discriminate between dogs and cats, to perform a different one, say to discriminate between cars and trucks, by keeping the lower layers and train or fine-tune the layers responsible for the high-level features. A natural question then arises if the same transferability happens on deep learning models used in HEP. More specifically, how transferable is a deep learning model trained on reconstructed physical observables in the discriminative task of separating signal from background when different signals are considered? Since the task is the same we will not retrain or fine-tune the last layers. Instead, we will assess how a trained model can perform the same task given a different signal sample.

The goal of this work is to study how deep learning models trained on a specific signal are transferable to new signals unseen during training. As such, a few benchmark signals were considered, having in common the presence of $tZ + X$ final states. Since the target topology determines which Standard Model (SM) backgrounds have to be considered, we focused on events with at least a pair of leptons, arising from the decay of the $Z$-boson or top quark, as well as with jets originating from $b$-quarks, from the decay of the top quark. Three classes of well motivated signals were considered in this context: (a) $tZ$ production via effective field theory operators inducing a flavor changing neutral current (FCNC) coupling $utZ$; (b) pair production of vectorlike $T$-quarks (VLT) via a SM gluon and (c) pair production of vectorlike $T$-quarks via a beyond-SM (BSM) heavy gluon. By testing the ability of the model to recognize

---
[*]mcromao@lip.pt
[†]nuno.castro@cern.ch
[‡]rute@lip.pt
[§]tiago.vale@cern.ch

these new signal events, we probe an intermediate step of generic detection, in the midway between fully signal-tailored classifiers and anomaly detectors.

The remainder of the paper is organized as follows. We start by briefly describing and motivating the BSM phenomenological grounds covering the new physics signals explored in the paper. The following section presents the data used in the analysis, from signal and background simulation to event selection and data preparation. Afterwards we present the deep neural network (DNN) model and the procedure used for its architecture optimization. Results are presented and discussed in the section that follows, with conclusions being drawn at the end.

## II. BEYOND STANDARD MODEL SIGNALS

FCNC interactions in single top-quark production can appear in dimension-four or dimension-five operators of an effective SM extension. The search for FCNC processes involving top quarks and $Z$-bosons at the Large Hadron Collider (LHC) has been performed considering both $t\bar{t}$ production with $t \to qZ$ decays (with $q$ being an $u$- or $c$-quark) [4,5] and $tZ$ production [6]. In the current paper we focus on the $tZ$ production via a $utZ$ FCNC coupling [7,8] since it produces a final state belonging to the $tZ + X$ category we are targeting.

Both the ATLAS and CMS collaborations have developed a comprehensive program to search for vectorlike quarks at the LHC, targeting an extensive list of different final states. Among those, multilepton final states originated by the pair production of $T$-quarks, where at least one of them decays into $tZ$, were performed and stringent limits on the mass of the new quarks as a function of the branching ratio (BR) for the considered decays were obtained [9,10]. 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 BR$(T \to tZ)$ is approximately $1/2$ [11,12]. In order to illustrate the effect on the signal kinematics, different $T$ masses were considered: 1.0, 1.2 and 1.4 TeV.

One class of models predicting the existence of vectorlike $T$-quarks is the composite Higgs models [13] with partial compositeness [14]. In such models, new massive color 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 [15,16]. In the current paper 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.

## III. SIMULATED SAMPLES

We use simulated samples of proton-proton collision events generated with MADGRAPH5_MCATNLO2.6.5 [17] at leading order with a center-of-mass energy of 13 TeV. The parton showering and hadronization was performed with PYTHIA 8.2 [18], using the CMS underlying event tune CUETP8M1 [19] and the NNPDF 2.3 [20] parton distribution functions. The detector simulation employs the DELPHES3 [21] multipurpose detector simulator with the default configuration, corresponding to the parameters of the CMS detector.

We target processes with a 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) [22], 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 ($t\bar{t}$) production and dibosons ($WW$, $WZ$ and $ZZ$). To obtain a robust statistical representation of the backgrounds across a significant region of the phase space, and especially in the high $H_T$ limit where most of the signal is expected to be, we generate each of the mentioned backgrounds in ranges of kinematic properties by applying an event filter at parton level according to the following:

  (i)  The top/anti-top $p_T$ ($p_T^{top}$) for $t\bar{t}$: $p_T^{top} < 100$ GeV, $p_T^{top} \in [100, 250]$ GeV, $p_T^{top} > 250$ GeV.

