Exploring the Higgs boson with Deep Learning.

Felipe Ferreira de Freitas

Towards recognizing the light facet of the Higgs Boson.

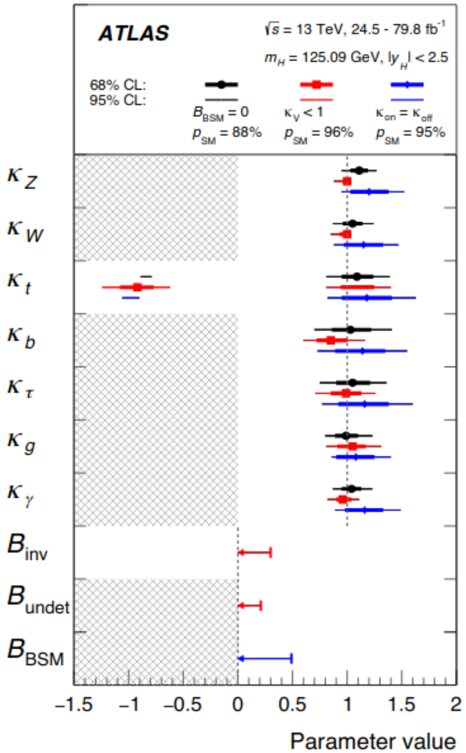
Alexandre Alves and Felipe F. Freitas

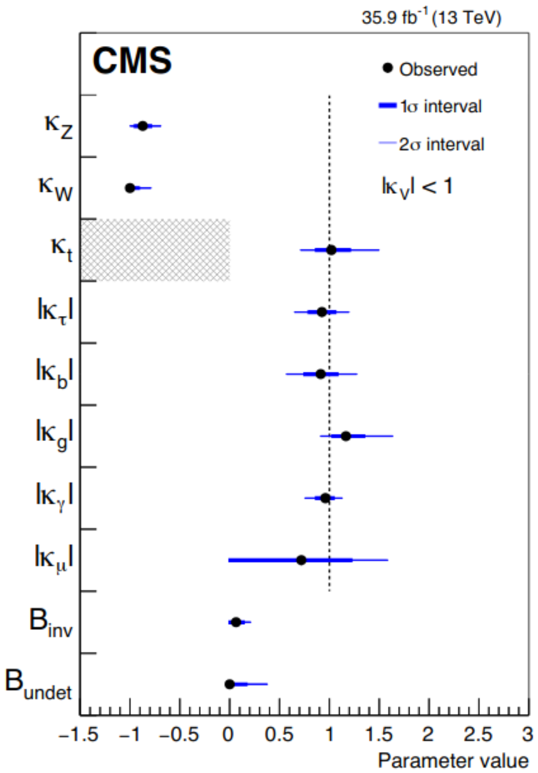
arXiv:1912.12532

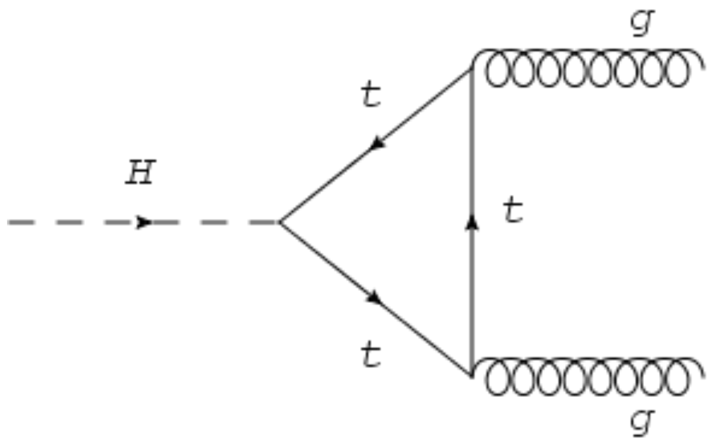
Overview

The Higgs boson couplings to bottom and top quarks have been measured and agree well with the Standard Model predictions.

