Dear Editor,

We would like to thank the referee for reviewing this paper, giving very useful comments and furnishing it. We have carefully considered all comments, and we have applied several changes to the original version of the paper to address the issues raised.

Detailed responses to all the comments can be found below. The referee comments are in **black**, our replies are in **blue**. List of changes are given in **green**.

We are at your disposal for any further clarifications and/or additional information.

Sincerely,
Authors.

1. There has been another study proposing to improve the search for top-quark FCNCs using neural-nets from Aguilar-Saavedra and Branco in 2000 (hep-ph/0004190). This paper only considers scalar contributions and does not consider the tri-lepton OS final state, but should still be mentioned.

   Thanks to the referee for drawing our attention to the reference. It is added in the manuscript within the following paragraph.

   An early study explored the use of neural networks to search for top-quark FCNCs mediated by scalar interactions, demonstrating the potential of machine learning techniques, although it focused on different final states, as detailed in Ref. [35].

2. In section 4.1, there needs to be much more quantitative detail on the neural network architecture (number of of layers, number of neurons in each layer), the loss function used (included as an equation maybe), the optimiser chosen, the learning rate, the dropout layers and dropout rate, activation functions, etc. They should also include any data preprocessing here, e.g. re-scalings. The results should be in principle reproducible by the reader.

   Based on the referee's recommendation:

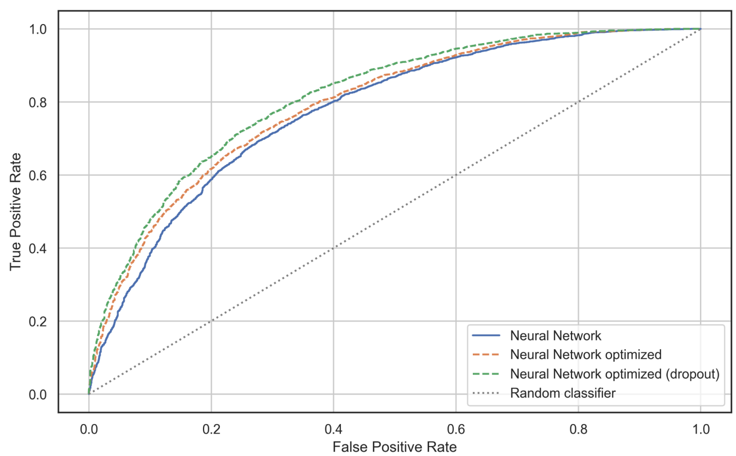
   The neural network used in this analysis consists of an input layer followed by three hidden layers, each containing 128 neurons. The scaled exponential linear unit (SELU) activation function is employed in all hidden layers to promote self-normalizing behavior. To mitigate overfitting, dropout regularization is applied after each hidden layer with a dropout rate of 10%. The output layer uses a sigmoid activation function to produce a binary classification score. The network is trained using the Adam optimizer with the learning rate provided by the Keras library (0.001), and the loss function is binary cross-entropy, defined as [55]:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{4}$$

where y_i and $\hat{y}_i$ denote the true and predicted labels, respectively. Prior to training, all input features are standardized using a StandardScaler transformation to zero mean and unit variance.

3. On the ROC plot, you should consider replacing 'Luck' with something like 'Random classifier', or explain what it means in the text.

   Thanks to the referee for the comment. It is corrected and plot has been replaced in the paper. It is also presented below:



4. The accuracy and the AUC for these neural-net classifiers are not presented, they should be either in the text or in a table. The accuracy of one of the neural nets is mentioned in 4.2, but only in the discussion of the other models.

   Done in comment 5.

5. The ROC curves show good performance, but robustness is mentioned later in the paper, and one way to demonstrate robustness in neural-net performance is to train more than one classifier per task and compute the average performance +- std dev. I would recommend doing this here for the AUC and accuracy since the performances of these networks are quite close.

We appreciate the referee for the comment. The table is corrected/added in the paper according to the referee’s request. The mean and std are derived after running each model ten times over the whole training set.