 (ii)  The scalar sum of the $p_T$ of the hard-scatter outgoing particles for $Z +$ jets: $H_T < 250$ GeV, $H_T \in [250, 500]$ GeV, $H_T > 500$ GeV.

(iii)  $W/Z$ $p_T$ ($p_T^{W/Z}$) for dibosons: $p_T^{W/Z} < 250$ GeV, $p_T^{W/Z} \in [250, 500]$ GeV, $p_T^{W/Z} > 500$ GeV.

Over 18 M events were simulated: 500 k per signal sample, 8 M for $Z +$ jets, 3 M for $t\bar{t}$ and 1.5 M per diboson sample. The background and signal samples are normalized to their expected yield after selection using each process generation cross section at leading order, computed with MADGRAPH5, matching a target luminosity of 150 fb$^{-1}$. The full dataset statistics is then split into three subsets according to the $0.3 : 0.2 : 0.5$ proportion, to be used by the DNN for training, validation and test, respectively.

Figure 1 shows the distribution of $H_T$ and leading large-radius jet $p_T$ for the total simulated background and each signal. $T\bar{T}$ production via the SM gluon exhibits significant kinematic dependency on the vectorlike $T$ mass, with heavier $T$ processes producing harder final state objects. On the contrary, the kinematics of $T\bar{T}$ production by the heavy gluon is determined by the 3.0 TeV BSM gluon, much heavier than any of the $m_T$ cases, and is thus less sensitive to the $T$-quark mass. Nonetheless, there are substantial kinematic differences between the $T\bar{T}$ production via a heavy gluon when compared to the standard production case. While $T\bar{T}$ is characterized by large momenta final states, FCNC is on average closer to the SM background processes in terms of $H_T$ and momenta of the final state objects.
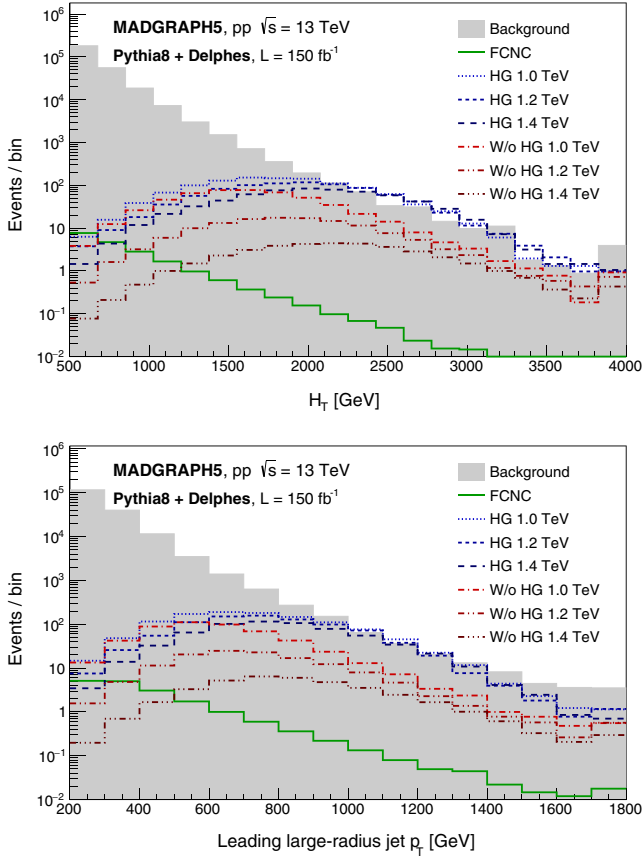
FIG. 1. Distribution of the (top) $H_T$ and (bottom) leading large-radius jet $p_T$ for total background and each signal type: $tZ$ production by FCNC, and $T\bar{T}$ production via heavy guon or without heavy gluon for $m_T = \{1.0, 1.2, 1.4\}$ TeV. The distributions are normalized to the generation cross section and to an integrated luminosity of 150 fb$^{-1}$.

## IV. DEEP NEURAL NETWORK ARCHITECTURE AND TRAINING

We train a DNN to distinguish each signal type from the set of backgrounds, using basic information constituted of the four-momenta of the reconstructed particles as provided by the Delphes simulation:

(i) $(\eta, \phi, p_T, m)$ of the five leading jets and large-radius jets;

(ii) $(\eta, \phi, p_T)$ of the two leading electrons and muons;

(iii) multiplicity of jets, large-radius jets, electrons and muons;

(iv) $(E_T, \phi)$ of the missing transverse energy (MET),

where $\phi$ is the azimuthal angle defined in the transverse plane, $\eta$ is the pseudorapidity and $m$ the invariant mass. $E_T$ is the energy in the transverse plane. Jets are built from calorimeter energy clusters grouped using the jet finder algorithm anti-$k_t$ with radius parameter $R = 0.4$ and $R = 1.0$ for jets and in large-radius jets, respectively. Additional inputs are the $H_T$ and the $N$-subjetiness of the leading large-radius jet ($\tau_N$ with $N = 1, 2, …, 5$) [23].

