2024 CERN Summer Student Project Report on

A Mass-Decorrelated Event Classifier for a Search for $\mathrm{t\bar{t}H}(\to \mathrm{c\bar{c}})$

Yuval Kay

Mathematical Institute

University of Oxford

Supervised by

Dr Huilin Qu

The CMS Collaboration

CERN

Dr Sebastian Wuchterl

The CMS Collaboration

CERN

Abstract

The following study contributes to the search for the decay of the Higgs boson into a charm quark-antiquark pair, $\mathrm{H \to c\bar{c}}$, produced in association with a top quark-antiquark ($\mathrm{t\bar{t}H}$). We analyse the fully hadronic, semileptonic and dileptonic channels to cover all decay modes and validate by measuring the $\mathrm{t\bar{t}Z}$ processes. To capture correlations within the events, we employ the Particle Transformer architecture for event classification. A significant challenge, however, is that the classifier's performance is correlated with the Higgs boson mass, which can distort the background dijet mass distribution and resemble a signal. Therefore, the goal of this project is to develop a mass-independent event classifier.

Contents

Acknowledgement

I would like to expresses my sincere gratitude for my supervisors - Dr Huilin Qu and Dr Sebastian Wuchterl for giving me the opportunity to spend the summer here at CERN. Their continuous support throughout the past 13 weeks has been invaluable.

1 Introduction

The discovery of the Higgs boson by both the ATLAS [1] and CMS [2] experiments at CERN LHC in 2012 was a momentous advancement towards understanding the electroweak symmetry-breaking mechanism. Studies have shown that the interactions between the Higgs boson and third-generation fermions are in agreement with the theoretically predicted standard model values [3–8]. In 2019, CMS reported direct evidence of the Higgs boson decaying into a pair of muons [9], opening the door for the observations of Higgs boson interactions with second-generation fermions. Hence, the next important target in Higgs physics is the measurement of the coupling to second-generation quarks.

Motivated by the fact that the coupling strength of the Higgs boson to fermions is proportional to their mass, across the second-generation of quarks, we begin by searching for the decay of the Higgs boson into a charm quark-antiquark pair, $\mathrm{H} \to \mathrm{c\bar{c}}$. However, a direct measurement of this process is highly challenging. Foremost, the branching fraction for the $\mathrm{c\bar{c}}$ decay channel, by SM calculations, $\mathcal{B}(\mathrm{H} \to \mathrm{c\bar{c}}) = 2.89\%$ [10] is a factor of 20 smaller than that of Higgs boson into a bottom quark-antiquark pair. Furthermore, it suffers from an extensive background from the QCD production of a $\mathrm{c\bar{c}}$ final state and poor resolution of the jets originating from the charm quarks, making the peak at the dijet invariant mass challenging to resolve.

Previous direct searches for the $\mathrm{H} \to \mathrm{c\bar{c}}$ decay channel involved $\mathrm{VH}(\to \mathrm{c\bar{c}})$, in which a Higgs boson is produced in association with a vector boson (W or Z). Despite the lower cross-section, the distinctive features of the leptonic decays of the vector bosons significantly reduce the background, giving the strictest limit to date on $\sigma(\mathrm{VH})\mathcal{B}(\mathrm{H} \to \mathrm{c\bar{c}})$ at 14 times the standard model prediction at 95% confidence level [11]. Another analysis involved searching for boosted $\mathrm{ggH}(\to \mathrm{c\bar{c}})$, in which the transverse momentum of the Higgs boson produced is greater than 450 GeV, giving an upper limit on $\sigma(\mathrm{ggH})\mathcal{B}(\mathrm{H} \to \mathrm{c\bar{c}})$ at 47 times the standard model prediction at 95% confidence level [12].

Influenced by the significance of the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ channel in previous analyses by the CMS Collaboration for probing the Higgs-bottom coupling [8], this report presents a search for $\mathrm{H} \to \mathrm{c\bar{c}}$, where the Higgs boson is produced in association with a top-antitop pair. Despite the lower cross-section of $\mathrm{t\bar{t}H}$ compared to VH [13], the $\mathrm{t\bar{t}H}$ process has a distinct signature from the top quark decays into high-energy jets and leptons (through $\mathrm{t} \to \mathrm{Wb} \to \mathrm{q\bar{q}b}$ or $\mathrm{t} \to \mathrm{Wb} \to \ell\bar{\nu}_\ell \mathrm{b}$), which allow for better discrimination against a large background from standard model $\mathrm{t\bar{t}}$ production with additional heavy flavour jets, thereby reaching comparable sensitivity. By categorising events based on the presence of 0, 1, or 2 leptons, we investigate the fully hadronic, semileptonic, and dileptonic channels to encompass all possible

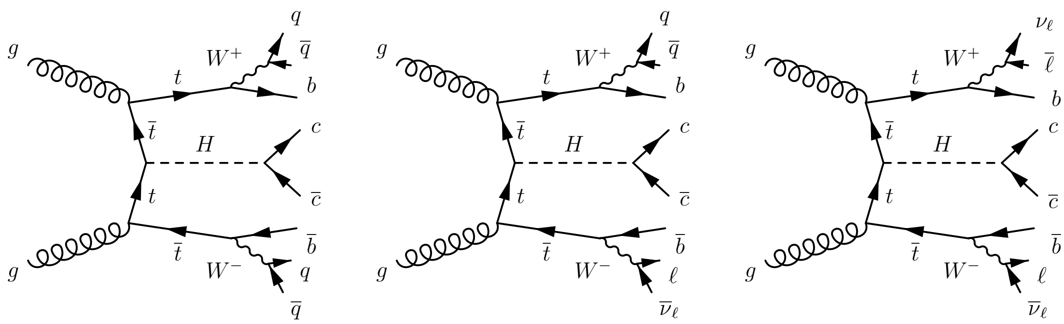

Figure 1: Leading-order Feynman diagrams for the associate production of a Higgs boson with a top quark-antiquark pair. (left) fully hadronic channel, (center) single leptonic channel, (right) dileptonic channel.

decay modes of the top-antitop pair [Fig 1]. This report primarily focuses on the dileptonic channel.

We utilise the ParticleTransformer architecture [14] - a transformer network classifier to capture correlations among objects within an event, thereby enabling to differentiate effectively between the $\mathrm{H} \to \mathrm{c\bar{c}}$ signal from backgrounds. One key challenge, however, is that the classifier's performance is correlated with the Higgs boson's mass since its response is to categorise events as belonging to the target $\mathrm{t\bar{t}H}$ signal. After applying selection criteria, this can lead to distortions in the background mass distribution, potentially mimicking a signal. Therefore, the goal of this project is to develop a mass-decorrelated event-classifier. This report also discusses the implementation of mass-decorrelation techniques on regression and segmentation models.

1.1 Particle Transformer Architecture

The Particle Transformer (ParT) architecture is a Transformer-based model designed for event reconstruction. An overview of the ParT architecture is presented in the appendix 5. Initially, ParT represents jets as particle clouds, which are unordered, permutation-invariant sets of particles. For each particle i, the corresponding feature vector $\mathbf{x}_i$ includes kinematic variables (the 4-vector (E, p_x, p_y, p_z)) and particle identifications (e.g. charge). These features form an array $\mathbf{X} \in \mathbb{R}^{N\times C}$ where N is the number of particles in a jet and C is the number of features. Then, for every pair of particles (i, j), C' features are collected in a matrix $\mathbf{U} \in \mathbb{R}^{N\times N\times C'}$. While various features can be used, in our analysis, we compute the following interaction features:

$$\begin{aligned}
\Delta &= \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \\
k_T &= \min(p_{T,i}, p_{T,j})\Delta \\
z &= \min(p_{T,i}, p_{T,j})/(p_{T,i} + p_{T,j}) \\
m^2 &= (E_i + E_j)^2 - \|\mathbf{p}_i + \mathbf{p}_j\|^2
\end{aligned}$$

where y is the rapidity, ϕ is the azimuthal angle, p_T is the transverse momentum, and $\mathbf{p}$ is the momentum 3-vector. We take the logarithm $(\ln \Delta, \ln k_T, \ln z, \ln m^2)$ to avoid the long-tail of the distributions. By including both particle-specific and pairwise-interaction features we allow the model to capture both local and global structures within a jet.

Each particle's feature vector $\mathbf{x}_i$ is embedded into a d-dimensional space through an multi-layer perceptron (MLP) forming a particle embedding matrix $\mathbf{X}_0 \in \mathbb{R}^{N\times d}$. Pairwise-interactions $\mathbf{U}$ are also projected by an MLP to d'-dimensional embedding, $\mathbf{U}' \in \mathbb{R}^{N\times N\times d'}$. The particle embedding $\mathbf{X}_0$ is fed into a network of L particle attention blocks. For each particle embedding $\mathbf{X}_\ell$ at layer $\ell \in [0, L]$, a modified multi-head self-attention mechanism (MHA) updates the particle representation by considering the interaction $\mathbf{U}'$ between particles. The particle MHA mechanism computes attention as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \mathbf{U}'\right) V$$

where $Q, K \in \mathbb{R}^{N\times d_k}$, and $V \in \mathbb{R}^{N\times d_v}$ are linear projections of the particle embedding, and d_k is the dimension of the query/key vectors. The output of the attention mechanism is passed through residual connections, and normalised via LayerNorm. The key difference in ParT is the incorporation of an additional bias term, $\mathbf{U}'$, into the pre-softmax attention weights. By integrating particle interactions into the MHA mechanism, the scaled dot-product attention weights are modified to account for physical particle relationships. This enhances the expressiveness of the attention mechanism, allowing it to better capture the underlying physics.

Lastly, the final particle embedding $\mathbf{X}_L$ is fed into a network of two class attention blocks. During training, a global class token, x_{class}, is learned and used to extract information for jet classification by attending to all particles. The attention is computed using a standard MHA with $Q = W_q x_{\text{class}} + b_q$, $K = W_k\mathbf{z} + b_k$ and $V = W_v\mathbf{z} + b_v$ as inputs, where $\mathbf{z} = [x_{\text{class}}, \mathbf{X}_L]$ is the concatenation of the class token and the final particle embedding. The resulting class token is then passed to a single-layer MLP, followed by a softmax, to yield a set of final classification scores that can be used for physical event categorisation.