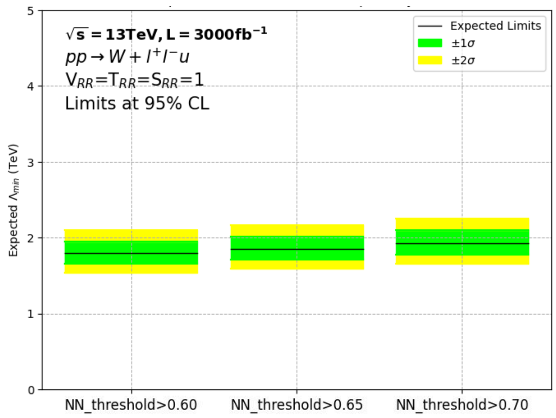
Model	AUC	Accuracy
Baseline DNN	0.7805 ± 0.0069	0.7310 ± 0.0040
DNN with Dropout	0.8102 ± 0.0064	0.7537 ± 0.0019
DNN with Optimizer Tuning	0.7958 ± 0.0025	0.7427 ± 0.0009

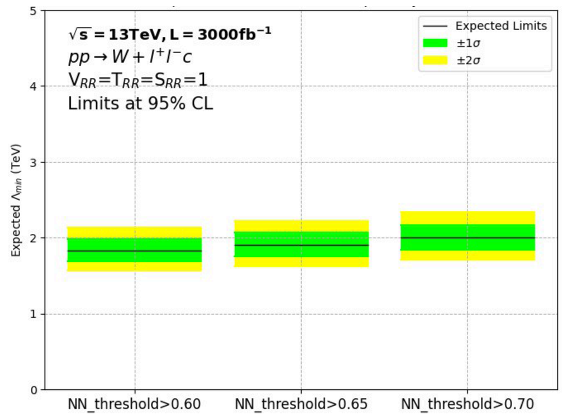
6. What is the motivation for using the AMS (eq. 4) to choose the working point? A well-motivated choice of metric for choosing the working-point is usually the significance improvement, as discussed in 1909.03081. You should consider this, or motivate why using AMS is more suitable.

   We acknowledge the referee's point regarding the choice of metric for optimizing the working point and the common utility of significance improvement metrics, such as those discussed in arXiv:1909.03081. In our analysis, which employs machine learning classifiers to distinguish signal from background, we selected the Approximate Median Significance (AMS) to determine the optimal threshold on the classifier's output. The AMS metric, AMS=2((TPR+FPR+br)ln(1+FPR+brTPR)−TPR), is designed to balance signal efficiency (represented by the True Positive Rate, TPR) against background contamination (related to the False Positive Rate, FPR). A key feature of the AMS metric is the inclusion of a regularization term, , which ensures stability and prevents the selection of overly narrow regions, particularly in scenarios with low background statistics that can arise after applying the stringent selections often achieved with powerful classifiers. This regularization is crucial for robustly optimizing the working point. The AMS metric was notably utilized in the context of the Higgs boson discovery analyses and provides a well-established and practical approach for maximizing the expected sensitivity in searches for new phenomena by optimizing the trade-off between retaining signal events and rejecting background. While direct optimization of significance improvement is a valid approach, AMS offers a robust and widely adopted method for defining the signal region in analyses like ours, particularly when dealing with the continuous output of a machine learning model.

7. In addition to the previous point, different terminology for the classifier working point is used at different points in the paper, e.g. threshold and NN_weight. NN_weight (p. 12) is misleading because the network weights are what are optimised during the network training. Something consistent should be used throughout.

Thanks to the referee for the comment. All NN_weight terms are changed to NN_threshold like the plot below for tW channel.