TABLE I. Hyperparameters used by all DNNs.

| Hyperparameter | Value |
| --- | --- |
| Hidden layers | 3 |
| Units | 352 |
| Unit activation function | Selu |
| Unit weights initializer | LeCun Normal |
| Dropout rate | 10% |
| Initial learning rate | $10^{-3}$ |
| Optimizer | Nadam |
| Maximum epochs | 1000 |

DNNs are defined by a set of parameters, called hyperparameters, which specify their architecture. Although bigger networks (those with more consecutive layers and units per layer) provide greater representational and approximating power, they can also overfit to training data and therefore lose generalization on new, unseen, data. Therefore, it is an important step in the DL workflow to assess the best configuration for the DNN for the task at hand. In order to find the best architecture, we implemented a Bayesian optimization procedure using SCIKIT-OPTIMIZE [24]. The deep neural networks themselves were implemented in KERAS [25] on top of TensorFlow [26]. Preprocessing and general data manipulation were performed using NUMPY [27], PANDAS [28], and SCIKIT-LEARN [29].

At each iteration of the Bayesian optimization loop, the performance of the network was assessed by the area under the receiver operating characteristics (ROC) curve on the valuation set. We compared the best architecture for each signal and found them to be very similar across the different signals. As a result, we fixed the architecture present in Table I for all signals, and trained a network on each signal. Training was performed on batches of 2048 events and made use of early stopping after 15 epochs without improvement, while the learning rate was reduced by a factor of 3 after five epochs without improvement. Physical weights were used as training sample weights after being classwise normalized, i.e.,

$$\sum_i^{N_b} \tilde{w}_i^b = \sum_i^{N_s} \tilde{w}_i^s, \tag{1}$$

where $\tilde{w}_i^b$ ($\tilde{w}_i^s$) and $N_b$ ($N_s$) are the training weights and number of events of the background (signal). We notice that this retains the relative weight ratios for different subsamples within each class in the sample. The value of the sum is irrelevant, as it can be absorbed into the learning rate, and we set it to 1.

## V. RESULTS OF EXCLUSION LIMITS

With the trained networks, we proceeded to compute the predictions on all test samples, each comprised of the same

background sample and a signal sample. The output is read as the probability of an event being of the signal in which the network was trained, i.e.,

$$\mathrm{DNN}_S(X_i) = p(\mathrm{S}|X_i), \tag{2}$$

where $\mathrm{DNN}_S$ is the deep neural network trained on signal $S$ and $X_i$ the reconstructed variables of the event $i$.

The prediction distributions of the network trained on the $T\bar{T}$ produced via SM with $m_T = 1.4$ TeV for the background and each test signal are shown in Fig. 2. Although the network is unequivocally more performant in classifying the signal it was trained on, it is still able to identify other signal types as nonbackground. The quality with which this is done for each signal matches what is expected from the kinematic differences of the signals, presented in Fig. 1. In Fig. 3 we present receiver operating characteristic (ROC) curves that show how the discriminating power degradates as we change the model used to predict signal, which is quite evident for FCNC signal. The bottom figure corresponds to the ROC curves of the predictions shown in Fig. 2.

These distributions are used as the inputs of OpThyLic [30], a software to compute limits using the $\mathrm{CL}_s$ method [31]. Poissonian statistical uncertainties on each bin of the distributions were included in the limit computation. From this we obtained the 95% upper limit on the signal strength, defined as the signal cross section normalized to the predicted cross section, computed with MADGRAPH5, $\mu$. This procedure was done for each combination of train and test samples and the results are presented in Table II. The central values presented in the table are used to produce the heatmap in Fig. 4. The quoted uncertainties correspond to $1\sigma$ variations around the expected limit and incorporate the statistical uncertainty on the generated signal and



FIG. 3. Receiver operating characteristic (ROC) curves for the FCNC signal as predicted by all models (up), and for all signals as predicted by the model trained on $T\bar{T}$ produced via SM with $m_T = 1.4$ TeV signal (down), where we notice that the curve for the FCNC signal is not shown with this level of zoom.

background samples. We also present the value of the limit obtained for the fit to the $H_T$ distribution as a baseline to highlight how the discriminative power of the DNNs improves the sensitivity.

