



Constraining the Charm-Yukawa coupling at the Large Hadron Collider

Joseph Walker^{a,b,*}, Frank Krauss^{a,b}

^a Institute for Particle Physics Phenomenology, Durham University, United Kingdom

^b Institute for Data Science, Durham University, United Kingdom

ARTICLE INFO

Article history:

Received 18 March 2022

Received in revised form 8 June 2022

Accepted 15 June 2022

Available online 17 June 2022

Editor: G.F. Giudice

ABSTRACT

We present theoretical results for the sensitivity of charm Yukawa coupling measurements in future high-luminosity LHC runs in three channels: Vector Boson Fusion (VBF), W Higgs-strahlung and Z Higgs-strahlung production of a Higgs boson and its subsequent decay into charm quarks. To reduce the overwhelmingly large backgrounds and to reduce false positives, we apply a set of simple kinematic and jet feature cuts and feed neural network data structures of three types; jet features, jet images and particle level features. To facilitate straightforward comparison with experimental studies [1,2], we express our results in terms of signal strengths [3].

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP³.

1. Introduction

The 2012 discovery of the Higgs boson H with a mass of 125 GeV by the ATLAS and CMS collaborations at the LHC [4–6] completed the Standard Model (SM) of particle physics, and initiated a step-change in our quest to understand nature at its fundamental scale. Since this discovery the focus has shifted to measuring the new boson's properties and interactions, and to constrain effects of possible new physics manifesting itself through subtle deviations from SM predictions. A central part of these efforts has been the determination of the Higgs boson couplings to the other SM particles. By the end of Run 2 couplings to the SM gauge bosons [7–9] and to the massive third-generation fermions [7,10–15] have been measured, and the first coupling to the second-generation fermions, i.e. to the muon, has been constrained [7,16,17].

Taking advantage of the significantly higher luminosity of the forthcoming phases of LHC data taking, the precision of these measurements will be further improved and couplings that have hitherto been inaccessible will become subject to scrutiny. This includes, in particular, a determination of the Higgs Yukawa coupling to charm quarks as a further, stringent test of the universal role of the Higgs boson in the generation of fundamental masses. Various channels have been suggested, some of them using the larger electromagnetic charge of the charm quark with respect to the bottom quark [18], or the decay into $c\bar{c}$ quarkonia, for example in decays such as $H \rightarrow J/\psi + \gamma$ [19–21], or by recasting measurement of the

$H \rightarrow b\bar{b}$ branching ratio, for example in [22–24]. We will analyse the prospects for such a measurement in three channels, namely the production of Higgs bosons and their subsequent decay into charm quarks in association with Z or W bosons and in vector boson fusion,

$$pp \rightarrow H_{\rightarrow c\bar{c}} jj + X \quad (0\text{-lepton channel, } 0L)$$

$$pp \rightarrow W_{\rightarrow \ell\nu_\ell} H_{\rightarrow c\bar{c}} + X \quad (1\text{-lepton channel, } 1L)$$

$$pp \rightarrow Z_{\rightarrow \ell\bar{\ell}} H_{\rightarrow c\bar{c}} + X \quad (2\text{-lepton channel, } 2L)$$

In all processes the relevant quantity is the branching ratio of the Higgs boson to the charm quark, in the SM given by $Br_{H \rightarrow c\bar{c}} = 2.89\%$ [7] and significantly smaller than its counterpart, the branching ratio to the b quarks, around 58.2% [7]. In previous studies [1,2], therefore, the production of b quarks has been identified as a significant non-trivial background, along with signatures from QCD, vector boson and $t\bar{t}$ production.

To provide a simple interpretation of our results we will express the signal strength μ in the κ -framework [3] and use the CL_s method [25] to estimate the upper bound for κ_c at the 95% level. Modifying our observed signal counts, s_x by the signal strength μ_i in each bin, x of a given distribution,

$$N_x(\mu) = b_x + \mu s_x, \quad (1)$$

and assuming an unmodified background, $\mu = 1$ defines the SM value and varying is equivalent to varying κ_c in some parametric way derived below. For different signal processes i , i.e. for the 0-lepton, 1-lepton, and 2-lepton signatures, the μ_i represent ratios of cross sections σ_i times branching ratios in the narrow-width approximation:

* Corresponding author.

E-mail address: j.j.walker@durham.ac.uk (J. Walker).

$$\mu_i = \frac{\sigma_i \cdot Br_{H \rightarrow c\bar{c}}}{\sigma_i^{SM} \cdot Br_{H \rightarrow c\bar{c}}^{SM}}, \quad (2)$$

where the σ_i denote the cross sections of the Higgs boson production processes i and $Br_{H \rightarrow c\bar{c}}$ is the branching ratio of the Higgs boson to charm quarks, $\Gamma_{H \rightarrow c\bar{c}}/\Gamma_H$. The superscript ‘‘SM’’ indicates the SM values, its absence refers to values obtained under variations of the signal strength. In the κ framework all couplings of particles X to the Higgs boson are independently modified by multiplying them with some κ_X ; here and in the following we will assume that only the charm Yukawa coupling is modified by a factor κ_c , which in turn will result in modified partial and total decay widths of the Higgs boson, and thus

$$\mu_i = \left[\frac{\kappa_c^2 \Gamma_{H \rightarrow c\bar{c}}^{SM}}{\Gamma_H^{tot}} \right] / \left[\frac{\Gamma_{H \rightarrow c\bar{c}}^{SM}}{\Gamma_H^{tot, SM}} \right] \quad (3)$$

Current analysis with such a framework sets the expected limit at 95% confidence at $\mu_{VH(c\bar{c})} \leq 31^{+12}_{-8}$ at ATLAS [26] measured at an integrated luminosity of 139 fb^{-1} and $\mu_{VH(c\bar{c})} \leq 37^{+16}_{-11}$ at CMS [2] measured at 35.9 fb^{-1} . The quadratic dependence of κ_c on μ leads to some limitations in the maximal resolvable value of μ that leads to a meaningful value of κ_c . This value sits at $\mu = 1/Br_{H \rightarrow c\bar{c}}^{SM} = 34.6$.

2. Simulation

For our analysis, signal and background samples are generated with SHERPA 2.2.9 [27], using LO-merged samples throughout [28–30]. For all signal and background processes considered, namely;

- vector boson fusion (0L),
- WH -associated production (1L),
- ZH -associated production (2L),
- vector boson ($W \rightarrow \ell\nu$, $Z \rightarrow \ell\bar{\ell}$) production + jets,
- vector-boson pair production + jets, where one of the two bosons decays hadronically,
- pure QCD multijets,
- $t\bar{t}$ production + jets,

we merge up to two additional jets to the core process. We use the NNPDF 3.0 PDF [31] from LHAPDF [32], the COMIX matrix element generator [33] for the LO matrix element, the CSSHOWER [34] for the simulation of QCD radiation, AHADIC++ as hadronisation model [35], PHOTONS++ [36] for the emission of photons in the decays of the W and Z bosons, and SHERPA’s built-in models for the underlying event and hadron decays. Using the default prescription for setting the renormalisation and factorisation scales in multi-jet merging, we obtain theoretical uncertainties from their variation by a factor of $f_{R,F} = 2$ in both directions and forming the envelope of the 7-points, schematically

$$\{f_R, f_F\} = \left\{ \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, 1; 1, \frac{1}{2}; 1, 1; 1, 1; 2; 2, 1; 2, 2 \right\}. \quad (4)$$

The rationale for using this approach is two-fold: First of all, including higher-order QCD corrections does not induce any sizable change in the shape of distributions, but only alters the overall cross section which can usually be captured by applying a flat K -factor to the overall sample. For the processes we consider, this K -factor is of the order of 1.3 or below, thereby increasing the total number of events by up to 30%, which in turn translates to a decrease of the statistical uncertainty by about 10%. Secondly, apart from increasing total

Table 1
Cut flow for each channel.