2 Mass-Decorrelation

As illustrated in the ParT architecture (1.1), the mass is embedded within the particle feature vector and utilised during the calculation of the pairwise interaction matrix through the 4-momentum. Furthermore, the classifier's correlation with the Higgs boson's mass also arises as an artifact of the training process itself. Since the Standard Model $\mathrm{t\bar{t}H}$ signal corresponds to events with a specific Higgs boson mass,

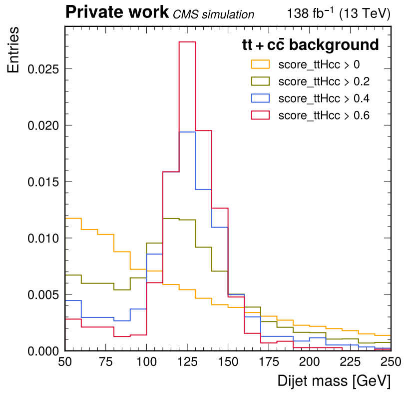

Figure 2: Normalised mass distribution of the dijet system for the $t\bar{t} + c\bar{c}$ event background. We can see that as we take tighter cuts on the discriminating score `score_ttHcc` (which is mass-dependent) the background distribution no longer appears smooth but starts to peak - resembling a signal.

the network learns to associate this mass feature with signal events and distinguish them from background events lacking this characteristic mass. Consequently, after applying a selection based on the classifier's output, distortions can occur in the background mass distribution, potentially mimicking a signal. Mass decorrelation helps avoid this problem by ensuring that the discriminating score is not correlated with the mass. This approach ensures that the background mass distribution remains smooth even at higher score thresholds, making any peaks in the mass distribution more likely to represent true signals. This, in turn, enables the use of the mass variable in further analyses.

To achieve mass independence, we train the model using Monte Carlo (MC) samples of the standard model $t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}$ events. Additionally, we generate $t\bar{t}H$ signal samples with masses ranging from 50 to 200 GeV, in 10 GeV steps. Then, in the training we sample the events so that the neural network sees a flat distribution in mass. This variation allows the model to learn to detect jets emerging from the Higgs boson across a range of masses, rather than only at the standard model mass. As a result, the model's classification becomes mass-decorrelated.

Motivated by the event topology of the dileptonic channel, possible sources of background, $t\bar{t}Z$ processes, the base ParT model is trained to classify events into 10 categories. These include two categories for the $t\bar{t}H$ signal: $t\bar{t}H(c\bar{c})$ and $t\bar{t}H(b\bar{b})$. For the $t\bar{t}Z$ background, there are three categories: $t\bar{t}Z(c\bar{c})$, $t\bar{t}Z(b\bar{b})$, and $t\bar{t}Z(q\bar{q})$, where q represents any light-flavor quark. Additionally, there are five categories for the multijet background - defined based on the additional presence of the quarks in the $t\bar{t}$ event sample: $t\bar{t} + b$, $t\bar{t} + b\bar{b}$, $t\bar{t} + c$, $t\bar{t} + c\bar{c}$, and $t\bar{t} + LF$, where LF stands for light-flavor jets.

Hyperparameter	**Value**	**Information**
`pair_input_dim`	4	Input dimension for pairwise interaction features
`remove_self_pair`	False	Whether to remove particle self-pairs
`use_pre_norm_pair`	False	Apply normalisation before pairwise computations
`embed_dims`	(128, 128, 128)	Embedding dimensions for particles features
`pair_embed_dims`	(64, 64, 64)	Embedding dimensions for pairwise features
`num_heads`	8	Number of heads in particle multi-head self attention
`num_layers`	8	Number of particle attention blocks
`num_cls_layers`	2	Number of class attention blocks
`activation`	GELU	Use Gaussian Error Linear Unit for activation function

Table 1: Hyperparameters for Particle Transformer Architecture

2.1 Model Setup

The ParT network architecture remains unchanged, and the hyperparameters used are listed in the Table. 1. We used k-fold cross-validation approach in the ParT training, which involves splitting the dataset into k subsets. In this method, we train the model on 80% of the events from each fold and validate it on the remaining 20% events of the fold, repeating the process k times so that each fold serves as the training-validation set exactly once. By combining the performance metrics across all folds, we gain a more reliable estimate of the model's classification performance, avoid any overfitting, as well as allow to use the full data sample in both the training and prediction stages. Overall, the model was trained for 20 epochs, 5-fold cross-validation, a starting learning rate of 10^{-4}, Automatic Mixed Precision (AMP) employed, and the standard binary cross-entropy (BCE) loss,

$$\mathrm{BCE}(p,t) = -t\ln(p) - (1-t)\ln(1-p)$$

where t is the true class distribution and p is the predicted class distribution.

Mass variable `mixedMassVar`

In this analysis, we consider the reconstruction-level mass variables: `mass_minDR_b`, `mass_minDR_c`, and `mass_minDR_bc`. The variable `mass_minDR_b` refers to the invariant mass of the two b-tagged jets, that are closest in ΔR. Similarly, `mass_minDR_c` and `mass_minDR_bc` represent the invariant mass for c-jets and the combination of b- and c-jets, respectively. The variation of these mass variables across all signal classes is presented in Figure 3. As expected, we observe that processes involving c-quarks primarily resonate in the `mass_minDR_c` distribution, in contrast to the `mass_minDR_b` distribution, which is dominated by b-quark processes. However, we see that `mass_minDR_bc` distribution does not produce a distinct peak around the Higgs mass. Hence, we discontinue its use in further analysis.

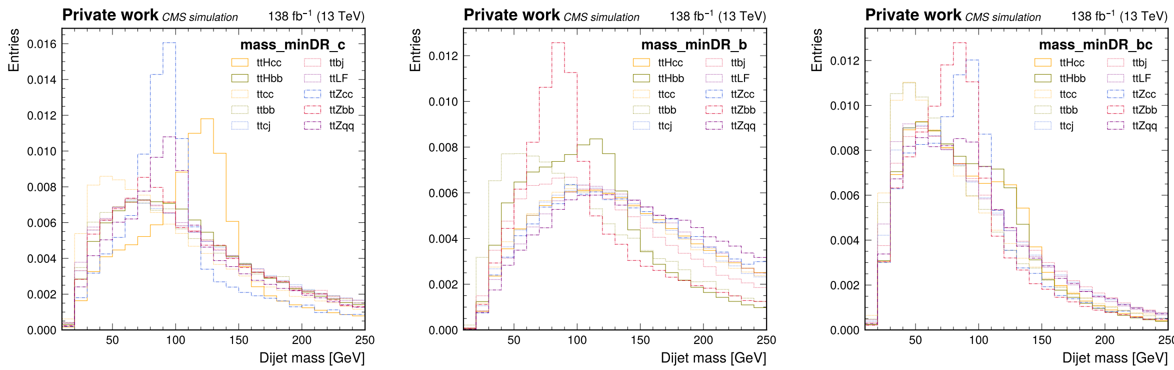

Figure 3: Normalised invariant mass distributions of the dijet system in terms of the mass variables (left) `mass_minDR_c`, (center) `mass_minDR_b`, (right) `mass_minDR_bc`.

To ensure that the classifier becomes mass-agnostic and decorrelated, we treat the processes $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}H}$ as effectively equivalent, as the primary difference between them is their mass. Our objective is for the network to not learn or exploit this mass distinction, but instead respond similarly to both processes. To achieve this, we introduce two new classes: $\mathrm{t\bar{t}X(c\bar{c})}$ and $\mathrm{t\bar{t}X(b\bar{b})}$. They are defined as the sum of the classes $\mathrm{t\bar{t}Z(c\bar{c})}$ with $\mathrm{t\bar{t}H(c\bar{c})}$, and $\mathrm{t\bar{t}Z(b\bar{b})}$ with $\mathrm{t\bar{t}H(b\bar{b})}$, respectively. The classification labels are updated accordingly.

Motivated by the behavior of the $\mathrm{t\bar{t}H(c\bar{c})}$ and $\mathrm{t\bar{t}H(b\bar{b})}$ signals in the `mass_minDR_b` and `mass_minDR_c` distributions, we define two new variables to better distinguish between the two signal sources. The `isBBProcess` flag is defined using a sequence of bitwise `or` operators, covering the following processes $\mathrm{t\bar{t}H(b\bar{b})}$, $\mathrm{t\bar{t}Z(b\bar{b})}$, $\mathrm{t\bar{t}+b\bar{b}}$, $\mathrm{t\bar{t}+b}$, $\mathrm{t\bar{t}+LF}$, and $\mathrm{t\bar{t}Z(q\bar{q})}$. Similarly, the `isCCProcess` flag is defined for c-related processes, covering $\mathrm{t\bar{t}H(c\bar{c})}$, $\mathrm{t\bar{t}Z(c\bar{c})}$, $\mathrm{t\bar{t}+c\bar{c}}$, and $\mathrm{t\bar{t}+c}$. Furthermore, we introduce a new mass variable, `mixedMassVar`. This variable is defined as `mass_minDR_b` for b-related processes and `mass_minDR_c` for c-related processes, thereby combining the two mass distributions into a single variable. This approach ensures that `mixedMassVar` is calibrated across all signal classes, making it a powerful tool for the reweighting procedure.

$$\texttt{mixedMassVar} = \texttt{np.where}(\texttt{isBBProcess}, \texttt{mass\_minDR\_b}, \texttt{mass\_minDR\_c})$$

For the training selection, we require at least three AK4 jets [15, 16], at least one tagged b-jet, and that the total number of tagged c- and b-jets be at least three. Additionally, we impose a selection cut of $30 < m < 250\,\mathrm{GeV}$, where m refers to `mixedMassVar`. This mass variable is incorporated into the reweighting procedure, with events sampled in 10 GeV bin steps from 30 to 250 GeV, ensuring the neural network observes a flat mass distribution. The reweighting classes are the entire classification network, with the exception of keeping $\mathrm{t\bar{t}H(c\bar{c})}$, and $\mathrm{t\bar{t}Z(c\bar{c})}$ separate so that they are reweighted separately to a flat spectrum. Likewise for $\mathrm{t\bar{t}H(b\bar{b})}$, and $\mathrm{t\bar{t}Z(b\bar{b})}$.