As expected, we observe that the signals with stricter limits are those with lower VLT masses for each case of either SM or heavy gluon production. A more interesting observation is that the limits seem to be reasonably insensitive to which VLT signal we used to train the discriminative DNN.

The limits computed on the sample with FCNC signal show clear degradation as we use any network trained on VLT signals. This is easily understood as the FCNC does not produce new heavy states, and as such its kinematics are manifestly different from those produced by VLT signals, being instead very similar to other SM processes.

For comparison, Table II also shows the limits for each signal type obtained by fitting the $H_T$ distribution as a simpler, but commonly used [9], alternative to the use of
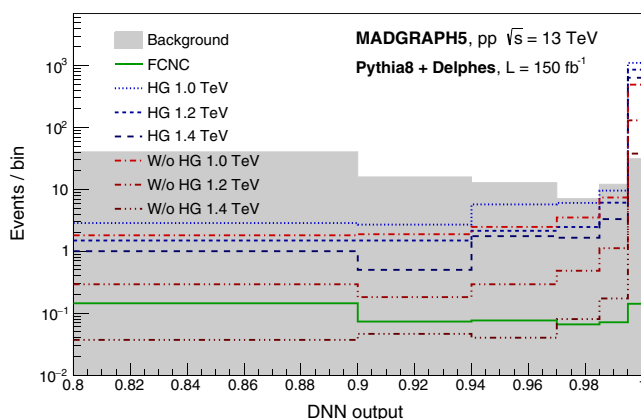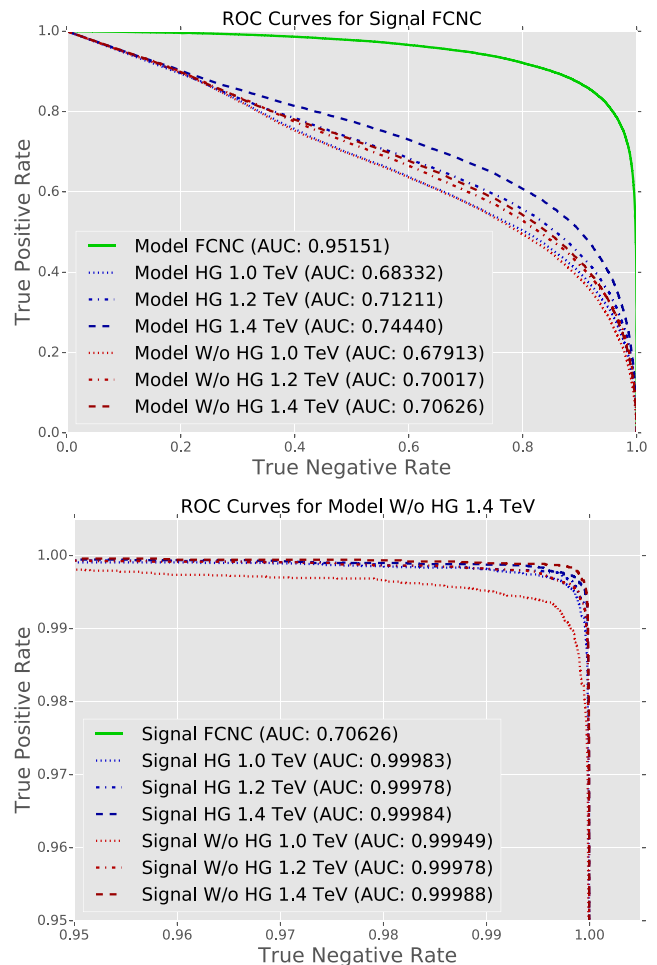


FIG. 2. Distribution of the predictions for total background and each signal type, for the network trained on the $T\bar{T}$ signal produced via SM with $m_T = 1.4$ TeV. The distributions are normalized to the generation cross section and to an integrated luminosity of 150 fb$^{-1}$.