Cut	#	‘‘0L’’	‘‘1L’’	‘‘2L’’
MET $\leq 30 \text{ GeV}$	1	✓	X	✓
MET $\geq 30 \text{ GeV}$	1	X	✓	X
0 Isolated Leptons	2	✓	X	X
1 Isolated Leptons	2	X	✓	X
2 Isolated Leptons	2	X	X	✓
1+ Fat jet	3	✓	✓	✓
Candidate Fat jet	4	✓	✓	✓
2 forward QCD jets	5	✓	X	X
$p_T + P_{T,L_1}$ and J B2B	5	X	✓	X
$P_{T,L_0} + P_{T,L_1}$ and J B2B	5	X	X	✓
1+ Secondary vertices	6	✓	✓	✓
2 sub-jets	7	✓	✓	✓
Simple Vertex cuts	8	✓	✓	✓
Machine Learning cuts	9	✓	✓	✓

event numbers, the higher-order corrections reduce scale uncertainties, typically by a factor of 2 or more. As we are only able to roughly estimate experimental uncertainties, our approach of potentially overestimating the theory uncertainties therefore merely translates into our results being more conservative.

3. Analysis strategy

3.1. Initial cuts

For each of the three signal topologies, there is a unique set of cuts, summarised in Table 1. They are encoded in a RIVET [37] analysis and detailed by:

1. MET is reconstructed from the total sum of visible particles, with

$$|\eta| < 4 \text{ and } p_T > 100 \text{ MeV} \quad (5)$$

and it is particularly powerful in enhancing or suppressing events with (1L) and without (0L, 2L) decaying W bosons.

2. to isolate leptons, we demand that the total transverse energy of all particles in a cone of size $R_{\text{iso}} = 0.2$ around the lepton direction is constrained by 5% of the lepton transverse energy $E_{T,\ell}$,

$$\sum_{\Delta R_i < R_{\text{iso}}} E_{T,i} \leq 0.05 \cdot E_{T,\ell} \quad (6)$$

We require the exact number of 0, 1, or 2 isolated leptons for the 0L, 1L, and 2L topologies.

3. we demand the Higgs decay products to form a fat jet, defined by the anti- k_T algorithm [38,39] with $R = 1.0$ and $p_T > 250 \text{ GeV}$, and we require events to contain at least one such fat jet.
4. to identify the required single candidate fat jet, the highest- p_T fat jet must contain at least three particles, but no isolated lepton, and its invariant mass must satisfy

$$75 \text{ GeV} < m_J < 175 \text{ GeV}, \quad (7)$$

cf. Fig. 1 for an illustration that motivates our choice.

5. in addition, we place some cuts that uniquely identify specific signal topologies:

- for the VBF (0L) topology, we require two forward anti- k_T jets ($R = 0.4$, $p_T > 20 \text{ GeV}$), and with a minimal rapidity separation and combined invariant mass,

$$\Delta y_{jj} > 2.5 \text{ and } m_{jj} > 400 \text{ GeV}. \quad (8)$$

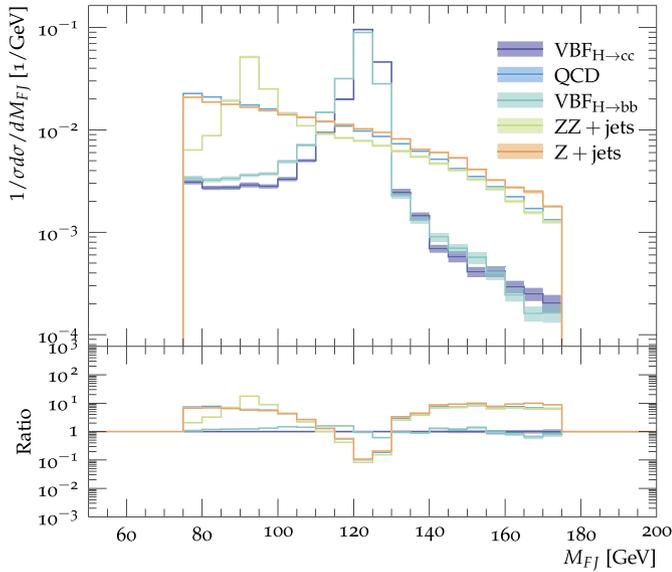


Fig. 1. Normalised histogram of the reconstructed fat jet mass in the “0L” channel. Shaded region indicates the error estimated from PDF4LHC scale variation convention and statistics.

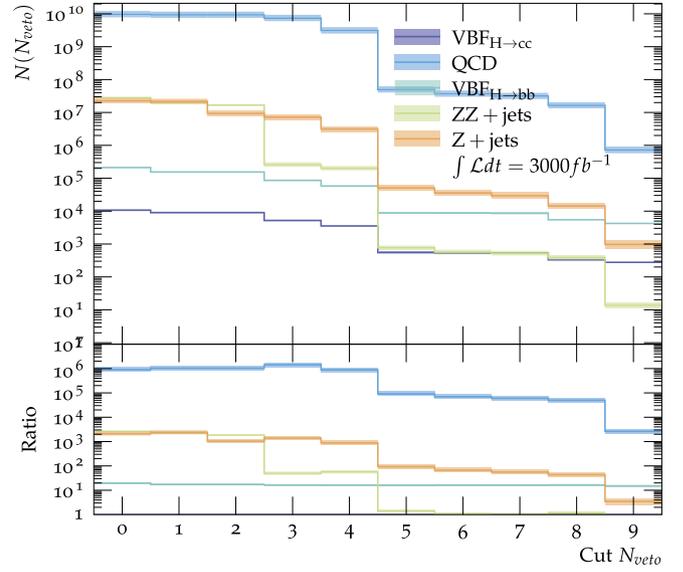


Fig. 2. Example cut flow for the “0L” channel. The Nth cut, N_{veto} is performed in the Nth bin of the x-axis. Shaded region indicates the error estimated from PDF4LHC scale variation convention and statistics. Backgrounds where $N_{\text{process}} << N_{\text{VBF}_{H \rightarrow c\bar{c}}}$ are omitted.

- for the two VH topologies (1L and 2L), we demand that the combined momentum of the two isolated leptons (2L) or of the isolated lepton and MET (1L) is anti-parallel to the fat-jet momentum within a tolerance of $R = 0.4$.
- 6. The fat jet must contain at least one reconstructed secondary vertex with at least two charged tracks. An adapted vertex fitter [40,41] performs a minimisation of weighted impact parameters, d_i^2 over points of closest approach of charged particles contained within the fat jet. Vertices within 1 mm of each other are considered unresolved and are merged.
- 7. The fat jet must contain at least two sub-jets (anti- k_T : $\Delta R = 0.4$, $p_T > 20$ GeV).
- 8. Further cuts are applied on the reconstructed primary vertex, namely mass flowing through it and the root mean square distance (RMSD) of particle tracks to the vertex,

$$M_{v_p} < 1 \text{ TeV and } RMSD_{v_p} < 3 \text{ mm,} \quad (9)$$

which are determined empirically studying histograms of distributions over signal and background.

In Fig. 2 we exhibit, as an example, the resulting cut flow for the 0L channel (vector boson fusion).

4. Machine learning improvements

4.1. ML “booster”

In a second step, we boost the cut-based analysis through a set of multivariate neural networks (MVA) trained with TensorFlow [42] on $\sim 10,000$ events of each background and signal processes that have survived the initial cuts:

1. “Observable”: a dense fully connected network trained with global event and fat jet features. A large selection of features is fed into the multivariate neural network. A number of global event and jet observables can be considered in order to best distinguish between the classes, and only the features with the strongest Shapley values [43] for the signal classes are used (see Fig. 3):

- jet mass: m_J ,
- missing transverse energy: \cancel{E}_T ,
- total perpendicular momentum in jet: $p_{J\perp}$,

$$p_{\text{jet}\perp} = \sqrt{p_x^2 + p_y^2} \ni \underline{p}_{\text{jet}} \cdot \hat{z} = |\underline{p}_{\text{jet}}|, \quad (10)$$

- 2-subjettiness: τ_2 [44], The sub-jets required for the N -subjettiness calculation are clustered by the k_t algorithm with $R = 1.0$ where clustering is stopped when exactly two sub-jets remain.
- sub-jet energy fraction: z_1 and z_2 ,

$$Z_i = \frac{E_{\text{sub-jet},i}}{E_J} \quad (11)$$

- planar flow: P_f [45]. These observables are standardised and normalised,

$$O'_i = \frac{O_i - \bar{O}}{\sigma_O} \quad (12)$$

$$O''_i = \frac{O'_i - \min(O)}{\max(O) - \min(O)}$$

This improves the gradient decent performance since it works more efficiently over variables with roughly equal ranges and magnitudes.