On the other hand, decays to lighter quarks and gluons remain uncovered. Observing these decays is essential to complete the picture of the Higgs boson interactions.







$$\Gamma_{\mathrm{LO}}[H \to gg] = \frac{G_F \alpha_S^2 M_H^3}{36\sqrt{2}\pi^3} |\sum_Q \mathrm{A}_Q(\tau_Q)|^2$$

The gluons pair decay is completely buried beneath a huge background of jet pairs from leading order QCD interactions turning its observation practically impossible in the gluon fusion channel.

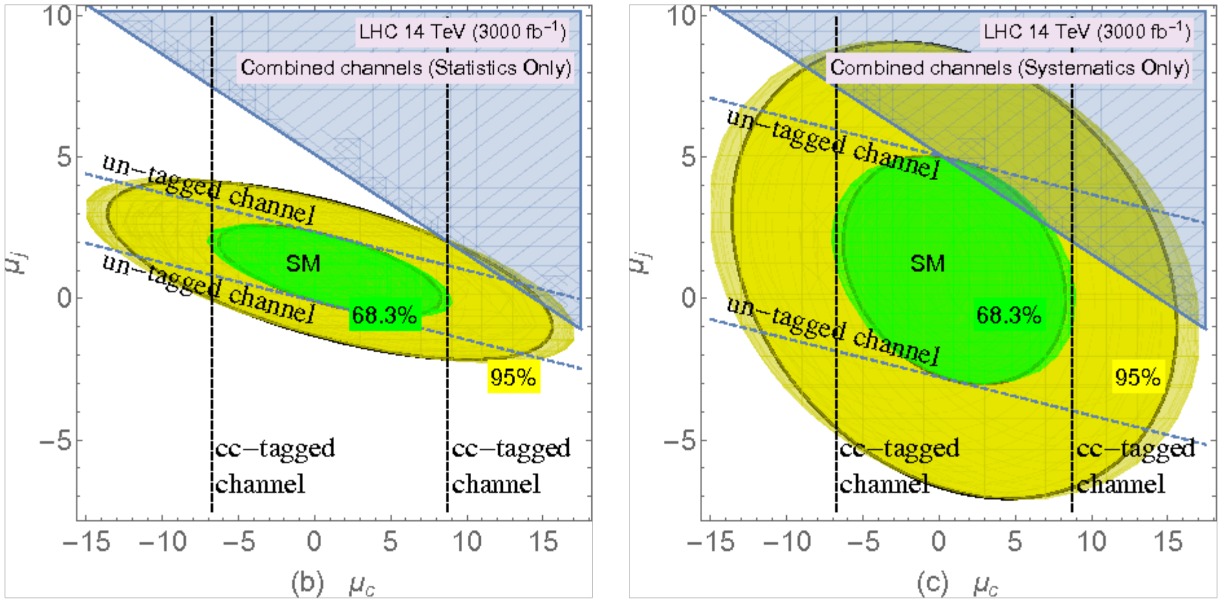
This motivates the search for untagged jets in Higgs decays in cleaner production channels:

- $pp \to VH, V = W, Z$

observation of the Higgs decay to light jets were found to be rather difficult:

- 1σ sensitivity after an integrated luminosityof 3000 fb^{1} and an upper bound of $BR(H \to jj) < 4 \times BR^{SM}(H \to gg)$, at 95% CL

Linda M. Carpenter, Tao Han, et al., Phys.Rev. D95 (2017) no.5, 053003 (2017-03-09)



- The authors suggested a multivariate analysis might improve the sensitivity of the LHC searches compared to the standard cut-and-count analysis.

Following that suggestion, we use machine learning (ML) in combination with Computer Vision (CV) techniques in order to improve the prospects to observe light quark and gluon jet pairs from Higgs boson decays.