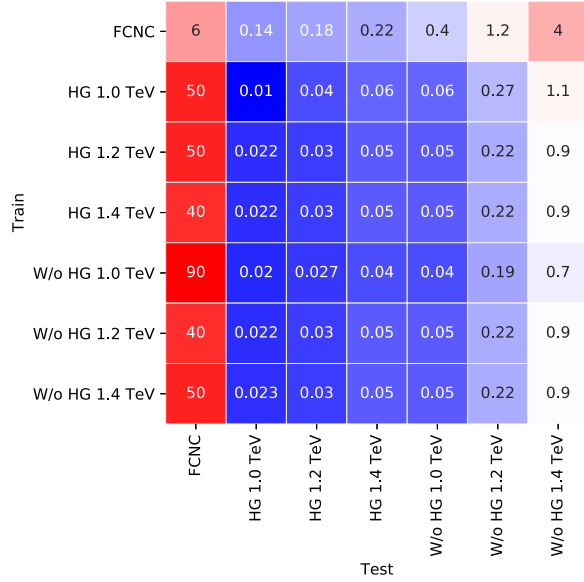|  | FCNC | HG 1.0 TeV | HG 1.2 TeV | HG 1.4 TeV | W/o HG 1.0 TeV | W/o HG 1.2 TeV | W/o HG 1.4 TeV |
|---|---|---|---|---|---|---|---|
| FCNC | 6 | 0.14 | 0.18 | 0.22 | 0.4 | 1.2 | 4 |
| HG 1.0 TeV | 50 | 0.01 | 0.04 | 0.06 | 0.06 | 0.27 | 1.1 |
| HG 1.2 TeV | 50 | 0.022 | 0.03 | 0.05 | 0.05 | 0.22 | 0.9 |
| HG 1.4 TeV | 40 | 0.022 | 0.03 | 0.05 | 0.05 | 0.22 | 0.9 |
| W/o HG 1.0 TeV | 90 | 0.02 | 0.027 | 0.04 | 0.04 | 0.19 | 0.7 |
| W/o HG 1.2 TeV | 40 | 0.022 | 0.03 | 0.05 | 0.05 | 0.22 | 0.9 |
| W/o HG 1.4 TeV | 50 | 0.023 | 0.03 | 0.05 | 0.05 | 0.22 | 0.9 |

FIG. 4. Heatmap of the central value of the limits on $\mu$ obtained for all combinations of training and testing signals, as presented in Table II.

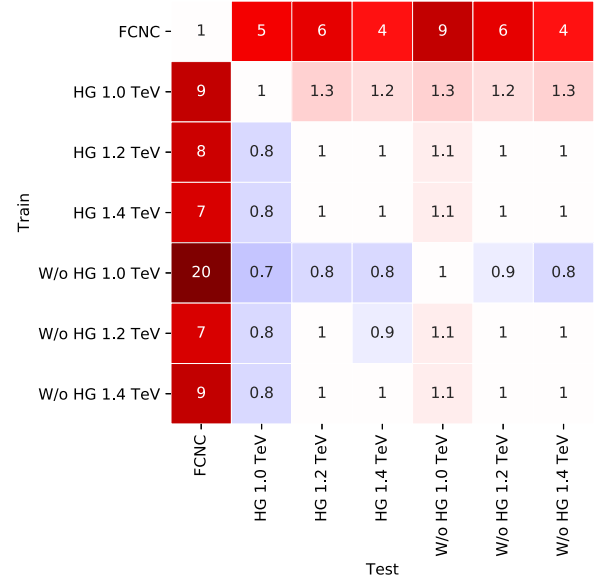|  | FCNC | HG 1.0 TeV | HG 1.2 TeV | HG 1.4 TeV | W/o HG 1.0 TeV | W/o HG 1.2 TeV | W/o HG 1.4 TeV |
|---|---|---|---|---|---|---|---|
| FCNC | 1 | 5 | 6 | 4 | 9 | 6 | 4 |
| HG 1.0 TeV | 9 | 1 | 1.3 | 1.2 | 1.3 | 1.2 | 1.3 |
| HG 1.2 TeV | 8 | 0.8 | 1 | 1 | 1.1 | 1 | 1 |
| HG 1.4 TeV | 7 | 0.8 | 1 | 1 | 1.1 | 1 | 1 |
| W/o HG 1.0 TeV | 20 | 0.7 | 0.8 | 0.8 | 1 | 0.9 | 0.8 |
| W/o HG 1.2 TeV | 7 | 0.8 | 1 | 0.9 | 1.1 | 1 | 1 |
| W/o HG 1.4 TeV | 9 | 0.8 | 1 | 1 | 1.1 | 1 | 1 |

FIG. 5. Heatmap of the central value of the normalized limits obtained for all combinations of training and testing signals, as presented in Table III.