2. “Image”: a dense fully connected convolutional neural network trained on rotated “jet images”. We create “2D calorimeter” images for the fat jet centred on its axis, and apply simple pre-processing steps to standardise the images using standard computer vision techniques,
 - (a) Centre: Rotate the jet in (η, ϕ) such that the jet axis lies at $(0,0)$,
 - (b) Rotate: Rotate (η, ϕ) such that any sub-jets align on the ϕ -axis,
 - (c) Image set-up: An ‘image’ which spans $\eta', \phi' \in (-R, R)$ with 21×21 pixels,
 - (d) Build: For each particle, i in the jet add some variable x in the bin (η'_i, ϕ'_i) ,
 - (e) Scale: For each image scale such that $0 < I_{\eta', \phi'} < 255$,

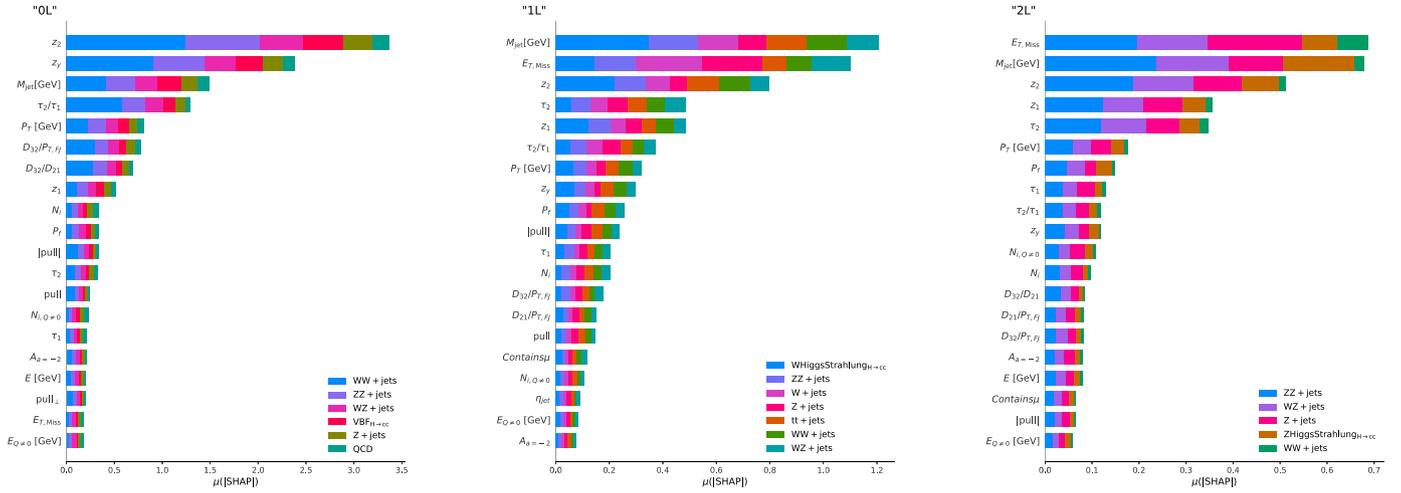


Fig. 3. Absolute Shapley contributions to classification of each coloured class in each channel "0L", "1L", "2L".

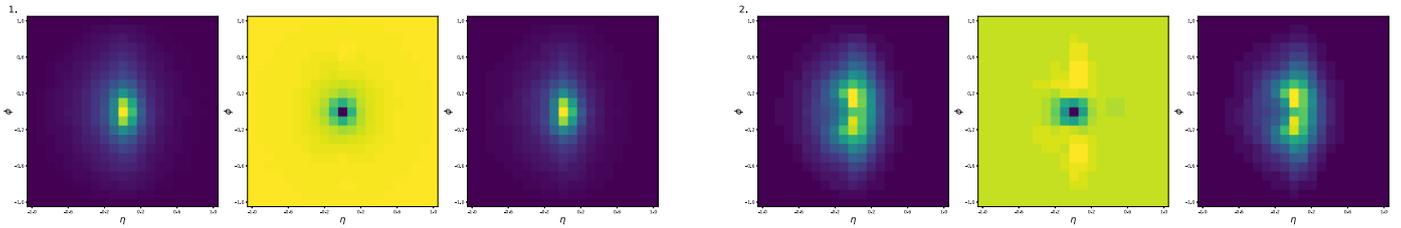


Fig. 4. Example mean fat jet images for signal and background: QCD (left) vs. VBF($H \rightarrow c\bar{c}$) (right).

The algorithm above can be extended from grey-scale to a colour image, in this work we consider RGB images where the pixel variable x is,

$$R = -\sum_{\text{particles}} \log(E_i/E_{\text{jet}})$$

$$G = -\sum_{\text{particles}} \log(p_{T,\text{jet}})$$

$$B = \sum_{\text{particles}, Q \neq 0}$$

Jet images exemplified in Fig. 4 are fed into the CNN. A pre-processing function centres and normalises the images to contain integer values between 0-255. Similarly to the observables dataset this improves the gradient decent performance, allowing the network to be more sensitive to the full range of pixel values and enables faster learning during back-propagation. The Image datasets also undergo some on-the-fly augmentation; once imported, there is a random chance of pixel shifting and randomised horizontal and vertical flipping. These changes artificially generate more data and enforce discrete symmetries and any make the network robust to any centring issues encountered.

3. "Flow": a dense fully connected recursive neural network trained on ordered particle level features within the fat jet. Particle level features are fed into a recursive neural network. To provide the neural network with a structured sequence of particles, up to 10 fat jet constituents are ordered in energy. The particle level features are:

- η displacement w.r.t. fat jet axis: $\Delta\eta = \eta_p - \eta_j$,
- ϕ displacement w.r.t. fat jet axis: $\Delta\phi = \phi_p - \phi_j$,
- Perpendicular momentum: $\log(p_{T,p})$,

- Perpendicular momentum fraction: $\log(\frac{p_{T,p}}{p_{T,J}})$,

- Energy: $\log(E_p)$,

- Energy fraction: $\log(\frac{E_p}{E_j})$,

- R displacement w.r.t. fat jet axis, $\Delta R = R_p - R_j$,

Again we standardise these particle features across the training data set and all particles,

$$F'_i = \frac{F_i - \bar{F}}{\sigma_F}$$

$$F''_i = \frac{F'_i - \min(F)}{\max(F) - \min(F)}. \quad (13)$$

The structure of each of these parts is shown in Appendix B, Fig. B.9.

The neural networks are trained over many epochs with a test-validation split of 90%-10%, and the network with the highest validation accuracy from any epoch is kept. We note that at this stage there is confusion between the $H \rightarrow c\bar{c}$ and $H \rightarrow b\bar{b}$ class as this ML "booster" makes no attempt to build a b and c jet classifier. The distinction between $H \rightarrow bb$ and $H \rightarrow cc$ is addressed with another network architecture, see below. In each channel we report overall retention of signal excluding $H \rightarrow b\bar{b}$. The results can be summarised in Table 2 for ϵ_s and ϵ_b the signal acceptance and background rejection efficiencies (excluding $H \rightarrow b\bar{b}$). Only the events which are predicted to be the signal class for the appropriate channel are kept. The trained models are converted into a format suitable to run natively in RIVET [37] using the Frugally Deep header library [46].

Table 2
Neural network efficiencies for each channel.

NN	"#L"	ϵ_s	ϵ_b
MVA	0	82.0%	77.8%
MVA	1	82.4%	76.0%
MVA	2	69.0%	82.0%

Table 3
Neural Network efficiencies for each channel.