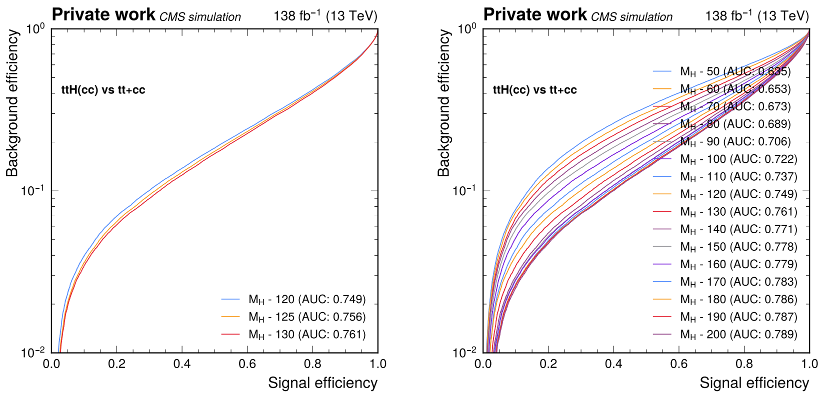

Figure 4: Receiver operating characteristic (ROC) curves for the $c\bar{c}$ discriminant. The performance is evaluated in terms of the efficiency in correctly classifying events containing a pair of c-quarks from the decay of the Higgs boson, as opposed to the efficiency of misidentifying the events as $\mathrm{t\bar{t}+c\bar{c}}$ (left) ROC curves for (blue) 120 GeV, (orange) 130 GeV, and (red) standard model, 125 GeV, $\mathrm{t\bar{t}H}$ signal samples. (Right) ROC curves for all 50 to 200 GeV signal samples together.

Mass variable `mixedGenMassVar`

In subsequent trainings, in addition to the standard model $\mathrm{t\bar{t}H}$ signal samples, we include dileptonic $\mathrm{t\bar{t}H}$ signal samples featuring Higgs boson masses ranging from 50 to 200 GeV. The performance of the variable mass signal samples, evaluated in terms of receiver operating characteristic curves for the $\mathrm{c\bar{c}}$ discriminant, are shown in Figure 4. The standard model sample lies directly in between the 120 GeV and 130 GeV samples. Particularly of note, is the monotonically increasing Area Under the Curve (AUC) score for 50 to 200 GeV signal samples. This trend can be attributed by the fact that a larger Higgs mass corresponds to higher momentum, which makes the signal easier to separate from the backgrounds. The higher momentum events are less likely to resemble background processes, leading to improved discrimination performance as reflected in the increasing AUC.

Due to the heterogeneous nature of the training samples, separate mass definitions were selected for the $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}H}$ samples. To address this, we introduce two new mass variables, `genH_mass` and `mass_add_genjets`, representing the generator-level constructed masses of the Higgs boson and background jets, respectively. Additionally, we define new labels, `sigMass` and `bkgMass`, using a sequence of `or` operators, to differentiate between signal events ($\mathrm{t\bar{t}H(c\bar{c})}$ and $\mathrm{t\bar{t}H(b\bar{b})}$) and background events ($\mathrm{t\bar{t}+b\bar{b}}$, $\mathrm{t\bar{t}+b}$, $\mathrm{t\bar{t}+c\bar{c}}$, $\mathrm{t\bar{t}+c}$, and $\mathrm{t\bar{t}+LF}$).

In the updated setup, we also introduce a new mass variable, `mixedGenMassVar`, defined as `genH_mass` for signal processes and `mass_add_genjets` for background processes. This new variable replaces `mixedMassVar` in the reweighting and selection procedures. We further add the bins at 0, 10, and 20, GeV in the reweighting.

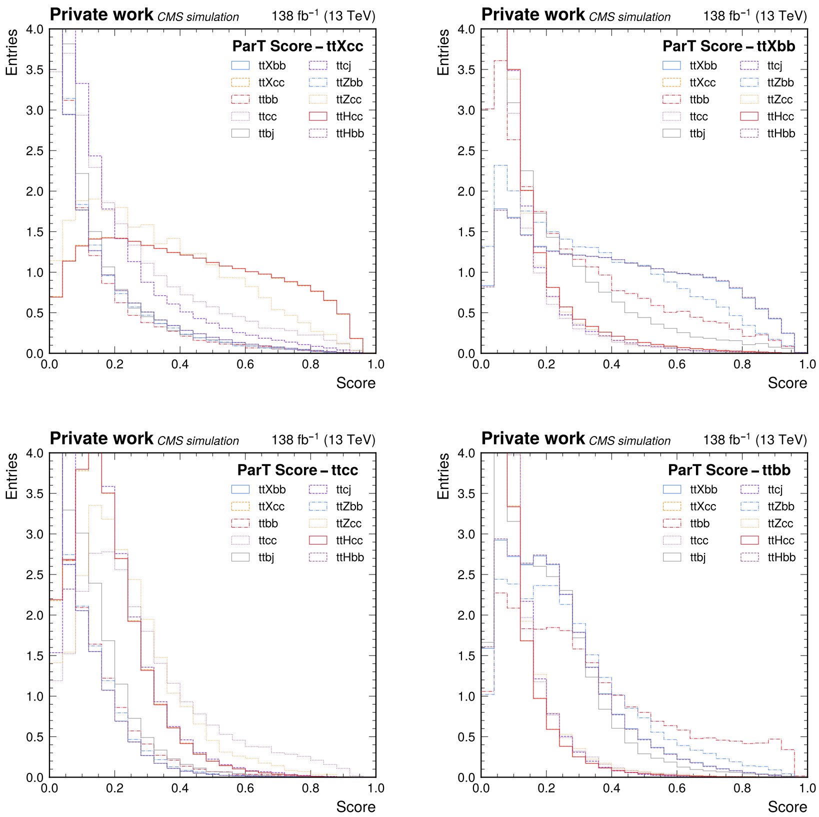

Figure 5: Normalised ParT output score distributions for each physical process x, with a selection based on whether a given label y is valued as `True`. (top) output for signal processes $p(\mathrm{X} \to \mathrm{c}\bar{\mathrm{c}})$ and $p(\mathrm{X} \to \mathrm{b}\bar{\mathrm{b}})$, (bottom) output for selected background $p(\mathrm{t}\bar{\mathrm{t}} + \mathrm{c}\bar{\mathrm{c}})$ and $p(\mathrm{t}\bar{\mathrm{t}} + \mathrm{b}\bar{\mathrm{b}})$.

We exclude the $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}(\mathrm{c}\bar{\mathrm{c}})$, $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}(\mathrm{b}\bar{\mathrm{b}})$, and $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}(\mathrm{q}\bar{\mathrm{q}})$ classes from both the labels and reweighting variables, and remove the $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}(\mathrm{q}\bar{\mathrm{q}})$ sample files. The $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}$ signals, being similar to $\mathrm{t}\bar{\mathrm{t}}\mathrm{H}$ signals with a Higgs mass of 90 GeV, could potentially skew the mass reweighting, thus excluding $\mathrm{t}\bar{\mathrm{t}}\mathrm{Z}$ ensures more consistent mass reweighting.

2.2 Results

The output from ParT, `score_x`, is a probability-like distribution which represents whether a given event is associated with the specified physics process, x. ParT provides two score distributions for the signal: $p(\mathrm{X} \to \mathrm{b}\bar{\mathrm{b}})$ and $p(\mathrm{X} \to \mathrm{c}\bar{\mathrm{c}})$ and additional distributions for the various background multijet. These distributions are summarised in Figure. 5. For each score distribution corresponding to process x, we define the discriminant for another process y as `score_x`, given that the label for y is set to `True`. As anticipated, the plots consistently show that the principal process distribution (`score_x` provided x is `True`) dominates over the rest of the processes, particularly in the high score region.

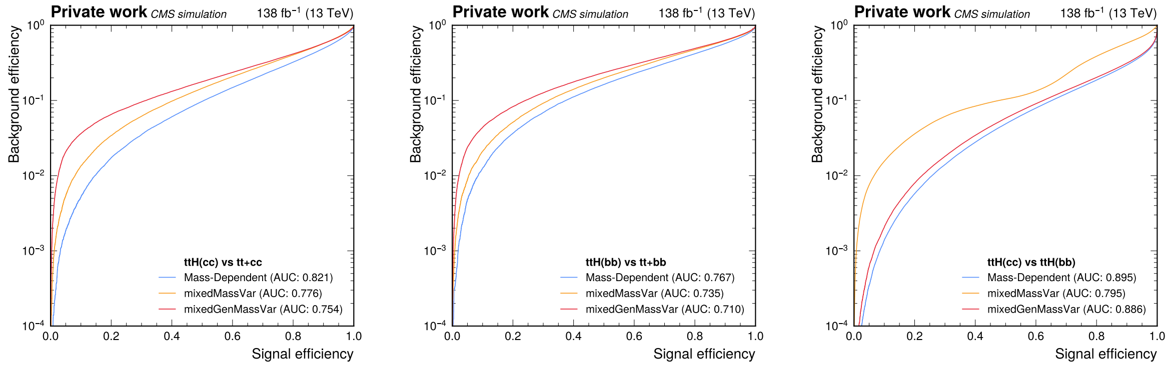

Figure 6: Receiver operating characteristic curves for the ParT discriminant. (left) efficiency in correctly identifying events containing a pair of c-quarks from the decay of the Higgs boson, as opposed to the efficiency of misidentifying the events as $\mathrm{t\bar{t}+c\bar{c}}$ (centre) Likewise for the b-quarks (right) efficiency for correctly identifying events with $\mathrm{c\bar{c}}$ pair from Higgs decay versus the efficiency of mistakenly identifying b-quark pairs instead.