DNN. We observe that for each signal, the limits obtained by employing a DNN are always better or compatible within the $1\sigma$ uncertainty than the ones obtained through the $H_T$ fit, regardless of the signal the model was trained on. This provides evidence that the usage of DNNs as discriminants can still provide impactful discrimination over signals not used during training.

In order to better assess the impact on the limits as we use a network trained on a different signal, we normalize the columns with respect to the value of $\mu$ obtained by the network trained on the respective signal. These results are presented in Table III. We also present the central values of this table as a heatmap in Fig. 5. With the normalized results it becomes more clear that for the VLT signals the limits are insensitive within uncertainties to the specific

signal used to train the network. Perhaps more surprisingly, the same holds when we use DNN trained on signals with heavy gluon on signals without and vice versa, as signals produced from a heavy gluon will have different kinematics to those produced from a SM gluon.

Although there is a clear degradation of the upper limits of the FCNC signal when using a network trained on VLT signals, the change of the upper limits on those signals using the FCNC trained models is not as severe. This suggests that even in a limiting case where the signal is very similar to SM processes, such as the FCNC, it can still be used to train a network that can discriminate background from other much different signals. This might indicate that, for a DNN, signals are more alike to each other than to background, which can be interpreted as follows. During

TABLE II. Upper limits on signal strength, $\mu$, from the fit to the DNN output distribution for all combinations of train and test signals, and from the fit to the $H_T$ distribution.

|  |  |  |  | Test | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  | HG | | | No HG | | |
|  |  |  | FCNC | HG, 1.0 TeV | HG, 1.2 TeV | HG, 1.4 TeV | 1.0 TeV | 1.2 TeV | 1.4 TeV |
| Train | FCNC |  | $6^{+2}_{-2}$ | $0.14^{+0.07}_{-0.04}$ | $0.18^{+0.08}_{-0.06}$ | $0.22^{+0.10}_{-0.06}$ | $0.4^{+0.2}_{-0.1}$ | $1.2^{+0.5}_{-0.4}$ | $4^{+1}_{-2}$ |
|  | HG | 1.0 TeV | $50^{+20}_{-20}$ | $0.03^{+0.01}_{-0.01}$ | $0.04^{+0.02}_{-0.01}$ | $0.06^{+0.04}_{-0.02}$ | $0.06^{+0.03}_{-0.02}$ | $0.27^{+0.15}_{-0.09}$ | $1.1^{+0.6}_{-0.3}$ |
|  |  | 1.2 TeV | $50^{+20}_{-20}$ | $0.022^{+0.011}_{-0.007}$ | $0.03^{+0.02}_{-0.01}$ | $0.05^{+0.03}_{-0.02}$ | $0.05^{+0.02}_{-0.02}$ | $0.22^{+0.11}_{-0.07}$ | $0.9^{+0.5}_{-0.3}$ |
|  |  | 1.4 TeV | $40^{+20}_{-10}$ | $0.022^{+0.012}_{-0.007}$ | $0.03^{+0.02}_{-0.01}$ | $0.05^{+0.03}_{-0.02}$ | $0.05^{+0.02}_{-0.02}$ | $0.22^{+0.11}_{-0.07}$ | $0.9^{+0.5}_{-0.3}$ |
|  | No HG | 1.0 TeV | $90^{+50}_{-30}$ | $0.020^{+0.010}_{-0.007}$ | $0.027^{+0.014}_{-0.009}$ | $0.04^{+0.02}_{-0.01}$ | $0.04^{+0.03}_{-0.01}$ | $0.19^{+0.09}_{-0.07}$ | $0.7^{+0.4}_{-0.2}$ |
|  |  | 1.2 TeV | $40^{+20}_{-10}$ | $0.022^{+0.011}_{-0.007}$ | $0.03^{+0.02}_{-0.01}$ | $0.05^{+0.02}_{-0.02}$ | $0.05^{+0.02}_{-0.02}$ | $0.22^{+0.11}_{-0.07}$ | $0.9^{+0.4}_{-0.3}$ |
|  |  | 1.4 TeV | $50^{+20}_{-20}$ | $0.023^{+0.012}_{-0.008}$ | $0.03^{+0.02}_{-0.01}$ | $0.05^{+0.03}_{-0.02}$ | $0.05^{+0.02}_{-0.02}$ | $0.22^{+0.11}_{-0.08}$ | $0.9^{+0.5}_{-0.3}$ |
| Fit to $H_T$ distribution |  |  | $90^{+40}_{-20}$ | $0.11^{+0.04}_{-0.04}$ | $0.11^{+0.05}_{-0.03}$ | $0.12^{+0.05}_{-0.04}$ | $0.3^{+0.1}_{-0.1}$ | $0.8^{+0.3}_{-0.2}$ | $1.7^{+0.7}_{-0.5}$ |