We study the following processes:

- $q\bar{q}\,, gg \to Z(\to \ell^+\ell^-)\; H(\to gg),\;\; \ell = e, \mu$
- $q\bar{q}\,, gg \to Z(\to \ell^+\ell^-)\; H(\to b\bar{b})$
- $q\bar{q}\,, gg \to Z(\to \ell^+\ell^-)\; H(\to c\bar{c})$
- $pp \to Z(\to \ell^+\ell^-)\; j(jj)$
- $pp \to Z(\to \ell^+\ell^-)\; W(\to jj)$
- $pp \to Z(\to \ell^+\ell^-)\; Z(\to jj)$
- $pp \to t(\to W^-\bar{b})\; \bar{t}\,(\to W^+b)$

we also imposed the following cuts to further eliminate backgrounds:

at least two same-flavour opposite-charge leptons with:

- $|\eta_\ell| < 2.5,\ p_T^\ell > 30$ GeV
- $M_{\ell\ell} > 80$ GeV, $p_T^{\ell\ell} = (p_T^{\ell_1} + p_T^{\ell_2}) > 100$ GeV

at least one central fat-jet with:

- $|\eta_j| < 2.0,\ p_T^j > 150$ GeV
- $|M_j - m_H| < 20$ GeV

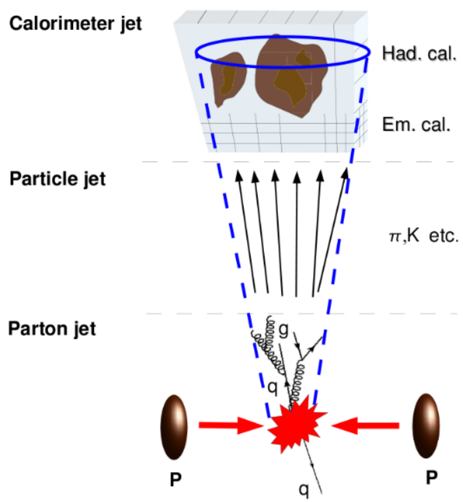
all the events must have

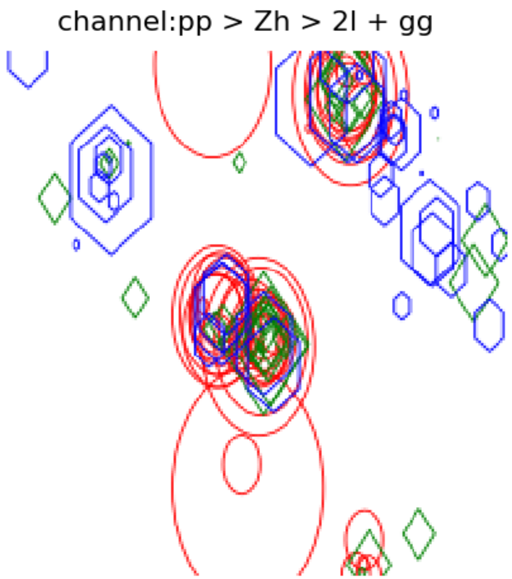
- $\not{E}_T < 40$ GeV

CONSTRUCTION OF ABSTRACT IMAGES

Delphes uses a particle-flow algorithm which produces two collections of 4-vectors:

- particle-flow tracks
- particle-flow towers





- EFlowTrack (b or c tagged quarks, charged pions, leptons, etc) $\longrightarrow$ Red circles.
- EFlowPhoton (photons)$\longrightarrow$ Green squares.
- EFlowNeutralHadrons (neutrons, neutral pions,...)$\longrightarrow$ Blue hexagons.
- centered at the $\eta \times \phi$ coordinates of the object and their radius are proportional to the logarithm of their transverse momentum.