The performance of the model is evaluated in terms of the ROC curves for the ParT discriminant scores. Firstly, we assess the efficiency in correctly identifying events containing a pair of c- or b-quarks from the decay of the Higgs boson, as opposed to the efficiency of misidentifying the events as $\mathrm{t\bar{t}+c\bar{c}}$ and $\mathrm{t\bar{t}+b\bar{b}}$ processes, respectively, as shown in Figure 6. The results are presented for three cases: mass-correlated training (blue), mass-decorrelation through the use of `mixedMassVar` in the reweighting procedure and selection (orange), and likewise using the `mixedGenMassVar` variable (red). As expected, there is some performance loss in the AUC score due to mass-decorrelation of the model. Additionally, when using the `mixedGenMassVar` variable in the reweighting, there is a higher misidentification rate for categorising events with c-quarks from Higgs decay as events from the $\mathrm{t\bar{t}+c\bar{c}}$ process, as well as for b-quarks. This is due to the fact that, `mixedGenMassVar` provides stronger decorrelation compared to `mixedMassVar`, leading to further performance trade-off. Nevertheless, as anticipated, the `mixedGenMassVar` parameter significantly improves the model's efficiency in correctly identifying events with c-quark pairs from Higgs decay compared to b-quark pairs, as illustrated in Figure 6 (right).

The correlation between the `mixedGenMassVar` and ParT classification scores `score_ttXcc` and `score_ttXbb`, are illustrated at Figure 7. Darker regions correspond to a greater density of `mixedGenMassVar` bins across a specific score value. We observe that for $\mathrm{t\bar{t}H(c\bar{c})}$ and $\mathrm{t\bar{t}H(b\bar{b})}$ distributions, there is narrow spread, centered around the Standard Model Higgs mass. Notably, there is a tail around the low score region of the $\mathrm{t\bar{t}H(b\bar{b})}$ distribution, possibly due to misidentification of jets. Some of the jets originating from light-flavor quarks or gluons can be misclassified as

b-jets, leading to a spread in the low-score region. In contrast, c-jets, being easier to separate from b-jets, are more often correctly identified leading to less pronounced tail. Additionally, a peak is evident around the Z mass for both the $t\bar{t}Z(c\bar{c})$ and $t\bar{t}Z(b\bar{b})$ distributions, with the latter exhibiting slightly more spread, again likely due to b-jet misidentification. Lastly, the background distribution appears as a vertical line at low scores, indicating a flat, uniform background as expected.

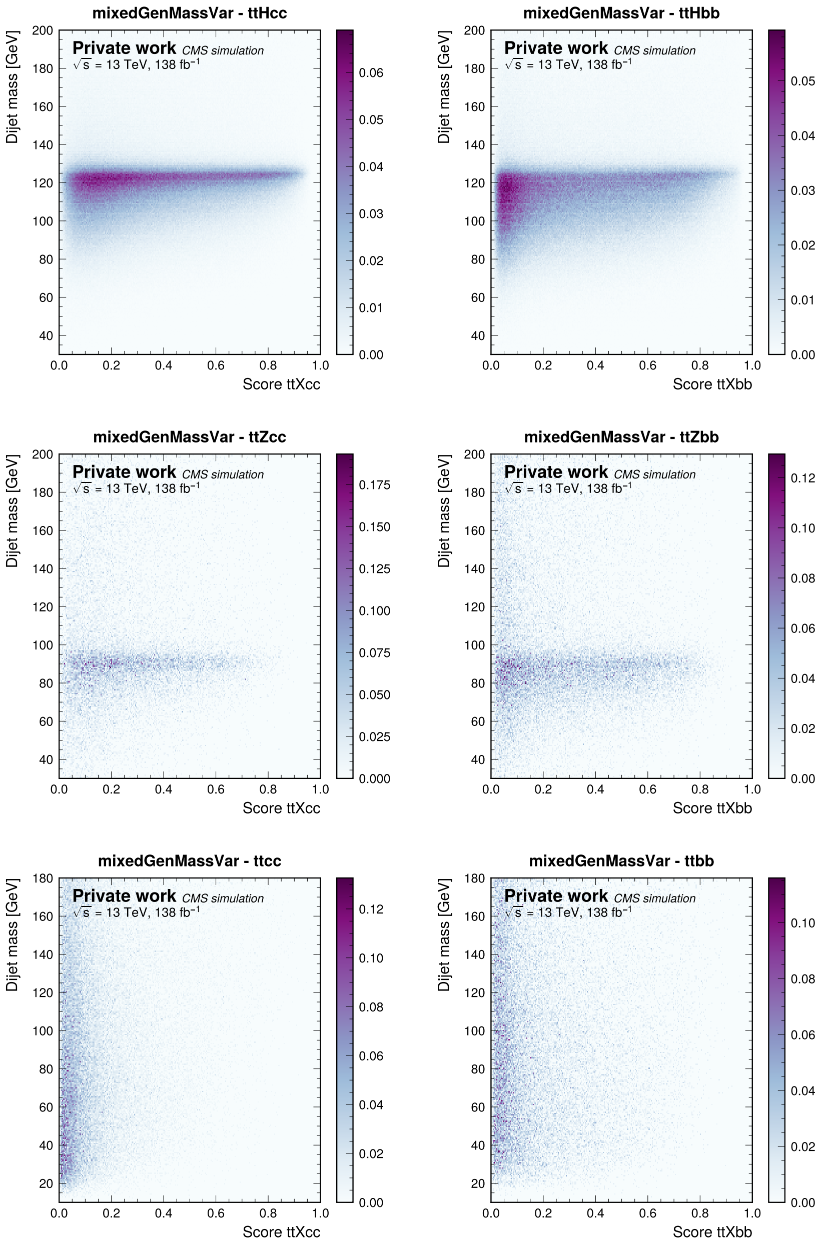

Figure 7: Normalised correlation between the `mixedGenMassVar` mass variable and ParT classification scores `score_ttXcc` (left) and `score_ttXbb` (right)

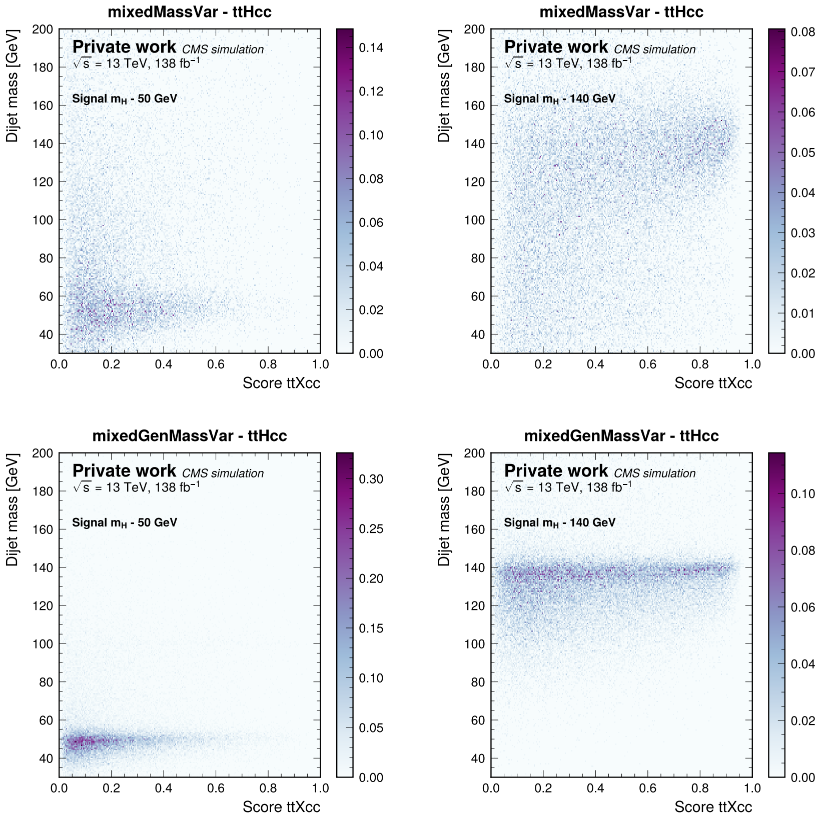

Figure 8: Normalised correlation between the `mixedMassVar` mass variable and ParT classification scores `score_ttXcc` (top) `mixedGenMassVar` (bottom) for 50 (left) and 140 GeV (right) $t\bar{t}H$ signal masses.

The correlation is also shown for both the 50 GeV and 140 GeV $t\bar{t}H$ signal masses for the variables `mixedMassVar` and `mixedGenMassVar` in Figure 8. In both cases, we observe that the 140 GeV signal sample exhibits a more dispersed distribution compared to the 50 GeV sample. This is likely due to the fact that, at higher mass, there is more phase space available for the decay products, resulting in a wider range of possible kinematic configurations and, consequently, lower resolution at the reconstructed dijet mass. Furthermore, we note that `mixedMassVar` shows a significantly broader spread at higher masses compared to `mixedGenMassVar`. This behavior is attributed to how `mixedMassVar` is defined, which relies on the minimum separation in ΔR between the *b*- or *c*-jets. For Higgs decays, ΔR approximately scales as $2m_H/p_T$, meaning that as the Higgs mass increases, the jets become less collimated. As a result, the $\texttt{min}\,\Delta R$ selection is less likely to correctly identify the two Higgs jets at higher masses.