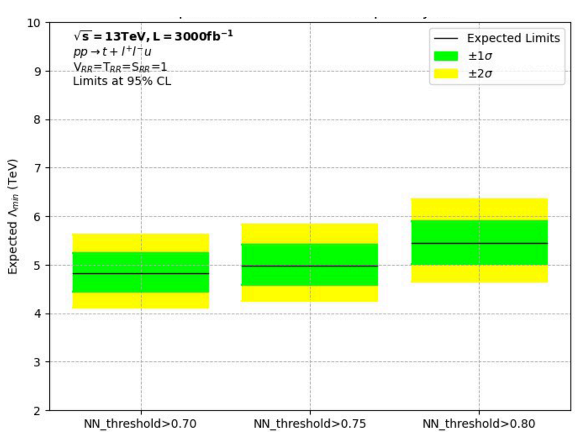


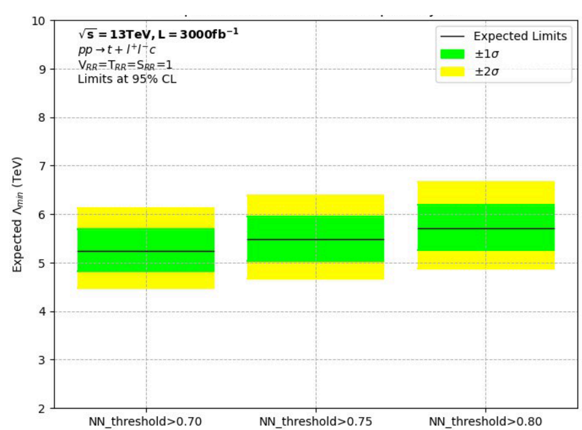


8. On p. 12 you say that the improvement in signal and background separation as you increase the classifier threshold/working-point gives greater sensitivity in distinguishing signal and background. But in fig 10 you show that the constraints on the NP scale for tW is relatively unaffected by moderate changes in this threshold, you should show the same plot and have a similar discussion for ttbar.

   This part is added based on the referee's recommendation:

   As shown below, for both the tuee and tcee operators in the $t\bar{t}$ channel, increasing the neural network threshold from 0.70 to 0.80 yields a modest but consistent improvement in the expected limits on the NP scale Λmin. While the central values increase slightly across the thresholds, the associated ±1σ and ±2σ bands remain stable, indicating improved signal-to-background separation without a significant increase in statistical uncertainty. This trend aligns with expectations: higher NN thresholds enhance purity, reducing background contamination and improving sensitivity.





9. The results for the two different NP couplings tuee and tcee are quite similar, but not exactly the same. Where do these differences arise?

   The observed differences between the tuee and tcee results primarily arise from both detector-level effects and subtle kinematic differences in the decay phase space.
   From a detector perspective, the c-quark jet is more likely to be mistagged as a b-jet compared to the u-quark jet. As specified in the Delphes card used in our simulation, the mistag rate for c-jets depending on the jet p_T exceeds 10%, whereas for u-jets it is closer to 1%. Since our selection criteria require exactly one b-tagged jet in the final state, this higher mistag probability for c-jets means that a larger fraction of $t \to ce^+e^-$ signal events pass the selection cuts compared to the $t \to ue^+e^-$ case. This effect is reflected in the post-selection signal event yields shown in Table 2.
   On the kinematic side, although both u and c quark masses are negligible compared to the top quark mass, the fact that m□ > m_u leads to small but non-negligible differences in the available phase space for the decay $t \to q\, e^+e^-$ (with q = u or c). These differences impact the momentum distributions and angular correlations of the final-state particles. As our deep neural networks are trained independently for tuee and tcee using datasets that reflect these differences, the learned decision boundaries; and consequently, the classification performance and extracted NP limits; show small discrepancies between the two channels. These are visible in the AUC values and in the slightly different limits on the NP scale Λ obtained for each scenario.

10. The benefits of the neural-net approach as opposed to the 'crude' traditional approach should be demonstrated by providing an example of the sensitivity to the NP scales obtained using the much simpler cut and count based approach on high-level observables.

    In revised version of the manuscript, to demonstrate the benefits of NN approach, we applied a simple cut-and-count approach and defined the signal region using the most powerful variables in signal-background discrimination like:

    150 GeV < M_{top}^{non-SM} < 200 GeV

    Njet >= 3

    And then the manuscript is updated by adding:

To benchmark the performance of our machine learning strategy, we performed a traditional cut-and-count analysis using high-level observables and a manually defined signal region (150 GeV < M_{non-SM} < 200 GeV and $N_{jet} \geq 3$). The resulting signal yields were 98.27 for the $_{top}$ tcee operator and 105.30 for tuee, with 98.53 background events (Table 4). This yields an NP scale sensitivity of approximately 3.6 TeV at 95% CL for both channels. While this demonstrates that even simple selection criteria can yield meaningful results, the deep neural network approach improves background suppression and signal discrimination, ultimately leading to stronger limits on the NP scale, surpassing the cut-and-count baseline.

Process	$\mathcal{N}_S$	$\mathcal{N}_B$	$\mathcal{N}_S/\mathcal{N}_B$
$pp \to t\bar{t} \to ue^-e^+\bar{t}$	105.3	98.5	1.07
$pp \to t\bar{t} \to ce^-e^+\bar{t}$	98.3	98.5	1.00