The abstract image data set consists of:

- Zj(jj): 4779 images with 224 x 224, 8-bit/color RGBA.
- WZ $\rightarrow \ell^+\ell^-$ jj: 2760 images with 224 x 224, 8-bit/color RGBA.
- ZZ $\rightarrow \ell^+\ell^-$ jj: 3164 images with 224 x 224, 8-bit/color RGBA.
- ZH $\rightarrow \ell^+\ell^-\ b\bar{b}$: 29663 images with 224 x 224, 8-bit/color RGBA.
- ZH $\rightarrow \ell^+\ell^-\ c\bar{c}$: 37887 images with 224 x 224, 8-bit/color RGBA.
- ZH $\rightarrow \ell^+\ell^-\ gg$: 35280 images with 224 x 224, 8-bit/color RGBA.
- $t\bar{t} \rightarrow \ell^+\ell^-\ \nu_\ell\bar{\nu}_\ell\ b\bar{b}$: 204 images with 224 x 224, 8-bit/color RGBA.

CNN ARCHITECTURE AND TRAINING METHODOLOGY

We want to classify whether a given abstract image belongs to one of the 7 classes:

- the signal class ZH(jj)
- back-ground classes ZZ, WZ, Z+j(jj), ZH $(b\bar{b})$,ZH $(c\bar{c})$, $t\bar{t}$.

We use a ResNet-50 as our base architecture, with some modifications:

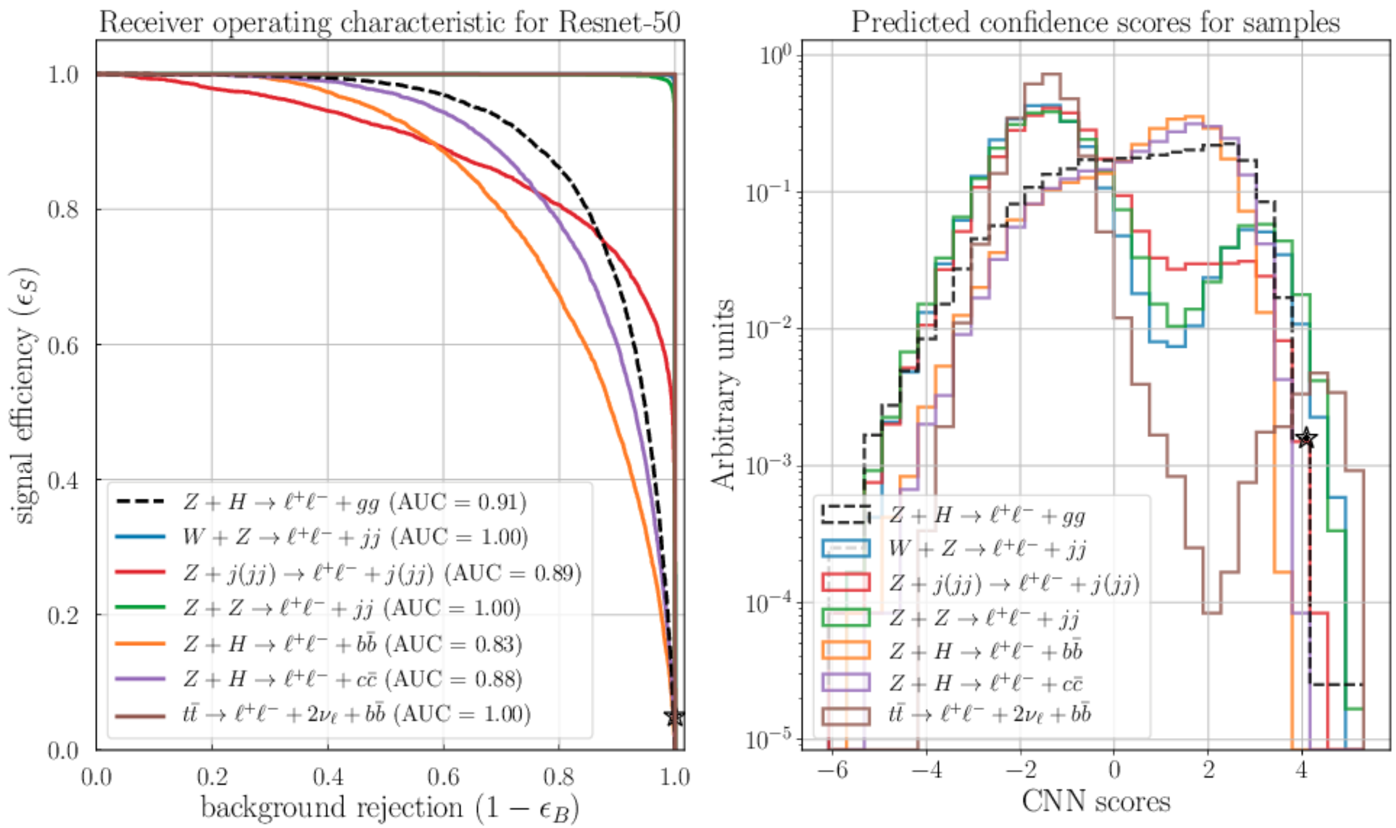
- an adaptive concatenate pooling layer (AdaptiveConcatPool2d)
- a flatten layer,
- a block with batch normalization, dropout, linear, and ReLU layers,
- a dense linear layer with 7 units as outputs, each unit corresponding to a class and a softmax activation function.