NN	"#L"	ϵ_s	ϵ_b
Vertex MVA	0	54.8%	89.5%
Vertex MVA	1	42.2%	90.1%
Vertex MVA	2	46.6%	90.0%

4.2. ML charm vs. bottom discriminator

Finally, we construct a neural network to discriminate the $H \rightarrow c\bar{c}$ signal from the $H \rightarrow b\bar{b}$ background, based on the structure of the displaced vertices. There are many examples of superior classifiers for c and b jet classifiers used by the experiments, including MV2 and DL1 [47]. However, here we construct our own multivariate neural network that uses primary and secondary vertex features to discriminate between fat jets with only light constituents, with c hadrons, and with b hadrons. This network has two inputs,

1. "vertex observable": a time-distributed fully connected network with input features describing up-to 5 reconstructed vertices, V_i with 10 features:
 - number of reconstructed vertices: N_{V_i} ,
 - total number of tracks: N_{p,V_i} ,
 - vertex invariant mass: M_{V_i} ,
 - vertex Energy: E_{V_i} ,
 - distance from primary vertex: $D(V_i, V_p)$,
 - transverse distance from primary vertex: $D_T(V_i, V_p)$,
 - RMSD of impact parameters: $\sqrt{\overline{d_i^2}}$,
 - polar angle of vertex: θ_{V_i} ,
 - order of vertex: \mathcal{O}_{V_i} . \mathcal{O}_{V_i} is necessarily 0 for the primary vertex; then any vertices within a cone with opening angle $\theta = \pi/4$ are subsequently numbered in order of distance from the primary vertex. This provides the neural network with reinforcement of a natural ordering in displaced vertices in any event.
2. "vertex flow" uses particle-level features of the 5 hardest particles of each vertex V_i with the following inputs,
 - longitudinal impact parameter: $d_{L,p}$,
 - transverse impact parameter: $d_{T,p}$,
 - energy fraction: $\log(\frac{E_p}{E_j})$,
 - η displacement w.r.t. fat jet axis: $\Delta\eta = \eta_p - \eta_j$,
 - ϕ displacement w.r.t. fat jet axis: $\Delta\phi = \phi_p - \phi_j$,
 - R displacement w.r.t. fat jet axis: $\Delta R = R_p - R_j$.

All of these features are normalised over a weighted mean over all features for all classes, leading to the results summarised in Table 3. The vertex booster network was also independently trained on a streamlined data set consisting of $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$ and QCD fat jets with a minimal cut flow for comparison to other analyses. We find $\epsilon_{H \rightarrow c\bar{c}} = 72\%$ and background rejection $\epsilon_b = 75\%$. Comparing directly with the JetFitterCharm Algorithm [48] and demanding similar signal efficiencies we obtain background rejection rates summarised in Table 4.

Table 4
Summary of neural network efficiencies selected ϵ_c on the streamlined fat jet data set, resulting in ϵ_l light jet, ϵ_b bottom jet and ϵ_c charm jet efficiencies.

	ϵ_c	$1/\epsilon_b$	$1/\epsilon_l$
"Loose"	0.95	1.65	1.03
"Medium"	0.21	13.2	149

5. Results

5.1. Limitations of the κ framework

To determine the 95% confidence limit on the signal strength, μ_{0L} , μ_{1L} and μ_{2L} we use a CL_s [25] frequentist approach implemented in RooFit/RooStats [49–51] and treat SHERPA as a standard model simulator. $\mu(\kappa, Br)$ is derived from equation (3) as

$$\mu_c = \frac{\kappa_c^2}{1 + Br_{H \rightarrow c\bar{c}}^{SM}(\kappa_c^2 - 1)} \quad (14)$$

The CL_s method uses a binned likelihood to determine the confidence limits on μ_i . This likelihood incorporates uncertainties due to statistics $\sigma_{N,x} = \sqrt{N}$, luminosity $\sigma_L = 2.5\%$ [52], and scale variations σ_α . Therefore,

$$\mathcal{L}(\mu, \mathbf{s}, \mathbf{b}) = \mathcal{N}(L', L, \sigma_L) \mathcal{N}(\alpha'_s, \alpha_s, \sigma_{\alpha_s}) \mathcal{N}(\alpha'_b, \alpha_b, \sigma_{\alpha_b}) \times \prod_x \mathcal{P}(b_x + s_x, L'(\alpha'_b b'_x + \mu \alpha'_s s'_x)) \mathcal{N}(b'_x, b_x) \mathcal{N}(s'_x, s_x) \quad (15)$$

The uncertainties are parameterised with a Gaussian smearing over our expected values, and the priors \mathcal{N} and \mathcal{P} are Gaussian and Poisson distributions, respectively. Profiling the likelihood function for each channel determines the confidence limits as a function of μ_i , with $i = "0L", "1L"$ or $"2L"$. These independent channels are combined into one confidence limit which could be inverted to a confidence limit on κ_c . However, inverting equation (14) is not well defined for $\mu_c > 1/Br_{H \rightarrow c\bar{c}}$.

As we explore this region of μ -values with our analysis, and in order to avoid counter-intuitive results for κ_c , we only quote projections for limits on μ . Expanding on ideas formulated for example in [22,23], we also suggest an indirect measurement, in which $H \rightarrow c\bar{c}$ and $H \rightarrow b\bar{b}$ are combined to the signal class and modified together with μ_{cb} . This extension transforms Eq. (14) into,

$$\mu_{cb} = \frac{\kappa_c^2 Br_{H \rightarrow c\bar{c}}^{SM} + \kappa_b^2 Br_{H \rightarrow b\bar{b}}^{SM}}{(Br_{H \rightarrow c\bar{c}}^{SM} + Br_{H \rightarrow b\bar{b}}^{SM})} \times \frac{1}{(1 + Br_{H \rightarrow b\bar{b}}^{SM}(\kappa_b^2 - 1) + Br_{H \rightarrow c\bar{c}}^{SM}(\kappa_c^2 - 1))} \quad (16)$$

and the limitation to resolve κ_c becomes;

$$\frac{Br_{H \rightarrow b\bar{b}}}{(Br_{H \rightarrow b\bar{b}} + Br_{H \rightarrow c\bar{c}})} \frac{1}{1 - Br_{H \rightarrow c\bar{c}}} < \mu_{cb} < \frac{1}{Br_{H \rightarrow c\bar{c}} + Br_{H \rightarrow b\bar{b}}} \quad (17)$$

If we further introduce $H \rightarrow c\bar{c}$ and $H \rightarrow b\bar{b}$ discrimination in the cutflow with the $ML_{b \leftrightarrow c}$ network Eq. (16) becomes,

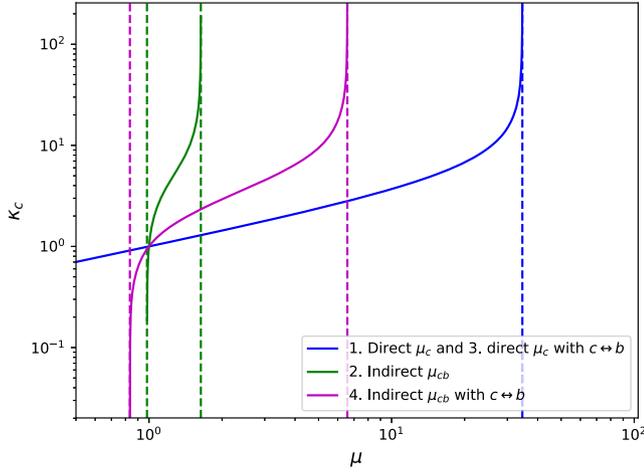


Fig. 5. $\kappa(\mu_i)$ plotted in three instances. Blue: direct $H \rightarrow c\bar{c}$ defined in equation (14). Green: The indirect measurement defined in Eq. (16). Purple: The indirect measurement with charm bottom discrimination defined in Eq. (18) evaluated with mean values of the efficiencies given in Table 3.

$$\mu_{cb} = \frac{\epsilon_c \kappa_c^2 Br_{H \rightarrow c\bar{c}}^{SM} + (1 - \epsilon_b) \kappa_b^2 Br_{H \rightarrow b\bar{b}}^{SM}}{(\epsilon_c Br_{H \rightarrow c\bar{c}}^{SM} + (1 - \epsilon_b) Br_{H \rightarrow b\bar{b}}^{SM})} \quad (18)$$

$$\times \frac{1}{(1 + Br_{H \rightarrow b\bar{b}}^{SM}(\kappa_b^2 - 1) + Br_{H \rightarrow c\bar{c}}^{SM}(\kappa_c^2 - 1))}$$

the factors ϵ_c and $1 - \epsilon_b$ re-weight the expected event counts contributing to the signal strength based on the performance of the charm bottom discriminator, $ML_{b \leftrightarrow c}$. Now the limitation on κ_c is given by;

$$\frac{(1 - \epsilon_b) Br_{H \rightarrow b\bar{b}}}{(1 - \epsilon_b) Br_{H \rightarrow b\bar{b}} + \epsilon_c Br_{H \rightarrow c\bar{c}}} \frac{1}{1 - Br_{H \rightarrow c\bar{c}}} < \mu_{cb} < \frac{\epsilon_c}{(\epsilon_c Br_{H \rightarrow c\bar{c}} + (1 - \epsilon_b) Br_{H \rightarrow b\bar{b}})}. \quad (19)$$

The limits of μ leading to a resolvable κ_c are determined for $\kappa_b = 1$. These expressions are illustrated in Fig. 5.