The normalised mass distributions of the dijet system as a function of `mixedMassVar` (top) and `mixedGenMassVar` (bottom), with progressively tighter cuts on the discriminating score, are presented in Figure 9. Solid lines represent the signal, while dashed lines correspond to the $t\bar{t}$ + jets background (specifically $t\bar{t} + c\bar{c}$ or

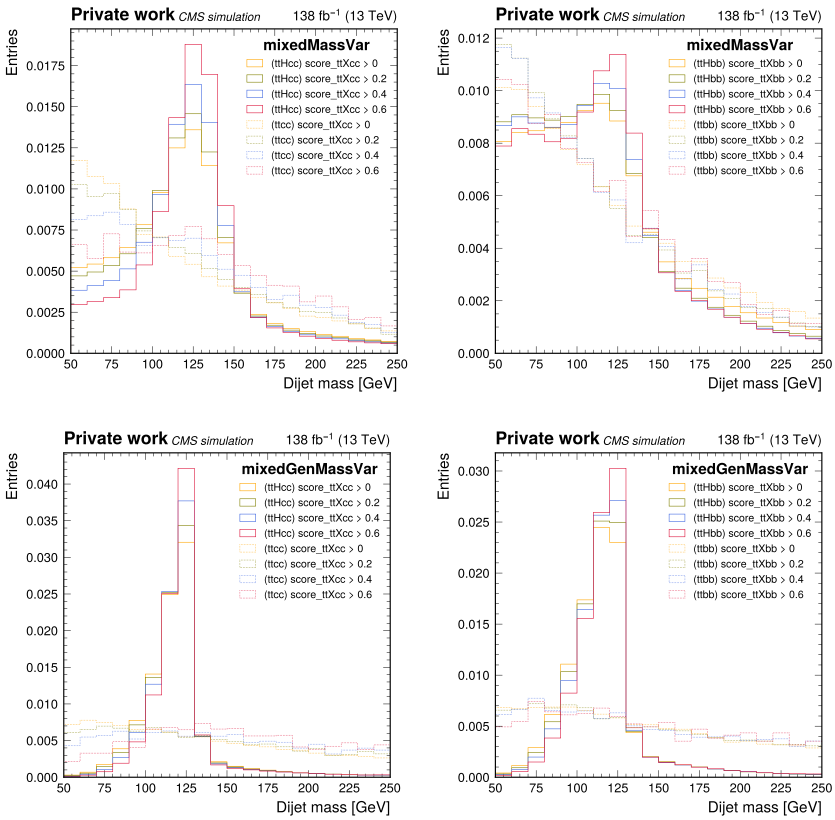

Figure 9: Normalised mass distribution of the dijet system, as a function of `mixedMassVar` (top) or `mixedGenMassVar` (bottom), with an increasing cut on the discriminating score. (left) reconstructed mass of a pair of c-quarks, (right) pair of b-quarks. Shown for the dileptonic channel. All plots correspond to using `mixedGenMassVar` in the reweighting.

$\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$). The left plots show distributions for a pair of b-quarks, while the right plots show those for a pair of c-quarks. All plots correspond to the dileptonic samples. As the selection increases, we observe that the signal distribution increases for both mass variables, while the background does not vary greatly and most significantly remains smooth. This demonstrates that the model has been effective in mass decorrelation. Notably, `mixedGenMassVar` shows better mass decorrelation performance compared to `mixedMassVar`, which is expected given that it is defined using generator-level constructed invariant mass, unlike `mixedMassVar` which is defined using reconstruction-level variables. For both mass variables, the $\mathrm{t\bar{t}H(c\bar{c})}$ signal has greater maximum bin entry density in contrast to $\mathrm{t\bar{t}H(b\bar{b})}$ distribution, possibly due to a greater source of combinatorial background in the b-quark channel. However, the background for c-related processes varies more strongly on higher selections than that of b-related processes.

Figures 10 and 11 display the mass distributions of all physical processes stacked together, with increasingly tighter cuts on the ParT discriminating score from

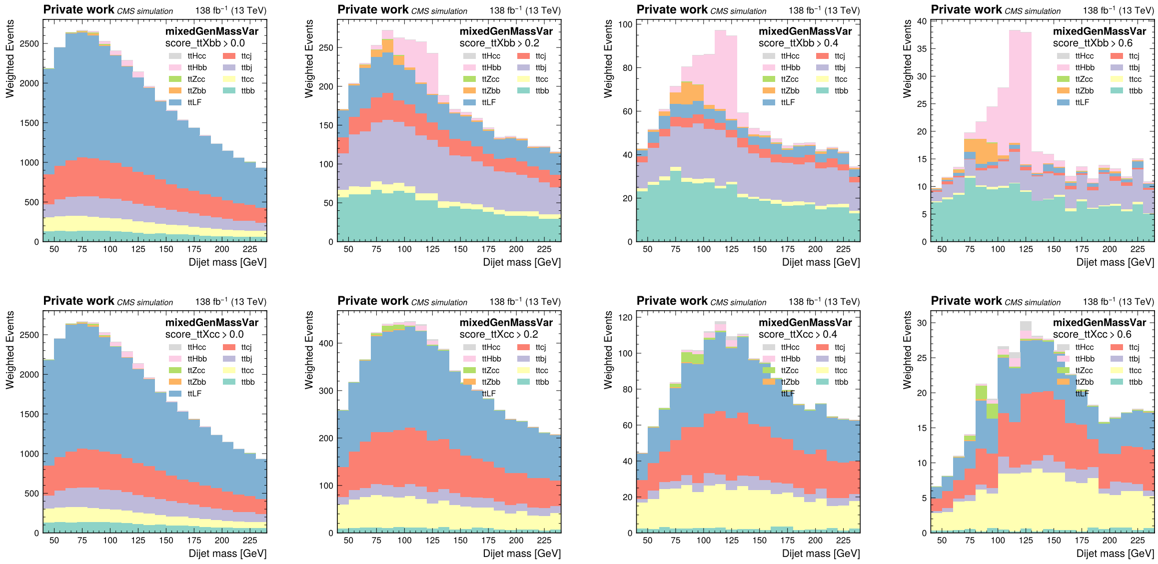

Figure 10: `mixedGenMassVar` weighted distribution for all of the physical processes stacked on top of each other, with an increasing cut on the ParT discriminating scores, `score_ttXbb` (top), and `score_ttXcc` (bottom)

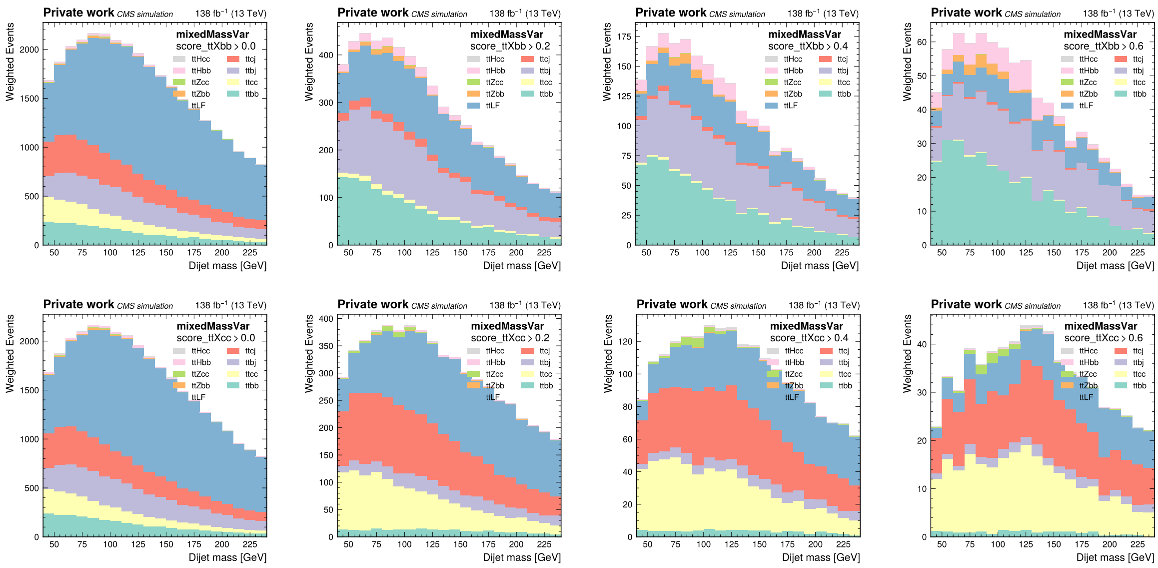

Figure 11: `mixedMassVar` weighted distribution for all of the physical processes stacked on top of each other, with an increasing cut on the ParT discriminating scores, `score_ttXbb` (top), and `score_ttXcc` (bottom)

left to right for `mixedGenMassVar` and `mixedMassVar`, respectively. In addition, a selection criterion requiring at least four AK4 jets has been applied on top. The events are weighted by the product of the `generator weight`, SM `cross-section weight`, and full Run 2 `Integrated luminosity` of 138 fb^{-1}. Notably, across all the plots, the signal for the $\mathrm{t\bar{t}H}$ processes is most pronounced around the Higgs boson mass.

We calculated the statistical significance of the final ParT event classifier for signal processes in the presence of the remaining background across a mass window of $100 < m < 150\,\mathrm{GeV}$ for $\mathrm{t\bar{t}H}$ and $65 < m < 115\,\mathrm{GeV}$ for $\mathrm{t\bar{t}Z}$, where m corresponds to `mixedMassVar` or `mixedGenMassVar`. The results are summarised in Table 2.

To achieve this, we divided the mass range into equally spaced bins. To ensure that the results for c- and b-like events were orthogonal, we imposed the condition that only bins where the ParT output `score_ttXcc` was greater than `score_ttXbb` were selected, and similarly for the b-related processes, thus eliminating any double-counting. For each mass bin, we computed the sum of event weights for background processes, denoted as B, and for the signal process, denoted as S. We then applied the Asimov approximation to determine the significance in each bin,

$$Z_{\mathrm{bin}}(S,B) = \sqrt{2\left((S+B)\log\left(1+\frac{S}{B}\right)-S\right)}$$

In cases where no background events were present ($B = 0$), we assigned a significance of zero to avoid division by zero. To obtain the total significance, we compute the square root sum of the significance value across all mass bins, $Z_{\mathrm{total}} = \sqrt{\sum_{\mathrm{bins}} Z^2_{\mathrm{bin}}}$.

Mass Variable		**Score Cut** (ttXcc/ttXbb)			
		0	0.2	0.4	0.6
`mixedGenMassVar`	ttHcc	0.119	0.190	0.263	**0.308**
	ttZcc	0.495	0.737	0.860	0.724
	ttHbb	3.123	5.598	6.964	**7.458**
	ttZbb	0.839	1.458	1.658	1.440
`mixedMassVar`	ttHcc	0.066	0.098	0.122	**0.133**
	ttZcc	0.565	0.765	0.774	0.569
	ttHbb	1.180	2.025	2.306	**2.261**
	ttZbb	0.744	1.118	1.109	0.861

Table 2: Significance for $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}Z}$ signals of the final ParT event classifier with a cut on the discriminating scores.