We trained our model in a 3-stage scheme:

1. end to the end (no transfer learning) for 50 epochs.
2. freeze the weights and biases up to the last 3 layers, and train for 25 epochs.
3. freeze all layers up to the last layer (the classification layer or header) and trained for 15 epochs.

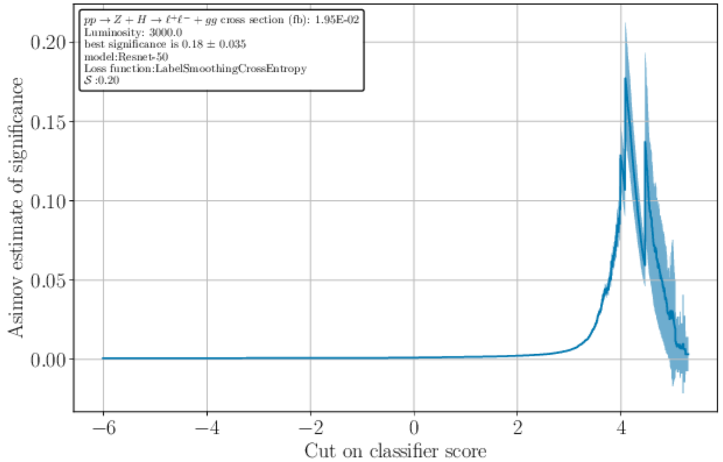
PERFORMANCE OF THE CLASSIFIERS

We can evaluate our CNN by looking into the ROC curves:



Or we can evaluate our quantitatively by using the Asimov Significance defined as:

$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$



Class	Number of events with ResNet-50
$ZH \to \ell^+\ell^- + gg$	2.3
$Zj(jj)$	76.4
$WZ \to \ell^+\ell^- + jj$	4.2
$ZZ \to \ell^+\ell^- + jj$	7.2
$ZH \to \ell^+\ell^- + b\bar{b}$	0
$ZH \to \ell^+\ell^- + c\bar{c}$	0
$t\bar{t} \to \ell^+\ell^- + \nu_\ell\bar{\nu}_\ell + jj$	38.0
total background	125.7