TABLE III.   Normalized limits obtained for all combinations of training and testing signals.

| | | | | Test | | | | | |
| | | | | HG | | | No HG | | |
| | | FCNC | HG, 1.0 TeV | HG, 1.2 TeV | HG, 1.4 TeV | 1.0 TeV | 1.2 TeV | 1.4 TeV |
|---|---|---|---|---|---|---|---|---|
| Train | FCNC | $1.0^{+0.4}_{-0.3}$ | $5^{+2}_{-2}$ | $6^{+2}_{-2}$ | $4^{+2}_{-1}$ | $9^{+4}_{-3}$ | $6^{+2}_{-2}$ | $4^{+2}_{-1}$ |
| HG 1.0 TeV | | $9^{+4}_{-3}$ | $1.0^{+0.5}_{-0.3}$ | $1.3^{+0.7}_{-0.4}$ | $1.2^{+0.6}_{-0.4}$ | $1.3^{+0.7}_{-0.4}$ | $1.2^{+0.6}_{-0.4}$ | $1.3^{+0.7}_{-0.4}$ |
| 1.2 TeV | | $8^{+4}_{-2}$ | $0.8^{+0.4}_{-0.2}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.1^{+0.5}_{-0.4}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ |
| 1.4 TeV | | $7^{+3}_{-2}$ | $0.8^{+0.4}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.1^{+0.6}_{-0.4}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.4}$ |
| No HG 1.0 TeV | | $20^{+9}_{-5}$ | $0.7^{+0.4}_{-0.2}$ | $0.8^{+0.4}_{-0.3}$ | $0.8^{+0.4}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $0.9^{+0.4}_{-0.3}$ | $0.8^{+0.4}_{-0.3}$ |
| 1.2 TeV | | $7^{+3}_{-2}$ | $0.8^{+0.4}_{-0.2}$ | $1.0^{+0.5}_{-0.3}$ | $0.9^{+0.5}_{-0.3}$ | $1.1^{+0.5}_{-0.4}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ |
| 1.4 TeV | | $9^{+4}_{-3}$ | $0.8^{+0.4}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.1^{+0.6}_{-0.3}$ | $1.0^{+0.5}_{-0.3}$ | $1.0^{+0.5}_{+0.3}$ |

training a DNN will learn a decision function that defines a separation hypersurface between the regions supporting signal and background events in the high dimensional input space. While different signals are supported at different regions the background is common to all of them, which leads to each DNN trained on a specific signal to learn similar ways of isolating large parts of the background. It is this shared learned representational capacity of the background that is being transferred when we apply a trained DNN on a new signal.

## VI. CONCLUSION

In this work we set out to explore the transferability of DNN trained to discriminate between signal and background using reconstructed physical observables. Here we approached the notion of transferability as how a model trained on a given signal performs when discriminating between signal and background on a different signal. The different signals fell under three classes: a FCNC model without new heavy states, VLT states produced by SM gluon, and VLT states produced by a new heavy gluon. For the latter two cases we considered three different points of the parameter space where we varied the masses of the VLT states.

We have shown that the upper limits for the physical models with the VLT states are insensitive within uncertainties to which network discriminator we trained, as long as it was trained on a signal with VLT states. These results highlight that discriminative DNN are highly transferable across signals with similar new heavy states across different points of their parameter space. It should also be noted that the same transferability is observed between signals corresponding to significantly different models, where important kinematic differences between the reconstructed final state objects are expected.

Furthermore, we showed that networks trained in any signal can provide similar or greater sensitivity than

performing limits on the fit to the $H_T$ distribution, demonstrating the aforementioned transferability of a trained DNN to new signals. More specifically, we showed that the FCNC signal, being the most different of the seven, can still be used to train a DNN that provides discriminative power to other signals that are kinematically very different. This suggests that in their training, DNNs learn features that are able to identify background, which is the same for all new physics. Ultimately, this means that the usage of DNNs in the discriminating step of an analysis might help to find new physics not assumed in its training. This fosters the idea that deep learning may contribute to a novel framework for generic searches, providing a powerful way to increase the sensitivity of searches for new physics phenomena at colliders. Future studies, beyond the scope of the present paper, can be done to verify the obtained results using detailed simulations of the LHC experiments. Also, it would be interesting to compare DNNs with other machine learning methods, also in terms of transferability of the results.