5.2. μ results

In each channel we consider the primary contributions of uncertainties in the determination of μ . This can be done by fixing nuisance parameters in turn and varying each quantity by its pre-fit (initial uncertainties fed into likelihood) and post-fit (post fitted values from maximisation of profiled likelihood) values. In Fig. 6 we exhibit four classes of uncertainties,

1. Statistical
2. Luminosity
3. Systematics (“Sys”)
4. Monte-carlo (“MC”)

Statistical uncertainties occur from the total counts in each bin in the likelihood function. The systematics from the 7-point envelope function in scale variations of f_F , f_R and lastly the Monte Carlo uncertainty from SHERPA. From this we can read off which backgrounds have the largest impact on the precise determination of μ : QCD, W +jets (W) production and Z +jets (Z) production processes for the “0L”, “1L” and “2L” channels, respectively. Profiling nuisance parameters to calculate their marginal error allows the determination of correlations between other parameters. It is

Table 5

Summary of the cut flows and neural networks architectures used in each of the methods and which of the κ_c, κ_b that are allowed to vary. Here $ML_{booster}$ refers to the ML “booster” network and $ML_{b \leftrightarrow c}$ the ML $b \leftrightarrow c$ discriminator network.

	Cut flow	$ML_{booster}$	$ML_{b \leftrightarrow c}$	κ dependence
1	✓	✓	X	$\mu_c(\kappa_c)$
2	✓	✓	X	$\mu_{cb}(\kappa_c, \kappa_b)$
3	✓	✓	✓	$\mu_c(\kappa_c)$
4	✓	✓	✓	$\mu_{cb}(\kappa_c, \kappa_b)$

worth stressing that measurements in the “2L” channel, while being the most sensitive channel in this analysis, are dominated by the statistics from the limited cross-section of the signal process and lower luminosity.

We explored four ways for the μ extraction, direct (1), indirect (2), direct with $b \leftrightarrow c$ discrimination (3), and indirect with $b \leftrightarrow c$ discrimination (4). We summarise the interplay of the different analysis steps and the variation of μ_c and μ_{cb} in Table 5. Fitting to distributions of various observables or to pairs of observables yields constraints of the signal strengths μ_c . The following observables showed the best discriminating power:

- planar flow: P_f ,
- 2-subjettiness: τ_2 ,
- boosted sub-jet separation angle: $\theta_{j1, j2}$,
- fat jet mass: M_J ,
- sub-jet energy fraction: Z_1 .

In Appendix A we show results for the four methods over all distributions for an integrated luminosity of 150 fb^{-1} in Fig. A.7, and for 3 ab^{-1} in Fig. A.8. The figures exhibit the 95% confidence limits of all considered 1 dimensional fits and 2 dimensional fits with the 1σ and 2σ uncertainty bands.

One way, in which the impact dominant statistical uncertainties can be counteracted, is by performing a two-dimensional fit and constraining the sum of bins on each axis to one another and therefore their uncertainty to one another. We summarise the μ limits we obtain in Table 6 for the standard choice $\mathcal{H}(M_J)$, used by the experimental analyses so far, and our most powerful combination of distributions, \mathcal{H}_{best} . Due to the low branching fraction of $H \rightarrow c\bar{c}$ we have a very low signal count at low luminosity, and moving forward into the high-luminosity phase of the LHC with around 3000 fb^{-1} we will be less limited by statistics. This impacts on the efficiency of the ML “booster” cuts, which greatly improve our confidence limit by a factor of 2 over the initial cuts at 150 fb^{-1} .

The two-dimensional fitting technique leads to an improvement in the obtained confidence limits by on average a factor of 2 over their one-dimensional counterparts. We also see that while distributions involving M_J provide good fits, other choices of observable work just as well or even better hinting at possible new avenues of exploration. The ML $b \leftrightarrow c$ discriminator network provides an improvement to the value of the confidence limit of a factor 2 in the direct case but only 1.1 in the indirect case. Our best fit result is $\mu_c \leq 8.0_{-2.3}^{+3.6}$ at 150 fb^{-1} ($\kappa_c \leq 3.18_{-0.60}^{+0.94}$) at the 95% confidence limit. This result is compatible to SM within 4.0 standard deviations and is competitive with current ATLAS and CMS values of 31_{-8}^{+12} [26] and 37_{-11}^{+16} [2]. These results may have scope for enhancement by considering a wider range of features and multi-dimensional fitting rather than the using the “standard” choice as we have demonstrated in our findings.

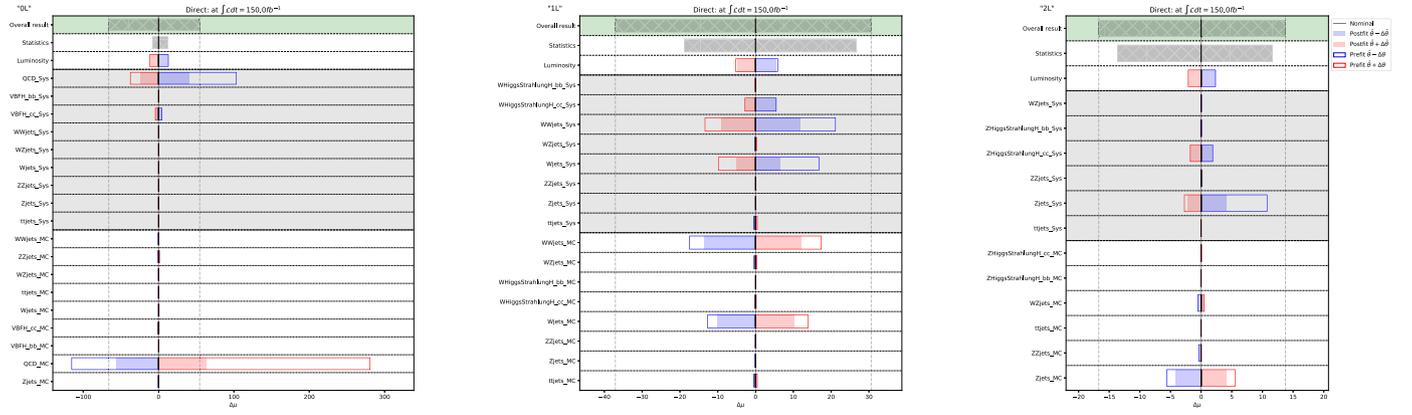


Fig. 6. Uncertainty contributions to the 95% confidence limit of signal strength μ for \mathcal{H}_{M_J} for the jet mass distribution. Uncertainties are shown for total statistics, luminosity, systematic uncertainty and Monte Carlo simulation uncertainty over all signal and background classes for each channel, “0L”, “1L”, “2L”, and compared with the total uncertainty. Pre-fit uncertainties shown in the outlined bars and post-fit uncertainties by the filled in bars.

Table 6

Combined μ bounds at 95% confidence for 4 methods for the \mathcal{H}_{M_J} and \mathcal{H}_{best} distributions. We compare results obtained from the fat jet distribution (\mathcal{H}_{M_J}) – the method of choice for the experimental analyses so far – with results we obtained from a best combination of two observables, \mathcal{H}_{best} and indicate them in the last column.