Classification with ResNet-50 and BDTs

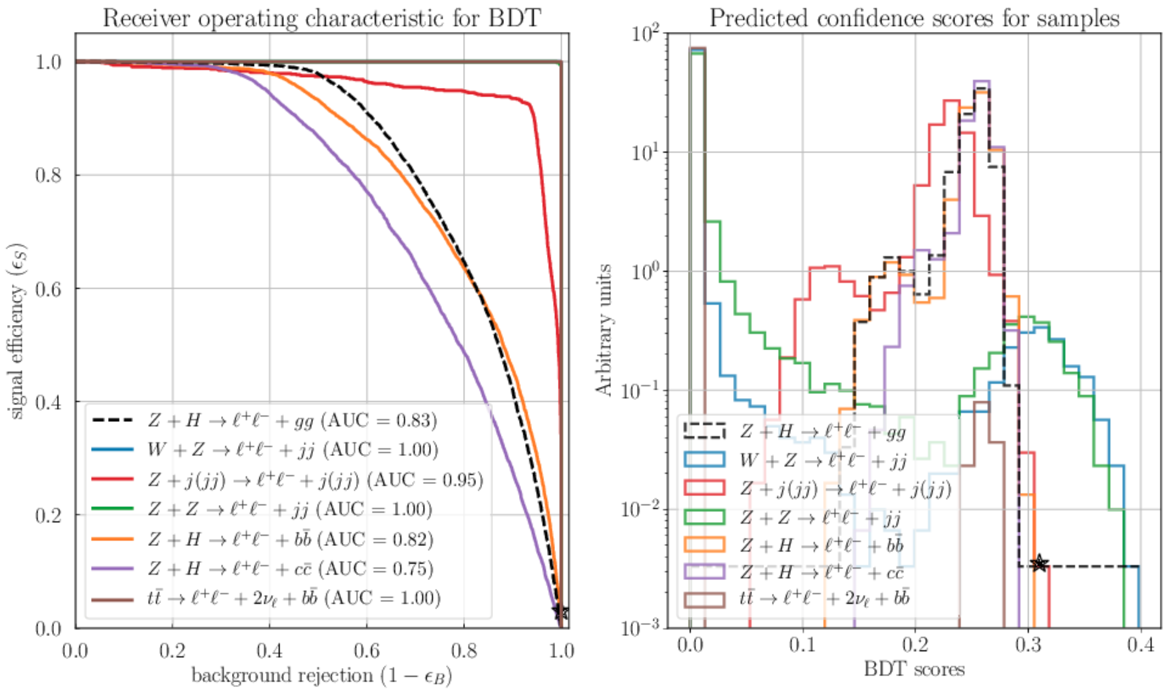
The representation of the data used to train the decision trees algorithm comprises the following variables:

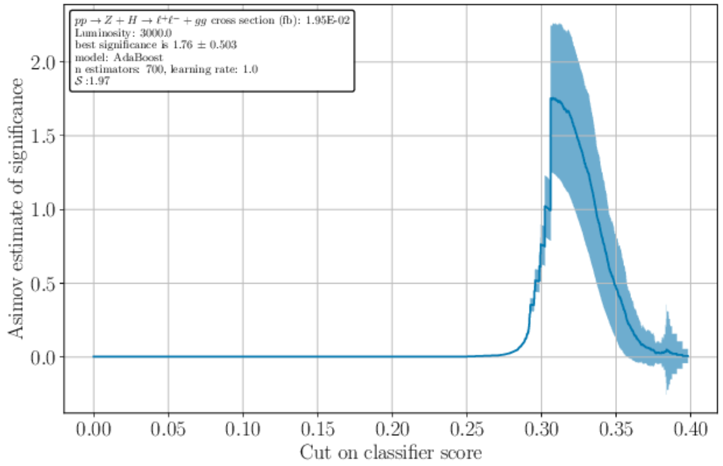
- $M_{\ell\ell}, M_{jj}, M_j, M_{j\ell\ell}$
- $p_T^{\ell\ell}, p_T^{j_1}, p_T^{j_2}, p_T^{b_1}, p_T^{b_2}, p_T^{\ell_1}, p_T^{\ell_2}$
- $\Delta R_{\ell\ell,j_1} = \sqrt{(\Delta\phi_{\ell\ell,j_1})^2 + (\Delta\eta_{\ell\ell,j_1})^2}, \Delta\phi_{\ell,\ell}, \cos(\Delta\phi_{\ell\ell,\ell_1}), \cos(\Delta\phi_{\ell\ell,j_1})$
- the score provided by the CNN classification for each one of the 7 classes.