For both `mixedGenMassVar` and `mixedMassVar`, increasing the ParT discriminating score generally leads to higher significance for the $\mathrm{t\bar{t}H}$ signals, with the `mixedGenMassVar` achieving the highest significance of 7.458 and 0.308 for $\mathrm{t\bar{t}H(b\bar{b})}$ and $\mathrm{t\bar{t}H(c\bar{c})}$ events at a score cut of 0.6, respectively. This trend underscores the classifier's ability to effectively distinguish signal from background when stricter selection criteria are applied. Conversely, for the $\mathrm{t\bar{t}Z}$ processes, the significance peaks at intermediate score cuts (0.4) before declining, suggesting that further improvement is required to refine the $\mathrm{t\bar{t}Z}$ processes. Additionally, the `mixedGenMassVar` consistently outperforms the `mixedMassVar` in terms of significance for both signal types, indicating its superior discriminative power, as expected since it is based on generator level variables.

3 Regression

In contrast to classification, where the objective is to learn a function $f\colon \mathbb{R}^n \to \{1, 2, \ldots, K\}$, where the input space is mapped to a finite set of discrete labels. Regression, on the other hand, seeks to predict a continuous, real-valued, function $f\colon \mathbb{R}^n \to \mathbb{R}$ that captures the relationship between the input features and the output. The performance of a classification model is assessed using metrics such as accuracy and binary cross-entropy loss, while regression model minimises a loss function - Log-Cosh Loss in our application,

$$L(\hat{y}, y) = \sum_{i=1}^{n} \log\big(\cosh(y_i - \hat{y}_i)\big)$$

to quantify the difference between predicted values $\hat{y}_i$ and the true values y_i.

3.1 Model Setup

We adopt a methodology similar to that of the ParT event classifier model. We use `mixedMassVar` in the reweighting process, sampled from 30 to 200 GeV to ensure the network sees a flat mass spectrum. A mass cut of $30 < m < 200\,\mathrm{GeV}$ is applied as part of the selection criteria. For consistent reweighting, only $\mathrm{t\bar{t}H}$ signal and $\mathrm{t\bar{t}}+\mathrm{jets}$ classes are considered, and all corresponding SM Monte Carlo samples are included. In this approach, the target mass is defined as the ratio of `mixedMassVar` to a newly introduced variable, `maxMassVar` - determined as the maximum value between `mass_minDR_b` and `mass_minDR_c`, thereby capturing the largest mass, and delivering improved representation for c- and b-related processes combined. Figure 12 shows the normalised distribution of `maxMassVar` (left), and ratio `mixedMassVar` / `maxMassVar` (right). Following this, the true mass distribution is then expressed as the output of the regression model multiplied by `maxMassVar`.

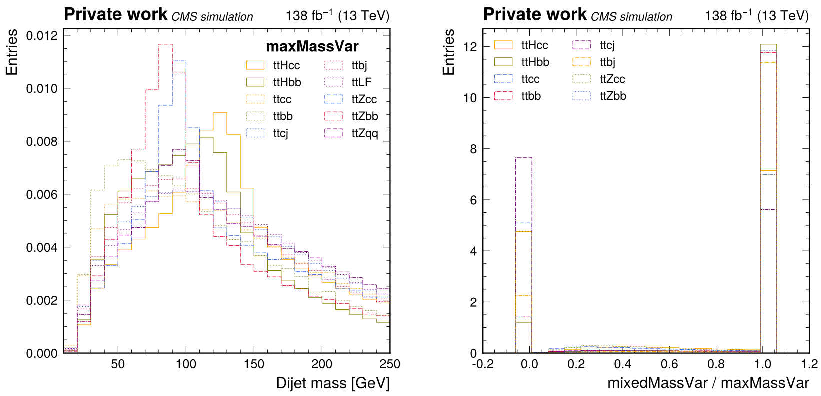

Figure 12: Normalised invariant mass distributions of the dijet system in terms of the mass variable `maxMassVar` (left). The ratio `mixedMassVar` / `maxMassVar` (right)

As seen in Fig 12, for the majority of events, `mixedMassVar` and `maxMassVar` are approximately equal. However, significant number of events also lie around 0. Applying the ratio as the target helps account for discrepancies, thereby improving over using only `mixedMassVar` as the target. Furthermore, by setting the regression target as the ratio between the two mass variables, we achieve a significant reduction in loss, which improves the overall model performance. The model was found to give best performance with 40 epochs, using 5-fold cross-validation and an initial learning rate of 10^{-4}. Moreover, loading the semi-leptonic final classifier model weights during training led to a more consistent and stable performance throughout.

3.2 Results

The performance of the regression model is evaluated in terms of the ratio of the regression output to the target mass, `mixedMassVar` / `maxMassVar`, illustrated in Figure 13 (right). The raw output of the regression is also presented on the left side of the figure. Notably, the large spike centered around 0 observed in the target variable, is no longer present in the regression output.

The normalised mass spectrum of the dijet system, obtained using the regression model with progressively tighter selections on the discriminating score, is presented in Figure 14. The discriminating score is derived from the output of the ParT mass-decorrelated training, with `mixedGenMassVar` used for the reweighting. As the selection tightens, we observe minimal mass sculpting in the background mass distribution. Additionally, Figure 15 displays the mass spectrums of all physical processes stacked together, with increasingly tighter cuts on the classification score. The events were selected to have at least four AK4 jets and weighted by the product of `generator weight`, SM `cross-section weight` and full Run 2 luminosity. The significance is summarised at Table 3.

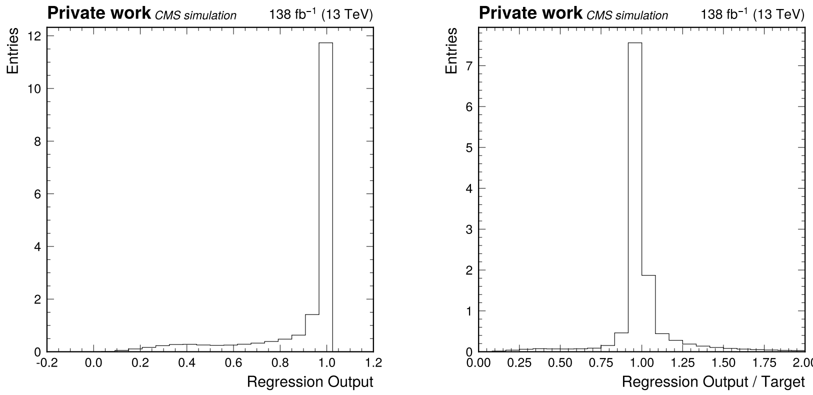

Figure 13: Normalised distribution of the regression output (left), and the ratio of regression output to the target variable (right) for the $t\bar{t}H(c\bar{c})$ dileptonic sample.

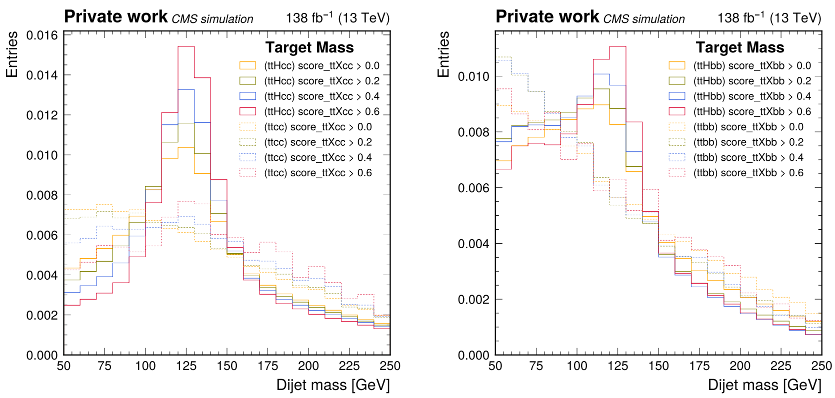

Figure 14: Normalised invariant mass distributions of the dijet system in terms of the regression output, with an increasing cut on the discriminating score from the ParT mass-decorrelated training. (left) reconstructed mass of a pair of c-quarks in the dileptonic channel, (right) pair of b-quarks.

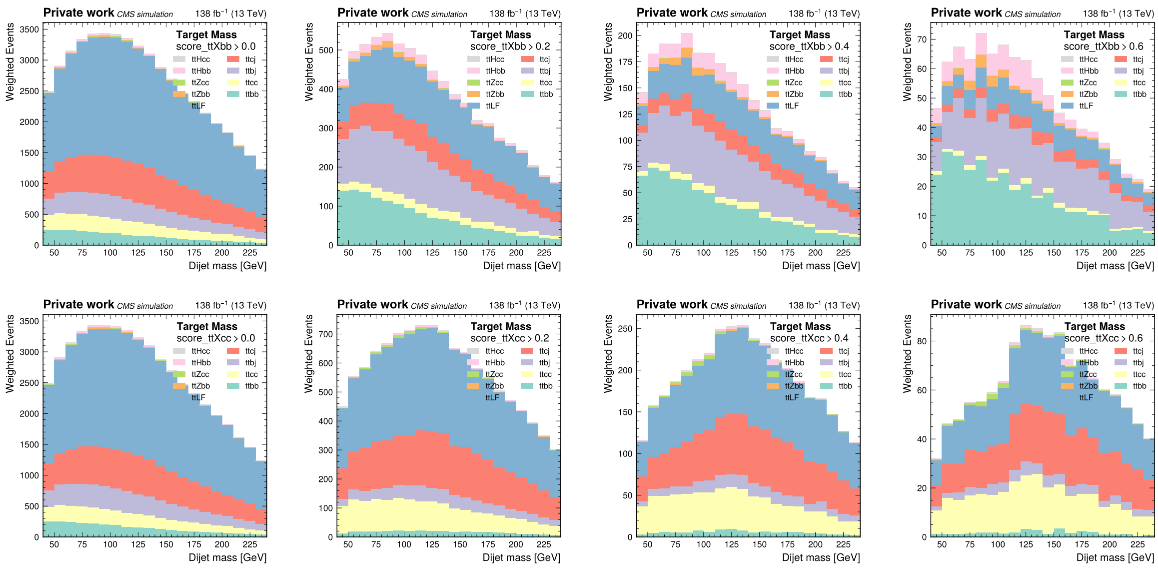

Figure 15: Weighted mass distribution obtained by the regression model for all of the physical processes stacked on top of each other, with an increasing cut on the ParT discriminating scores, `score_ttXbb` (top), and `score_ttXcc` (bottom). Further selection requiring at least four AK4 jets has been applied.