Search	μ w/ \mathcal{H}_{M_J}	μ w/ \mathcal{H}_{best}	\mathcal{H}_{best}
$\int \mathcal{L} dt = 150 \text{ fb}^{-1}$			
Direct, μ_c	$53.7^{+22.7}_{-15.9}$	$42.7^{+17.4}_{-12.4}$	$\mathcal{H}(M_J, P_f)$
Indirect, μ_{cb}	$4.0^{+1.3}_{-0.9}$	$3.1^{+1.0}_{-0.7}$	$\mathcal{H}(M_J, \theta_p)$
Direct, μ_c with $b \leftrightarrow c$ discrimination	$48.1^{+19.2}_{-13.8}$	$8.0^{+3.6}_{-2.3}$	$\mathcal{H}(Z_1, P_f)$
Indirect, μ_{cb} with $b \leftrightarrow c$ discrimination	$4.7^{+1.6}_{-1.1}$	$2.0^{+0.6}_{-0.4}$	$\mathcal{H}(Z_1, P_f)$
$\int \mathcal{L} dt = 3000 \text{ fb}^{-1}$			
Direct, μ_c	$35.5^{+14.0}_{-10.3}$	$12.1^{+5.1}_{-3.4}$	$\mathcal{H}(M_J, P_f)$
Indirect, μ_{cb}	$3.0^{+0.8}_{-0.6}$	$1.5^{+0.2}_{-0.2}$	$\mathcal{H}(M_J, \theta_p)$
Direct, μ_c with $b \leftrightarrow c$ discrimination	$33.9^{+13.2}_{-8.9}$	$2.1^{+0.6}_{-0.4}$	$\mathcal{H}(Z_1, P_f)$
Indirect, μ_{cb} with $b \leftrightarrow c$ discrimination	$4.0^{+1.2}_{-0.8}$	$1.1^{+0.1}_{-0.1}$	$\mathcal{H}(Z_1, P_f)$

Moving into the high-luminosity regime we see a again an enhancement in the benefit from 2D fits by a factor of about 2 over 1D fits. At 3 ab^{-1} our best fit μ values tighten and the limits are now resolvable under the κ framework. The direct measurement provides the best expected limit of $\kappa_c \leq 1.47^{+0.21}_{-0.16}$ at the 95% confidence limit.

6. Conclusions

We studied prospects for a determination of the charm Yukawa coupling or its constraints with present and future LHC data. We considered the production of the Higgs boson in Higgstrahlung and weak boson fusion processes, leading to final states with 0, 1, or 2 leptons, and the subsequent decay of the Higgs boson into a fat jet. We augmented a simple cut-based strategy with a multi-variate “booster” step and showed that this enhances the sensitivity of the analysis. We also investigated the impact of a neural-network based discriminator for fat jets containing only light partons, charm or bottom quarks and found a non-negligible impact. As a by-product we suggested an indirect measurement strategy where the branching ratio of a Higgs boson into heavy quarks – charm or bottom – is used in conjunction with the known value of the bottom Yukawa coupling to infer the charm Yukawa coupling. The signal strength in any case is extracted from fits to observable distributions, and we found that the fat jet mass is not necessarily the best-suited observable. We also found that two-dimensional fits further boost the sensitivity by about factors of

two to four compared to fits to a single observable, motivating further investigations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We are indebted to Marumi Kado and Francesco Di Bello for enlightening discussions and important feedback throughout the project, and we wish to thank the other members of the SHERPA team for ongoing collaboration.

This work was supported by the UK Science and Technology Facilities Council (STFC) under grant ST/P001246/1. FK acknowledges support from the European Union’s Horizon 2020 research and innovation programme as part of the Marie Skłodowska-Curie Innovative Training Network MCnetITN3 (grant agreement no. 722104), and support by the Royal Society and Wolfson Foundation under award RSWF\R1\191029.

Appendix A. Likelihood distributions

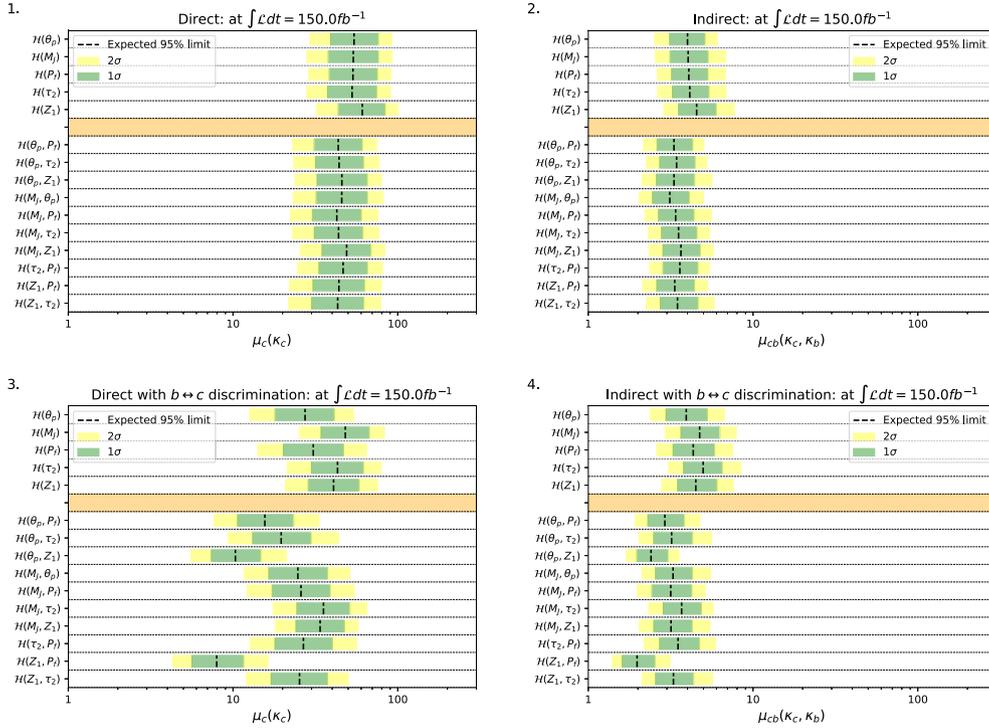


Fig. A.7. Comparison over binned likelihood distributions showing the 95% confidence limit on the signal strength μ .

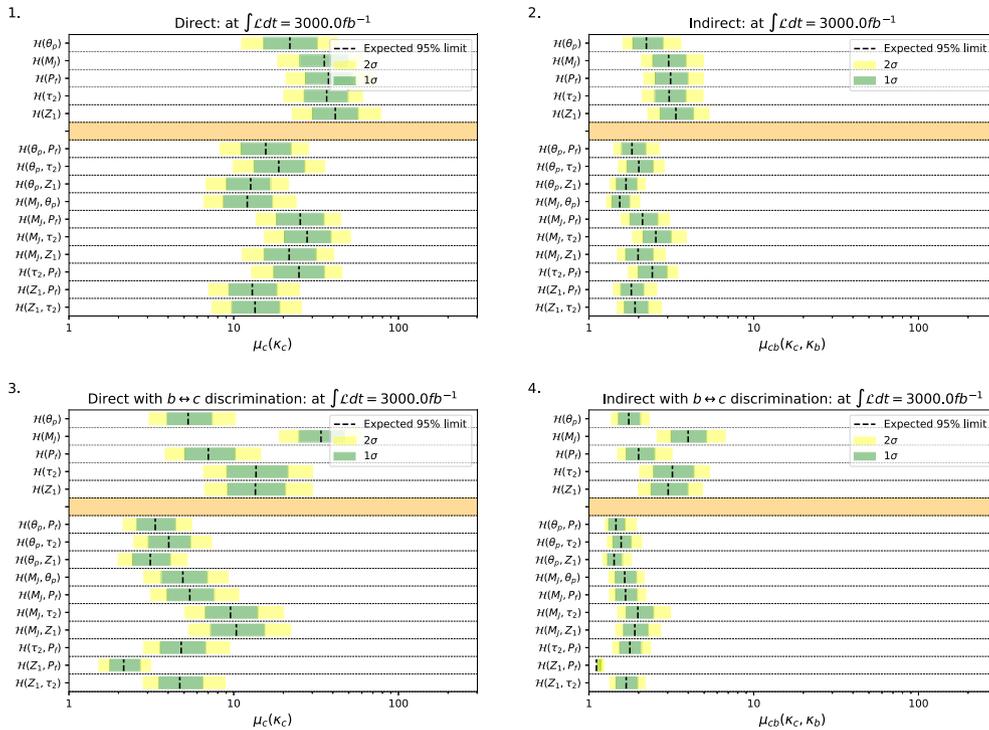


Fig. A.8. Comparison over binned likelihood distributions showing the 95% confidence limit on the signal strength μ .