We make use of Evolutionary Algorithm to search the best hyper-parameters which can achiev the highest significance for our signal. In our analysis we found that:

- multi-class AdaBoost classifier.
- 700 base estimators.
- a maximum tree depth of 5.
- a learning rate of 1.0.

Evaluating the BDT performance by looking into the ROC curves:





Class	Number of events with ResNet-50+BDT
$ZH \to \ell^+\ell^- + gg$	15.37
$Zj(jj)$	0
$WZ \to \ell^+\ell^- + jj$	31.2
$ZZ \to \ell^+\ell^- + jj$	29.7
$ZH \to \ell^+\ell^- + b\bar{b}$	0
$ZH \to \ell^+\ell^- + c\bar{c}$	0
$t\bar{t} \to \ell^+\ell^- + \nu_\ell\bar{\nu}_\ell + jj$	0
total background	60.9

where S_j is the mean signal significance obtained after the BDT classification computed with the simple significance metrics $S_j = \frac{s}{\sqrt{b+\sigma_b^2}}$

where $s = 15.37$ and $b = 60.9$ assuming $\sigma_b/b =$ 0(5\%)[10\%] uncertainties in the background normalization.

Signal significance and constraints on the light jet Higgs branching ratio:

We have to outline two major characteristics of our analysis:

1. we consider the two-lepton category only.
2. we include the signal contaminants $ZH(b\bar{b})$ and $ZH(c\bar{c})$ in the background category from the beginning.

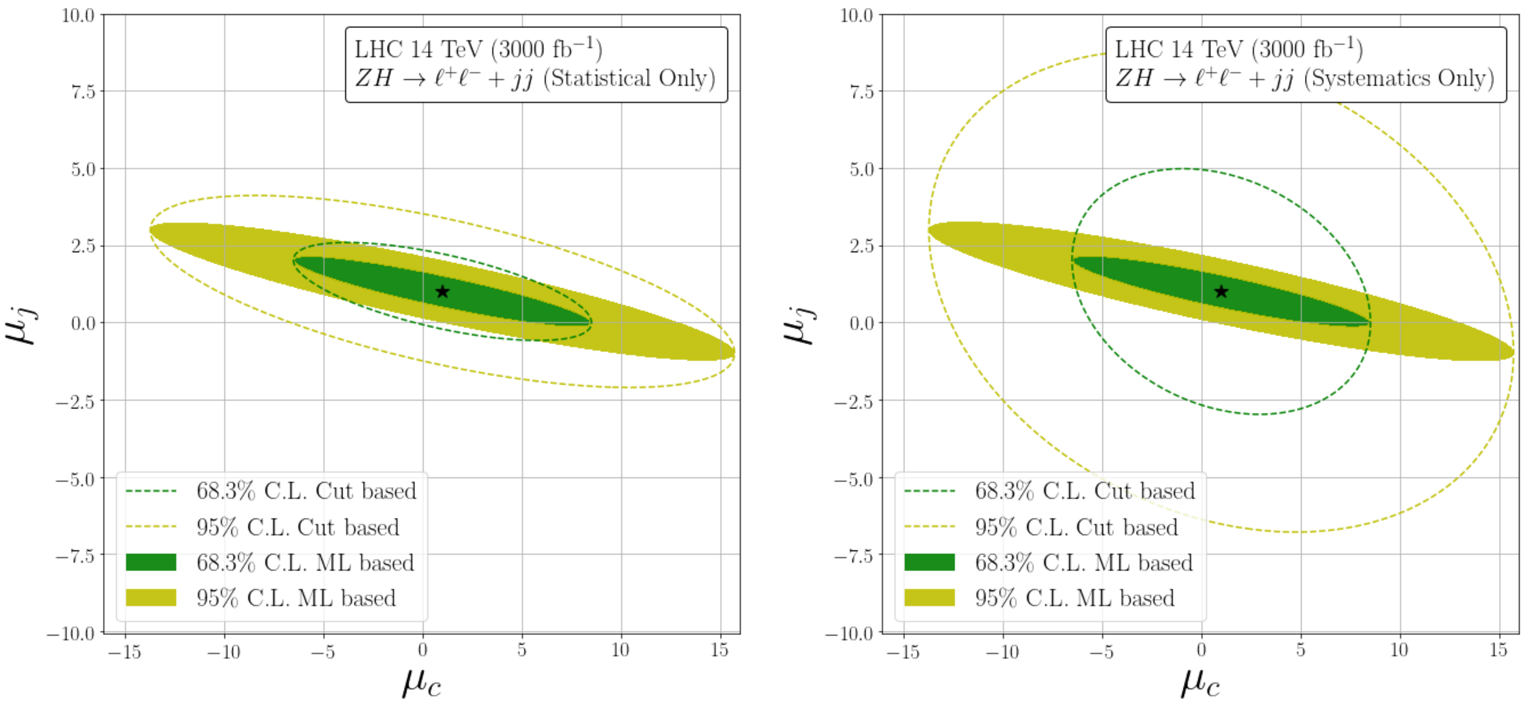
In previous work (Phys.Rev. D95 (2017) no.5), was considered:

1. one+two-lepton categories and WH is also take into account in the analysis.
2. The $ZH(b\bar{b})$ and $ZH(c\bar{c})$ categories are considered only in the statistical analysis to constrain the light jets branching ratio.

The efficient clearing up from $ZH(b\bar{b})$ and $ZH(c\bar{c})$ background events allows to place a direct upper bound on the Higgs to light jets branching ratio at the 95\% confidence level (CL)

$$\mu_j = \frac{\mathrm{BR}(H \to jj)}{\mathrm{BR}^{\mathrm{SM}}(H \to jj)} \leq 1 + \frac{\sqrt{\chi^2_{95\%}}}{S_j} = 2.0\ (2.07)\ [2.26],$$

$$\mathrm{BR}(H \to jj) \leq 2(2.07)[2.26] \times \mathrm{BR}^{\mathrm{SM}}(H \to gg)\ ,$$



We obtained the following 95\% CL upper bound limit for the Higgs branching ratio into untagged jets with 0\% and 1\% systematic errors, in parenthesis, for 3000 fb^{-1}

$$BR(H \to j' j') \leqslant 3.26(3.28) \times BR^{SM}(H \to gg).$$

Conclusions

- We employed several state-of-art ML techniques to improve the performance of the CNN algorithm in obtaining the highest signal significance possible.

- In spite of its power, the CNNs were not able to separate signal from backgrounds at the level we need, however, the output scores assigned by the CNNs to each event class is by themselves a very distinctive feature that can be combined with kinematic information of the particles of the event to train another ML algorithm

- Our methodology was able to reach $\sim 2\sigma$ in the statistics dominance scenario after 3000 fb^{-1}, despite even a 10% systematics on the backgrounds normalization

- the ML algorithm was able to eliminate the $Z(H \to b\bar{b})$ and $Z(H \to c\bar{c})$ contaminants allowing us to derive the following 95% CL bound directly on the light jets branching ratio:
  $$\mathrm{BR}(H \to jj) \leq 2(2.26) \times \mathrm{BR}^{SM}(H \to gg) \ ,$$
  assuming a 0(10)% systematic uncertainty on the background normalization.

- Combining the significance reached in this analysis with the ones in the search for $H \to b\bar{b}$ and $H \to c\bar{c}$ taking into account mixings of tagged and mistagged jet classes,
  $$\mathrm{BR}(H \to j'j') \leq 3.26(3.28) \times \mathrm{BR}^{SM}(H \to gg) \ ,$$
  which improves the results obtained exclusively with a dedicated cut-and-count analysis.