4 Segmentation

In segmentation, the task is to label each item within an event based on their origin or relevance to a particular process of interest. This is distinct from classification, where a single label is assigned to the event as a whole. Specifically, in the context of our studies, segmentation involves identifying jets produced by the decay of the Higgs boson and differentiating them from background jets. Formally, for an event with N jets, each with C features, the task can be described as learning a mapping,

$$f\colon \mathbf{X} \in \mathbb{R}^{N\times C} \to \mathbf{Y} \in \mathbb{Z}^{N}$$

where $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ represents the features of every jet (such as transverse momentum p_T, energy, and angular separation), and $\mathbf{Y} = \{y_1, y_2, \ldots, y_N\}$ represents the corresponding labels. Here, $y_i = 1$ indicates that jet $i \in [1, N]$ originated from the Higgs boson decay, and $y_i = 0$ means it came from a background process.

While regression predicts continuous values that may be useful for parameter estimation, it may not effectively distinguish between different jet origins. Hence, by focusing on the structural patterns within events through segmentation, we enhance our ability to discriminate between signal and background processes. This detailed labeling facilitates more precise event reconstruction and better background rejection, ultimately improving the sensitivity of the analysis by isolating the Higgs boson signal more effectively than in regression.

4.1 Model Setup

In the segmentation model, jets are indexed to assign a unique identifier to each jet within the event. This is done using the following code,

```
jet_idx = np.tile(np.arange(9), len(ak4_pt)).reshape(-1, 9)
```

This expression creates an array of indices from 0 to 8 for each event. The `np.tile` function replicates this sequence across the number of events, and `reshape(-1, 9)` ensures that each event has a fixed number of jets (9 in this case). This indexing allows the model to uniquely identify and keep track of the jets. Once indexed, the jets are labeled based on whether they originated from the Higgs boson decay. This is achieved by comparing each jet index to the indices of the Higgs daughter jets. If the current jet index matches either `hdau_jetidx1` or `hdau_jetidx2`, the jet is labeled as originating from the Higgs boson; otherwise, it is labeled as background. This labeling provides the ground truth against which the model can compare its predictions.

Events often contain a variable number of jets, which presents a challenge for neural the network that requires fixed input sizes. To address this, the segmentation model employs masking and padding. The `ak4_mask` variable, defined by `ak.ones_like(ak4_pt)`, is used to mask out irrelevant jets — that is, only jets that match the structure of `ak4_pt` are included in the segmentation learning. The jets are then padded to ensure that each event has a uniform number of jets for processing by the model. Padding is achieved using the function,

```
_pad(jet_label, 9)
```

Ensuring that all events are padded to have exactly 9 jets, even if the original event contains fewer, henceforth allowing the model to handle events of varying sizes while maintaining a consistent input shape. The loss function used for segmentation is a per-jet binary cross-entropy, since the task is to classify each jet as either originating from the Higgs boson decay or from a background process. The loss function is defined as,

$$H(\hat{y}, y) = -\sum_{i=1}^{N} \big(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\big)$$

where y_i is the true label of the i-th jet (1 if it comes from the Higgs, 0 otherwise), and $\hat{y}_i$ is the predicted probability that the jet originated from the Higgs boson. The labels in this setup are defined as either `other` — jets that do not originate from the Higgs boson — or `fromH` for jets that the model assigned as originating from the Higgs decay.

Mass Variables `mH`

Since the definition and resolution of the ΔR-based observable are not optimal, and we aim to achieve a sharper peak at m_H, we introduce a mass variable. `mH` is defined

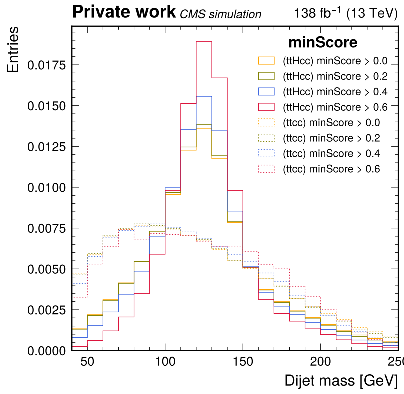

Figure 16: Normalised invariant mass distribution of the dijet system with a selection on the `minScore` discriminant, shown for the $t\bar{t}H(c\bar{c})$ signal and $t\bar{t}+c\bar{c}$ background.

as an integer corresponding to the specific mass used in our variable-mass $t\bar{t}H$ signal samples. We incorporate `mH` into the reweighting procedure, sampling events in 10 GeV increments from 50 to 210 GeV to ensure the model observes a flat mass spectrum. Since we are interested in classifying whether a jet originated from a Higgs boson or not, we limit to only $t\bar{t}H(c\bar{c})$ and $t\bar{t}H(b\bar{b})$ categories for reweighting. Consequently, we restrict the training inputs to variable-mass dileptonic $t\bar{t}H$ signal samples, in accordance with the mass variable used in reweighting.

4.2 Results

The output from ParT segmentation training, `score_fromH`, is a probability-like distribution that indicates whether a given jet originated from a Higgs boson decay or not. We use this score to identify and select the most probable jets associated with the Higgs boson and then calculate their four-momenta to compute the invariant mass of the system. To achieve this, we sort the jets within each event in descending order based on their `score_fromH` values and select the top two jets per event as most likely Higgs boson candidates. Additionally, we extract the second-highest score per event, denoted as `minScore`, to measure the confidence level of the less probable jet among the top two candidates. Next, we extract the momentum components p_x, p_y, p_z, and the energy E of the selected two jets. Using these components, we compute the invariant mass M of the dijet system for each event, yielding a mass distribution.

A plot of the normalised invariant mass distribution of the dijet system with a selection on the `minScore` discriminant is shown in Figure 16. By comparing the distributions, a choice has been made to require a cut on the segmentation `minScore` score such that `minScore` > 0.4. This constraint will enhance the purity of the selected proposed Higgs boson candidates in the invariant mass

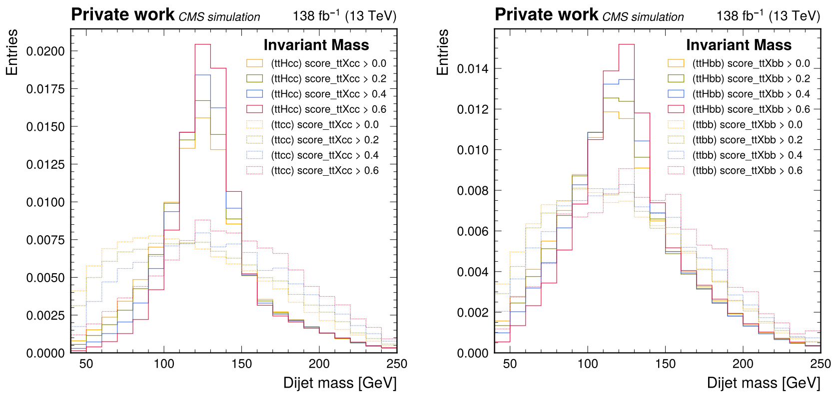

Figure 17: Normalised invariant mass distributions of the dijet system of the two highest `fromH` score jets within an event satisfying `minScore` > 0.4, with an increasing cut on ParT mass-decorrelated discriminating score. (left) reconstructed mass of a pair of c-quarks in the dileptonic channel, (right) pair of b-quarks

calculation by requiring both jets to have a high probability of originating from Higgs decay. Following this we compute the invariant mass, with tighter selections on the ParT mass-decorrelated event classifier discriminating score, `score_ttXcc` or `score_ttXbb`, with `mixedGenMassVar` used for the reweighting - illustrated in Figure 17. As the selection increases, we observe minimal mass sculpting in the background mass distribution. Additionally, Figure 18 displays the mass spectrums of all physical processes stacked together, with increasingly tighter cuts on the classification score. On top of the `minScore` constraint, the events were selected to have at least four AK4 jets and weighted by the product of `generator weight` and `cross-section weight`.

Table 3 summarises the significances obtained from both the regression output, scaled by `maxMassVar`, and the invariant mass derived from the segmentation model, across increasing cuts of the mass-decorrealted ParT discriminating score.

Notably, the segmentation approach consistently yields higher significance values than the regression method across all score cuts. Specifically, for $t\bar{t}H(b\bar{b})$, the significance peaks at 3.469 with a score cut of 0.4, while regression yields a slightly lower peak at 2.539. Similarly, $t\bar{t}Z(b\bar{b})$ follows a comparable trend, though the significance drops greatly on the last cut, suggesting that excessive background rejection also leads to signal loss. For $t\bar{t}H(c\bar{c})$ and $t\bar{t}Z(b\bar{b})$, the observed significances are lower, with segmentation still showing a slight advantage. In the $t\bar{t}H(c\bar{c})$ process, the segmentation approach peaks at 0.140 for the 0.6 cut, while regression reaches 0.114. In $t\bar{t}Z(c\bar{c})$, the trend is similar, with a peak significance of 0.771 for segmentation, although similarly to the b-corresponding process, the value drops significantly at the highest score cut. Regression performs similarly but peaks at a slightly lower value of 0.610. These observations align with the understanding that regression, though useful for parameter estimation, lacks the structural pattern recognition employed in segmentation, which enhances event classification, providing a better signal-to-background discrimination.

A comparison with the previous CMS $t\bar{t}H(b\bar{b})$ analysis [17] shows that the observed significance across all channels was around 1.3 standard deviations (SD) with an expected 4.1 SD. In contrast, the significance for $t\bar{t}H(b\bar{b})$ process in our analysis reaches up to 3.469 in the dileptonic channel, highlighting a notable improvement.