Appendix B. Neural network architectures

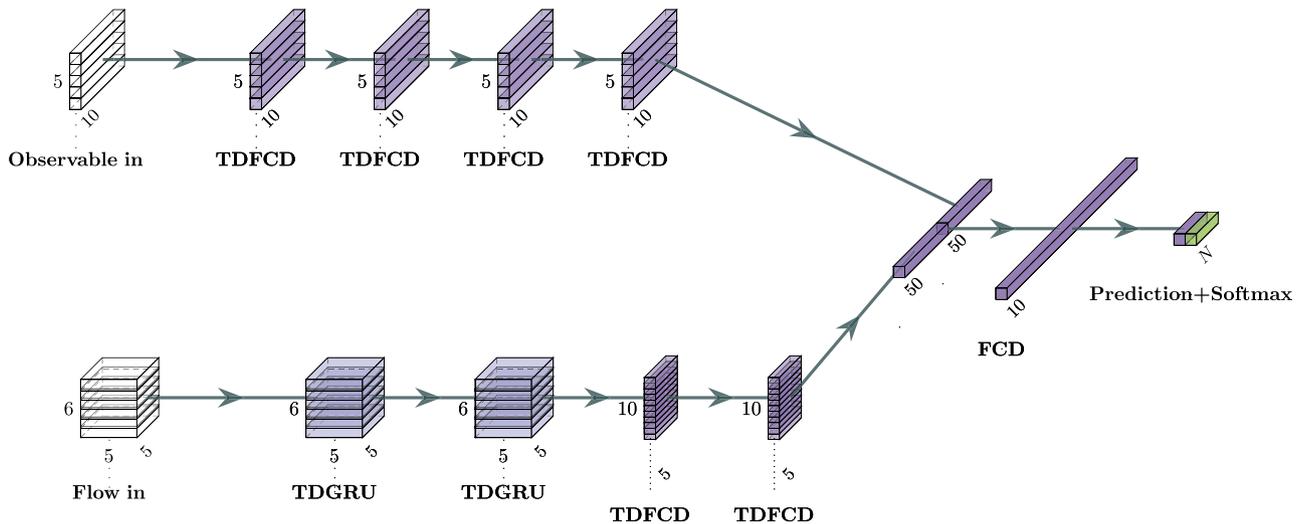
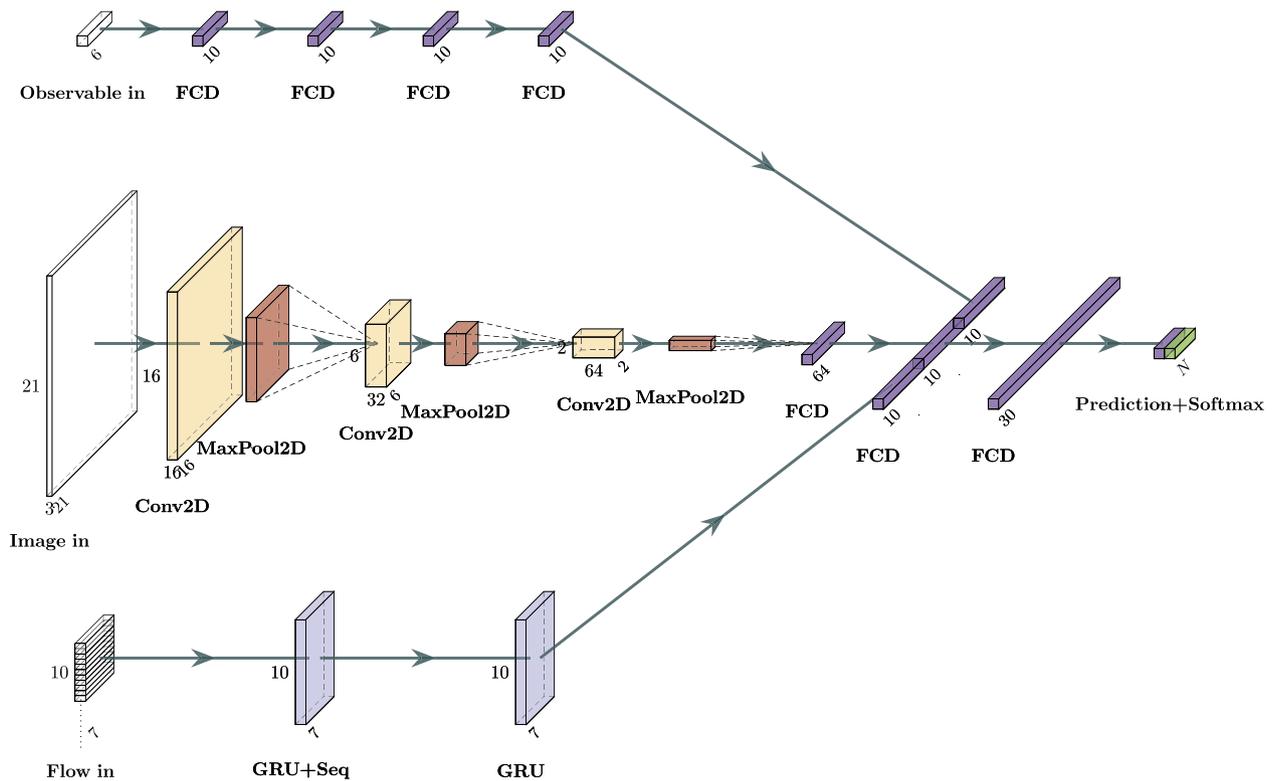


Fig. B.9. The neural network architectures for the $ML_{booster}$ (top) and $ML_{b \rightarrow c}$ (bottom). The abbreviation layer names are; **FCD**: Fully connected dense layer, **Conv2D**: A two dimensional convolutional layer, **MaxPool2D**: A two dimensional maximum pooling layer, **GRU+Seq**: A Gated recurrent unit which returns output from each unit not only the last, **GRU**: A gated recurrent unit, **TDFCD**: Time distributed fully connected dense layer and lastly **TDGRU**: Time distributed gated recurrent unit. Drop out layers used in training are omitted from these diagrams.

References

[1] ATLAS Collaboration, Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector, 2021.

[2] CMS Collaboration, Search for the standard model Higgs boson decaying to charm quarks, 2019.

[3] LHC Higgs Cross Section Working Group, LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle, 2012.

[4] ATLAS Collaboration, 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) (2012) 1–29, <https://doi.org/10.1016/j.physletb.2012.08.020>.

[5] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716 (2012) 30–61, <https://doi.org/10.1016/j.physletb.2012.08.021>, arXiv:1207.7235.

[6] L. Evans, P. Bryant, LHC machine, J. Instrum. 3 (08) (2008).

[7] Particle Data Group, Review of particle physics, Prog. Theor. Exp. Phys. 8 (2020), <https://doi.org/10.1093/ptep/ptaa104>.