[1] D. Guest, K. Cranmer, and D. Whiteson, Deep learning and its application to LHC physics, Annu. Rev. Nucl. Part. Sci. **68,** 161 (2018).

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[3] J. Yosinski *et al.*, How transferable are features in deep neural networks? in *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (Curran Associates, Inc., 2014), pp. 3320–3328.

[4] ATLAS Collaboration, Search for flavor-changing neutral current top-quark decays $t \rightarrow qZ$ in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, J. High Energy Phys. 07 (2018) 176.

[5] CMS Collaboration, Search for Flavor-Changing Neutral Currents in Top-Quark Decays $t \rightarrow Zq$ in $pp$ Collisions at $\sqrt{s} = 8$ TeV, Phys. Rev. Lett. **112,** 171802 (2014).

[6] CMS Collaboration, Search for associated production of a Z boson with a single top quark and for tZ flavor-changing interactions in pp collisions at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 07 (2017) 003.

[7] C. Degrande, F. Maltoni, J. Wang, and C. Zhang, Automatic computations at next-to-leading order in QCD for top-quark flavor-changing neutral processes, Phys. Rev. D **91,** 034024 (2015).

[8] G. Durieux, F. Maltoni, and C. Zhang, Global approach to top-quark flavor-changing interactions, Phys. Rev. D **91,** 074017 (2015).

[9] ATLAS Collaboration, Search for pair and single production of vectorlike quarks in final states with at least one Z boson decaying into a pair of electrons or muons in $p\bar{p}$ collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV, Phys. Rev. D **98,** 112010 (2018).

[10] CMS Collaboration, Search for vector-like quarks in events with two oppositely charged leptons and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C **79,** 364 (2019).

[11] M. Buchkremer, G. Cacciapaglia, A. Deandrea, and L. Panizzi, Model independent framework for searches of top partners, Nucl. Phys. **B876,** 376 (2013).

[12] J. A. Aguilar-Saavedra, Identifying top partners at LHC, J. High Energy Phys. 11 (2009) 030.

[13] D. B. Kaplan and H. Georgi, SU(2) × U(1) breaking by vacuum misalignment, Phys. Lett. **136B,** 183 (1984).

[14] D. B. Kaplan, Flavor at SSC energies: A new mechanism for dynamically generated fermion masses, Nucl. Phys. **B365,** 259 (1991).

[15] J. P. Araque, N. F. Castro, and J. Santiago, Interpretation of vector-like quark searches: Heavy gluons in composite Higgs models, J. High Energy Phys. 11 (2015) 120.

[16] J. P. Araque, N. F. Castro, and J. Santiago, Interpretation of vector-like quark searches: The case of a heavy gluon in composite Higgs models and vector-like quarks, Proc. Sci., TOP2015 (**2015**) 069.

[17] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.

[18] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to PYTHIA8.2, Comput. Phys. Commun. **191,** 159 (2015).

[19] V. Khachatryan *et al.*, Event generator tunes obtained from underlying event and multiparton scattering measurements, Eur. Phys. J. C **76,** 155 (2016).

[20] R. D. Ball *et al.*, Parton distributions with LHC data, Nucl. Phys. **B867,** 244 (2013).

[21] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, DELPHES3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.

[22] The transverse plane is defined with respect to the proton colliding beams.

[23] S. Marzani, G. Soyez, and M. Spannowsky, Looking inside jets: An introduction to jet substructure and boosted-object phenomenology, *Lecture Notes Physics* (Springer, Cham, 2019), p. 958.

[24] T. Head *et al.*, scikit-optimize/scikit-optimize: v0.5.2, 2018.

[25] F. Chollet *et al.*, Keras, https://keras.io (2015).

[26] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, https://www.tensorflow.org.

[27] T. Oliphant, *NumPy: A Guide to NumPy* (Trelgol Publishing, 2006).

[28] W. McKinney *et al.*, Data structures for statistical computing in python, in *Proceedings of the 9th Python in Science Conference* (2010), Vol. 445, pp. 51–56.

[29] F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12,** 2825 (2011).

[30] E. Busato, D. Calvet, and T. Theveneaux-Pelzer, OpTHyLiC: An optimized tool for hybrid limits computation, Comput. Phys. Commun. **226,** 136 (2018).

[31] A. L. Read, Presentation of search results: The CL(s) technique, J. Phys. G **28,** 2693 (2002).