	Mass Variable	**Score Cut** (ttXcc/ttXbb)			
		0	0.2	0.4	0.6
`Segmentation-based`	ttHcc	0.087	0.115	0.137	**0.140**
	ttZcc	0.575	0.736	0.771	0.598
	ttHbb	2.314	3.035	3.469	**3.362**
	ttZbb	0.976	1.220	1.229	0.929
`Regression-based`	ttHcc	0.063	0.090	0.109	**0.114**
	ttZcc	0.443	0.606	0.610	0.476
	ttHbb	1.307	2.235	2.539	**2.475**
	ttZbb	0.754	1.147	1.151	0.906

Table 3: Significance for $t\bar{t}H$ and $t\bar{t}Z$ signals of the invariant mass of the two highest `fromH` score jets (top), regression output times `maxMassVar` (bottom).

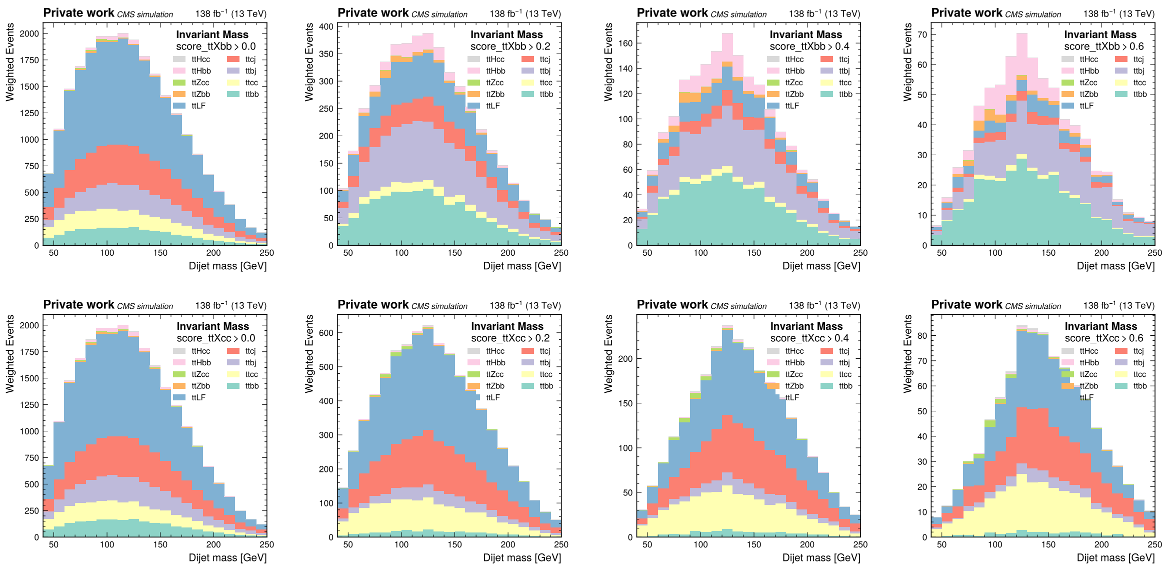

Figure 18: Weighted invariant mass distribution of the two highest `fromH` score jets for all of the physical processes stacked on top of each other, with an increasing cut on the ParT discriminating scores, `score_ttXbb` (top), and `score_ttXcc` (bottom).

5 Conclusion

The goal of this project was to develop a mass-decorrelated event classifier using the Particle Transformer architecture. To attain this, we produced $\mathrm{t\bar{t}H}$ signal samples with masses varying from 50 to 200 GeV, then in the training we sampled the events so that the neural network saw a flat distribution in mass, employing the Higgs candidate mass in the reweighting. Altogether, we have been able to achieve promising mass decorrelation results, with minimal mass sculpting, and a significance for the generator mass variable of up to 0.308 and 7.458 for the standard model dileptonic $\mathrm{t\bar{t}H(c\bar{c})}$ and $\mathrm{t\bar{t}H(b\bar{b})}$ samples, respectively.

Furthermore, we have investigated the implementation of mass-decorrelation techniques in regression and segmentation - where the classification is done on a per-jet basis. Going forwards, the next target would be to investigate across the semileptonic and fully-hadronic channels, thereby covering the whole event topology of the associated Higgs boson production with a top quark-antiquark pair to effectively capture the Higgs to charm quark-antiquark pair decay.

References

[1] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716: 1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.

[2] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.

[3] Morad Aaboud et al. Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 99:072001, 2019. doi: 10.1103/PhysRevD.99.072001.

[4] Morad Aaboud et al. Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett. B*, 786:59–86, 2018. doi: 10.1016/j.physletb.2018.09.013.

[5] M. Aaboud et al. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett. B*, 784: 173–191, 2018. doi: 10.1016/j.physletb.2018.07.035.

[6] Albert M Sirunyan et al. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Phys. Lett. B*, 779:283–316, 2018. doi: 10.1016/j.physletb.2018.02.004.

[7] Albert M Sirunyan et al. Observation of $t\bar{t}H$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018. doi: 10.1103/PhysRevLett.120.231801.

[8] A. M. Sirunyan et al. Observation of Higgs boson decay to bottom quarks. *Phys. Rev. Lett.*, 121(12):121801, 2018. doi: 10.1103/PhysRevLett.121.121801.

[9] Albert M. Sirunyan et al. Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(2):021801, 2019. doi: 10.1103/PhysRevLett.122.021801.

[10] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998. doi: 10.1016/S0010-4655(97)00123-9.

[11] Armen Tumasyan et al. Search for Higgs Boson Decay to a Charm Quark-Antiquark Pair in Proton-Proton Collisions at s=13 TeV. *Phys. Rev. Lett.*, 131(6):061801, 2023. doi: 10.1103/PhysRevLett.131.061801.

[12] Armen Tumasyan et al. Search for Higgs Boson and Observation of Z Boson through their Decay into a Charm Quark-Antiquark Pair in Boosted Topologies in Proton-Proton Collisions at s=13 TeV. *Phys. Rev. Lett.*, 131(4):041801, 2023. doi: 10.1103/PhysRevLett.131.041801.

[13] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016. doi: 10.23731/CYRM-2017-002.

[14] Huilin Qu, Congqiao Li, and Sitian Qian. Particle Transformer for Jet Tagging. 2 2022.

[15] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[16] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.

[17] Aram Hayrapetyan et al. Measurement of the $t\bar{t}H$ and tH production rates in the H $\rightarrow b\bar{b}$ decay channel using proton-proton collision data at $\sqrt{s} = 13$ TeV. 7 2024.

Appendix

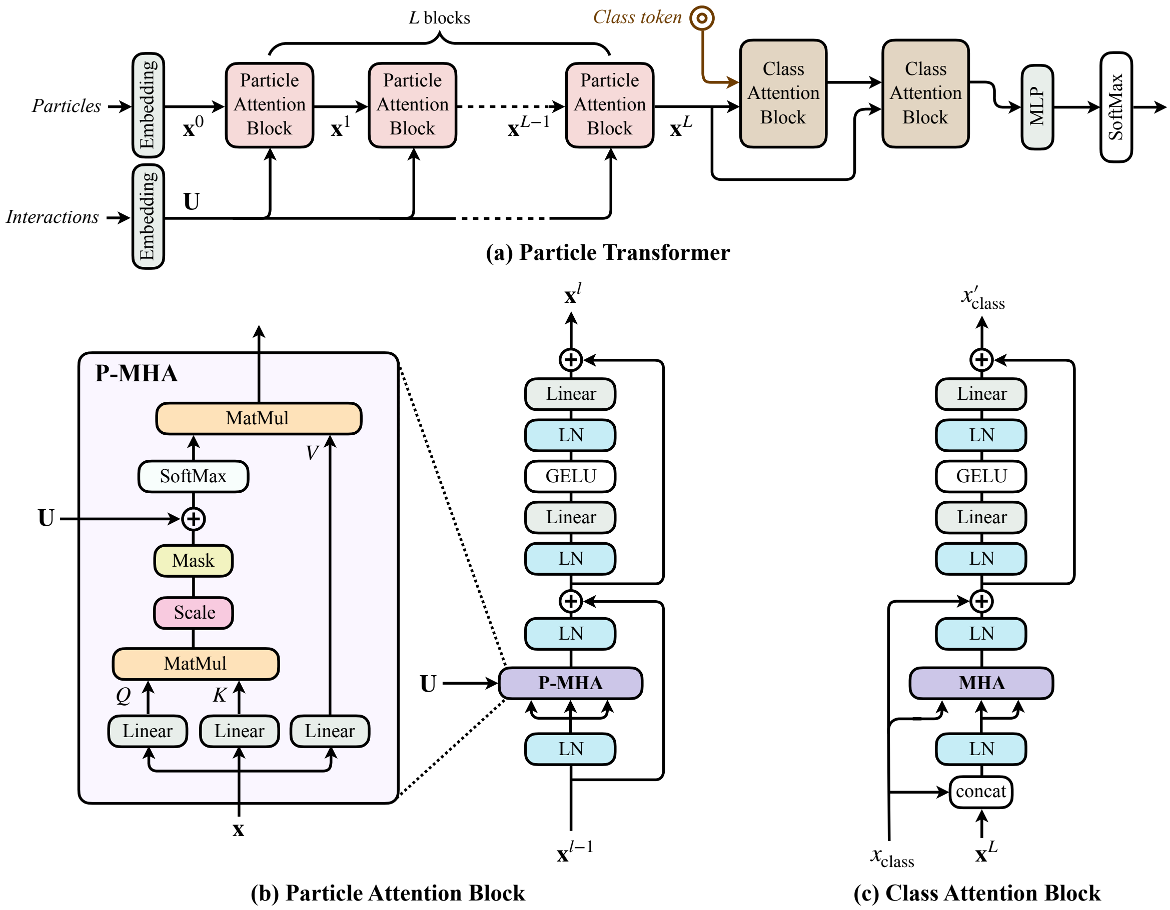

Figure 19: Structure of the Particle Transformer Network [14]