- [8] ATLAS Collaboration, Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \rightarrow WW_* \rightarrow l\nu l\nu$ channel in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector, Phys. Lett. B 798 (2019) 134949, <https://doi.org/10.1016/j.physletb.2019.134949>.
- [9] CMS Collaboration, Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state, Phys. Rev. D 104 (2021) 052004, <https://doi.org/10.1103/PhysRevD.104.052004>, arXiv:2104.12152.
- [10] ATLAS Collaboration, Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector, Phys. Lett. B 784 (2018) 173–191, <https://doi.org/10.1016/j.physletb.2018.07.035>.
- [11] CMS Collaboration, Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s}=13$ TeV, Eur. Phys. J. C 81 (2020) 378, <https://doi.org/10.1140/epjc/s10052-021-09014-x>, arXiv:2011.03652.
- [12] ATLAS Collaboration, 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 (2019) 072001, <https://doi.org/10.1103/PhysRevD.99.072001>.
- [13] CMS Collaboration, Observation of the Higgs boson decay to a pair of tau leptons with the CMS detector, Phys. Lett. B 779 (2018) 283–316, <https://doi.org/10.1016/j.physletb.2018.02.004>.
- [14] ATLAS Collaboration, Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector, Phys. Lett. B 816 (2021) 136204, <https://doi.org/10.1016/j.physletb.2021.136204>.
- [15] CMS Collaboration, Observation of Higgs boson decay to bottom quarks, Phys. Rev. Lett. 121 (2018) 121801, <https://doi.org/10.1103/PhysRevLett.121.121801>, arXiv:1808.08242.
- [16] CMS Collaboration, Evidence for Higgs boson decay to a pair of muons, J. High Energy Phys. 2101 (2020) 148, [https://doi.org/10.1007/JHEP01\(2021\)148](https://doi.org/10.1007/JHEP01(2021)148), arXiv:2009.04363.
- [17] ATLAS Collaboration, A search for the dimuon decay of the standard model Higgs boson with the ATLAS detector, Phys. Lett. B 812 (2021) 135980, <https://doi.org/10.1016/j.physletb.2020.135980>.
- [18] B. Carlson, T. Han, S.C.I. Leung, Higgs boson to charm quark decay in vector boson fusion plus a photon, Phys. Rev. D 104 (7) (2021) 073006, <https://doi.org/10.1103/PhysRevD.104.073006>, arXiv:2105.08738.
- [19] G.T. Bodwin, F. Petriello, S. Stoynev, M. Velasco, Higgs boson decays to quarkonia and the $Hc\bar{c}$ coupling, Phys. Rev. D 88 (5) (2013) 053003, <https://doi.org/10.1103/PhysRevD.88.053003>, arXiv:1306.5770.
- [20] G.T. Bodwin, H.S. Chung, J.-H. Ee, J. Lee, F. Petriello, Relativistic corrections to Higgs boson decays to quarkonia, Phys. Rev. D 90 (11) (2014) 113010, <https://doi.org/10.1103/PhysRevD.90.113010>, arXiv:1407.6695.
- [21] G.T. Bodwin, H.S. Chung, J.-H. Ee, J. Lee, New approach to the resummation of logarithms in Higgs-boson decays to a vector quarkonium plus a photon, Phys. Rev. D 95 (5) (2017) 054018, <https://doi.org/10.1103/PhysRevD.95.054018>, arXiv:1603.06793.
- [22] C. Delaunay, T. Golling, G. Perez, Y. Soreq, Enhanced Higgs boson coupling to charm pairs, Phys. Rev. D 89 (3) (2014) 033014, <https://doi.org/10.1103/PhysRevD.89.033014>, arXiv:1310.7029.
- [23] G. Perez, Y. Soreq, E. Stamou, K. Tobioka, Constraining the charm Yukawa and Higgs-quark coupling universality, Phys. Rev. D 92 (3) (2015) 033016, <https://doi.org/10.1103/PhysRevD.92.033016>, arXiv:1503.00290.
- [24] G. Perez, Y. Soreq, E. Stamou, K. Tobioka, Prospects for measuring the Higgs boson coupling to light quarks, Phys. Rev. D 93 (1) (2016) 013001, <https://doi.org/10.1103/PhysRevD.93.013001>, arXiv:1505.06689.
- [25] A.L. Read, Presentation of search results: the CLs technique, J. Phys. G, Nucl. Part. Phys. 28 (10) (2002) 2693–2704, <https://doi.org/10.1088/0954-3899/28/10/313>.
- [26] ATLAS Collaboration, Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector, 2022.
- [27] E. Bothmann, et al., Event generation with Sherpa 2.2, SciPost Phys. 7 (3) (2019) 034, <https://doi.org/10.21468/SciPostPhys.7.3.034>, arXiv:1905.09127.
- [28] S. Catani, F. Krauss, R. Kuhn, B.R. Webber, QCD matrix elements + parton showers, J. High Energy Phys. 11 (2001) 063, <https://doi.org/10.1088/1126-6708/2001/11/063>, arXiv:hep-ph/0109231.
- [29] F. Krauss, Matrix elements and parton showers in hadronic interactions, J. High Energy Phys. 08 (2002) 015, <https://doi.org/10.1088/1126-6708/2002/08/015>, arXiv:hep-ph/0205283.
- [30] S. Hoeche, F. Krauss, S. Schumann, F. Siegert, QCD matrix elements and truncated showers, J. High Energy Phys. 05 (2009) 053, <https://doi.org/10.1088/1126-6708/2009/05/053>, arXiv:0903.1219.
- [31] NNPDF Collaboration, Parton distributions for the LHC run II, J. High Energy Phys. 2015 (4) (2015), [https://doi.org/10.1007/jhep04\(2015\)040](https://doi.org/10.1007/jhep04(2015)040).
- [32] A. Buckley, et al., LHAPDF6: parton density access in the LHC precision era, Eur. Phys. J. C 75 (2015) 132, <https://doi.org/10.1140/epjc/s10052-015-3318-8>, arXiv:1412.7420.
- [33] T. Gleisberg, S. Höche, Comix, a new matrix element generator, J. High Energy Phys. 12 (2008) 039, <https://doi.org/10.1088/1126-6708/2008/12/039>, arXiv:0808.3674.
- [34] S. Schumann, F. Krauss, A Parton shower algorithm based on Catani-Seymour dipole factorisation, J. High Energy Phys. 03 (2008) 038, <https://doi.org/10.1088/1126-6708/2008/03/038>, arXiv:0709.1027.
- [35] J.-C. Winter, F. Krauss, G. Soff, A modified cluster-hadronisation model, Eur. Phys. J. C 36 (3) (2004) 381–395, <https://doi.org/10.1140/epjc/s2004-01960-8>.
- [36] M. Schönherr, F. Krauss, Soft photon radiation in particle decays in SHERPA, J. High Energy Phys. 12 (2008) 018, <https://doi.org/10.1088/1126-6708/2008/12/018>, arXiv:0810.5071.
- [37] A. Buckley, et al., Rivet user manual, Comput. Phys. Commun. 184 (2013) 2803–2819, <https://doi.org/10.1016/j.cpc.2013.05.021>, arXiv:1003.0694.
- [38] M. Cacciari, et al., The anti- k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063, <https://doi.org/10.1088/1126-6708/2008/04/063>, arXiv:0802.1189.
- [39] M. Cacciari, et al., FastJet user manual, Eur. Phys. J. C 72 (2012) 1896, <https://doi.org/10.1140/epjc/s10052-012-1896-2>, arXiv:1111.6097.
- [40] R. Frühwirth, W. Waltenberger, P. Vanlaer, Adaptive Vertex Fitting, 2007.
- [41] ATLAS Collaboration, Secondary Vertex Finding for Jet Flavour Identification with the ATLAS Detector, Jun. 2017.
- [42] M. Abadi, et al., TensorFlow: large-scale machine learning on heterogeneous systems, <http://tensorflow.org/>, 2015.
- [43] L.S. Shapley, A value for n-person games, in: Contributions to the Theory of Games (AM-28), vol. 2, 1952, pp. 307–318.
- [44] D. Napoletano, G. Soyez, Computing N -subjettiness for boosted jets, J. High Energy Phys. 2018 (2018) 1.
- [45] L.G. Almeida, et al., Substructure of high- p_T jets at the LHC, Phys. Rev. D 79 (7) (2009) 074017, <https://doi.org/10.1103/PhysRevD.79.074017>, arXiv:0807.0234.
- [46] T. Hermann, Frugally deep: header-only library for using Keras models in C++, <https://github.com/Dobiasd/frugally-deep>.
- [47] ATLAS Collaboration, ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}=13$ TeV, Eur. Phys. J. C 79 (11) (2019), <https://doi.org/10.1140/epjc/s10052-019-7450-8>.
- [48] ATLAS Collaboration, Performance and Calibration of the JetFitterCharm Algorithm for c-jet Identification, 2015.
- [49] L. Moneta, et al., The RooStats project, arXiv:1009.1003, 2011.
- [50] W. Verkerke, D. Kirkby, The RooFit toolkit for data modeling, 2003, arXiv:physics/0306116 [physics.data-an].
- [51] R. Brun, F. Rademakers, ROOT: an object oriented data analysis framework, Nucl. Instrum. Methods A 389 (1997) 81–86, [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X).
- [52] ATLAS Collaboration, Luminosity determination in pp collisions at $\sqrt{s}=13$ TeV using the ATLAS detector at the LHC, 2019.